JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 117, D23114, doi:10.1029/2012JD018192, 2012

Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models

Satoshi Watanabe,[1] Shinjiro Kanae,[2] Shinta Seto,[3] Pat J.-F. Yeh,[4] Yukiko Hirabayashi,[1] and Taikan Oki[3]

Received 28 May 2012; revised 20 October 2012; accepted 23 October 2012; published 13 December 2012.

[1] Bias-correction methods applied to monthly temperature and precipitation data simulated by multiple General Circulation Models (GCMs) are evaluated in this study. Although various methods have been proposed recently, an intercomparison among them using multiple GCM simulations has seldom been reported. Moreover, no previous methods have addressed the issue how to adequately deal with the changes of the statistics of bias-corrected variables from the historical to future simulations. In this study, a new method which conserves the changes of mean and standard deviation of the uncorrected model simulation data is proposed, and then five previous bias-correction methods as well as the proposed new method are intercompared by applying them to monthly temperature and precipitation data simulated from 12 GCMs in the Coupled Model Intercomparison Project (CMIP3) archives. Parameters of each method are calibrated by using 1948–1972 observed data and validated in the 1974–1998 period. These methods are then applied to the GCM future simulations (2073–2097) and the bias-corrected data are intercompared. For the historical simulations, negligible difference can be found between observed and bias-corrected data. However, the differences in future simulations are large dependent on the characteristics of each method. The new method successfully conserves the changes in the mean, standard deviation and the coefficient of variation before and after bias-correction. The differences of bias-corrected data among methods are discussed according to their respective characteristics. Importantly, this study classifies available correction methods into two distinct categories, and articulates important features for each of them.

Citation: Watanabe, S., S. Kanae, S. Seto, P. J.-F. Yeh, Y. Hirabayashi, and T. Oki (2012), Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models, *J. Geophys. Res.*, *117*, D23114, doi:10.1029/2012JD018192.

1. Introduction

[2] The procedures for correcting biases in the General Circulation Model (GCM) simulations ("bias-correction") are important. The impacts of changing climate on the Earth's environment are of increasing interest, and GCMs have enabled the projections of future climate change caused by natural variability or anthropogenic activities [*Intergovernmental Panel on Climate Change* (*IPCC*), 2007]. Despite continuous efforts to improve GCM's capability of simulating historical climates, the use of bias-correction methods is essential for the impact assessment studies of climate change. This importance was described in the special report of the IPCC [*Seneviratne et al.*, 2012].

[3] In assessing potential hydrologic impacts of climate change [e.g., *Arnell*, 2004; *Oki and Kanae*, 2006], a suitable correction of biases in climate model projected temperature (*Ta*) and precipitation (*Pr*) is the principal focus due to its significant influence on the hydrologic response. For example, *Lehner et al.* [2006] assessed the impacts of global climate change on the risk of flood and drought by applying a hydrologic model forced with the bias-corrected atmospheric data. *Dettinger et al.* [2004] investigated the impacts of climate change on river flow in the Sierra Nevada of California by using bias-corrected *Ta* and *Pr* data from GCM simulations. Bias-correction methods were also applied to *Ta* and/or *Pr* data of River Thames [*Diaz-Nieto and Wilby*, 2005], Mekong River basin [*Kiem et al.*, 2008], and Tone River

[1]Institute of Engineering Innovation, School of Engineering, University of Tokyo, Tokyo, Japan.
[2]Department of Mechanical and Environmental Informatics, Tokyo Institute of Technology, Tokyo, Japan.
[3]Institute of Industrial Science, University of Tokyo, Tokyo, Japan.
[4]International Centre for Water Hazard Risk Management, UNESCO, Tsukuba, Japan.

Corresponding author: S. Watanabe, Institute of Engineering Innovation, School of Engineering, University of Tokyo, 2-11-16 Yayoi, Bunkyo, Tokyo, 113-8656 Japan. (stswata@sogo.t.u-tokyo.ac.jp)

0148-0227/12/2012JD018192

basin in Japan [*Takara et al.*, 2009]. In addition, bias-correction has also been applied to the Regional Climate Model (RCM) simulations such as the studies conducted in four basins of the United States [*Hay et al.*, 2002], Ireland [*Steele-Dunne et al.*, 2008], and the Mediterranean [*Quintana Seguí et al.*, 2010].

[4] In this study, multiple bias-correction methods found in literature are applied to climate model simulated monthly *Ta* and *Pr* data in global terrestrial areas, excluding Antarctica and Greenland. The historical data are divided into two periods: the first half (1948–1972) for baseline calibration and the second half (1974–1998) for validation. Two main aspects of bias-correction are focused in the evaluation: (1) The accuracy of bias-corrected data as judged by comparisons with observations. (2) The differences in the 21st-century future simulations (2073–2097) among the bias-corrected data from using different methods, as well as of the changes of the statistics between uncorrected and bias-corrected data. For the latter purpose, we compare the changes in the mean *Ta* and *Pr*, the standard deviation (SD) of *Ta*, and the coefficient of variation (CV) of *Pr* from the baseline (1948–1972) to the projection (2073–2097) period (hereafter denoted as $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV, respectively). The variability of *Pr* is quantified by CV instead of SD since it is in general proportional to the mean of *Pr*, and this is consistent with that used by *Leander and Buishand* [2007].

[5] One of the distinctions among various bias-correction methods is related to the difference between simulated and observed data in the baseline period. The simplest method [e.g., *Graham et al.*, 2007; *Sperna Weiland et al.*, 2010] focuses on the mean difference between them. Since the correction in only the mean value is often insufficient to assess the impacts of climate change on potential hydrologic responses, some methods [e.g., *Wood et al.*, 2004; *Leander et al.*, 2008; *Piani et al.*, 2010] considering the difference in the statistical distribution in addition to mean have been proposed. However, these methods do not consider the changes in statistics between the historical and future simulations.

[6] Another important distinction among bias-correction methods is related to the difference in the changes of *Ta* and *Pr* between the historical and future simulations. *Haerter et al.* [2011] emphasized the significance of this difference and proposed an option to match the difference in $\Delta\mu_{\mathrm{ta}}$. They mentioned that $\Delta\mu_{\mathrm{ta}}$ is not identical between uncorrected and bias-corrected data if the SD of the simulated and observed data in the baseline period is different.

[7] Although not yet proposed in literature, it can be assumed that the change of simulated data from the baseline to projection period is true. For the multitude of bias-correction methods that have been proposed over last two decades [e.g., *Wood et al.*, 2004; *Leander et al.*, 2008; *Piani et al.*, 2010], the $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV are not identical between uncorrected and bias-corrected data because most of them assumed that the difference between simulated and observed data in the baseline period will not change in the future. However, the assumption that the changes in these statistical quantities are conserved between uncorrected and corrected data is not unreasonable, because most impact assessment studies [*IPCC*, 2007] were generally conducted under this critical assumption. Since in practice it is impossible to know the actual changes (of *Ta* and *Pr*) from the historical to projection period, to assume that these changes obtained from model simulations are true can be considered as a reasonable approach.

[8] One of the two main goals of this study is to evaluate existing bias-correction methods by using the data from multi-GCM simulations. Although some studies [e.g., *Christensen et al.*, 2008; *Terink et al.*, 2010; *Haerter et al.*, 2011; *Iizumi et al.*, 2011] have attempted to evaluate several correction methods, a systematic and thorough analysis on the differences among available methods have not been reported. In addition, although it has been a common practice to use multiple GCM data in climate change impact studies, the same recommendation has seldom been adapted in the evaluation of available bias-correction methods.

[9] The second goal of this study is to propose a new bias-correction method that enables the time changes of basic statistics ($\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV) to be nearly identical between uncorrected and bias-corrected data. Although several studies [e.g., *Li et al.*, 2010; *Haerter et al.*, 2011] have attempted to keep a subset of these time differences unchanged between uncorrected and corrected, none of them can conserve most of the changes in the statistics.

[10] This paper is organized into six sections. The global monthly *Ta* and *Pr* data sets used in this study are introduced in Section 2. Several bias-correction methods that will be intercompared in this study are briefly summarized in Section 3. The results of intercomparisons are presented in Sections 4, and the differences in the 95th percentile extreme values of the bias-corrected data among various bias-correction methods are presented and discussed in section 5. This is followed by a summary of the performance of each method, and the recommendations for future investigations in Section 6.

2. Data and Methods

[11] The data used in this study include the simulations from 12 GCMs in the Coupled Model Intercomparison Project (CMIP3) [*Meehl et al.*, 2007] archives of the Program for Climate Model Diagnosis and Intercomparison (PCMDI), as listed in Table 1. Although in general multiple ensemble members are available, only one ensemble member from each GCM is selected for this study. The variables to be corrected include monthly mean *Ta* and *Pr* from the 20C3M and the SRES A1B experiments.

[12] Global climate forcing data from 1948 to 2006 with the 0.5° grid resolution compiled by *Hirabayashi et al.* [2008] are taken as the observational data from which the bias-correction methods are calibrated and validated. An advantage of this global data set is that the statistical characteristics of climate variables are independent from the reanalysis data since this global forcing data were created by applying statistical methods to spatially and temporally interpolate monthly to 3-hourly observational data, including e.g., the PREC/L *Pr* data [*Chen et al.*, 2002], CRU TS 2.1 [*Mitchell and Jones*, 2005], Global Historical Climatology Network version 2 and the Climate Anomaly Monitoring System GHCN/CAMS [*Fan and van den Dool*, 2008], etc. Using this data set, *Hirabayashi et al.* [2010] was able to simulate reasonable global-scale glacier mass balance by using their glacier model.

Table 1. List of 12 GCMs Used in This Study

	CMIP3_ID	Originating Group
M1	BCCR-BCM2.0	Bjerknes Centre for Climate Research
M2	CGCM3.1(T63)	Canadian Centre for Climate Modeling & Analysis
M3	CCSM3	National Center for Atmospheric Research
M4	CNRM-CM3	Météo-France/Centre National de Recherches Météorologiques
M5	CSIRO-Mk3.5	CSIRO Atmospheric Research
M6	GFDL-CM2.1	U.S. Dept. of Commerce/NOAA/ Geophysical Fluid Dynamics Laboratory
M7	FGOALS-g1.0	LASG/Institute of Atmospheric Physics
M8	MIROC3.2(hires)	CCSR/NIES/FRCGC
M9	ECHAM5/MPI-OM	Max Planck Institute for Meteorology
M10	MRI-CGCM2.3.2	Meteorological Research Institute
M11	UKMO-HadGEM1	Hadley Centre for Climate Prediction and Research/Met Office
M12	INGV-SXG	Instituto Nazionale di Geofisica e Vulcanologia

[13] The following three time periods are defined in this study. The GCM-simulated monthly *Ta* and *Pr* data are calibrated against the corresponding observed data from 1948 to1972 (baseline period) and then validated from 1974 to1998 (validation period). Also, GCM simulations of the SRES A1B experiment are corrected using observed data during the baseline period. Since in some methods the length of data period during which the bias-correction is conducted must match that of the baseline period (25 years), the future simulation data of the SRES A1B experiment from 2073 to 2097 (projection period) are used for the intercomparison and evaluation of bias-correction methods.

3. Bias-Correction Methods

[14] In this section, five bias-correction methods found in literature briefly reviewed in Section 3.1 will be classified according to their respected characteristics with the two criteria (Section 3.2). Then, a new bias-correction method will be proposed in Section 3.3.

3.1. Previous Methods

[15] The simplest way for bias-correction is by adding (or multiplying) the $\Delta\mu_{\mathrm{ta}}$ to the observed data in the baseline period (i.e., the "delta method" [e.g., *Graham et al.*, 2007; *Sperna Weiland et al.*, 2010]):

$$x_{cor,i} = x_{o,i} + \mu_p - \mu_b \tag{1}$$

$$x_{cor,i} = x_{o,i} \times \frac{\mu_p}{\mu_b} \tag{2}$$

where $x_{cor,i}$, $x_{o,i}$ (i = 1, 2, ···, 25) denotes the bias-corrected data and observed data in the baseline period, respectively. The subscript b, p and o indicate the simulated data in the baseline period and projection period, and observed data, respectively. Previous studies often used equation (1) for *Ta* and equation (2) for *Pr*, since negative values may result from equation (1). Because only the average change (or ratio) is added to (or multiplied with) the observed data, the SD (CV) of the statistical distribution of bias-corrected data is identical to that of observed data if equation (1) (equation (2)) is adopted. In other words, the difference in SD (CV) between two time periods does not affect the result of bias-correction in equation (1) (equation (2)).

[16] For *Pr*, a slightly more advanced nonlinear bias-correction method was proposed by *Leander and Buishand* [2007], and subsequently adopted in several studies [*Leander et al.*, 2008; *van Pelt et al.*, 2009; *Hurkmans et al.*, 2010; *Terink et al.*, 2010]. Since the delta method corrects only the mean of *Pr*, the following method [*Leander et al.*, 2008] that corrects both mean and CV is considered more preferable:

$$x_{cor,i} = b\left(x_{p,i}\right)^a \tag{3}$$

where $x_{p,i}$ is the simulated data in the projection period, and a and b are the parameters obtained from calibration in the baseline period and subsequently applied to the projection period. They are determined by matching the mean and CV of simulated data with that of observed data [*Leander et al.*, 2008].

[17] Another approach to correct both mean and SD (or CV) is to apply the quantile-based mapping [*Panofsky and Brier*, 1968] as applied in several recent studies [e.g., *Wood et al.*, 2004; *Ines and Hansen*, 2006; *Lopez et al.*, 2009; *Maurer* 2007; *Maurer et al.*, 2009]. For using this method, the statistical distribution of the data must be specified. Normal distribution is often assumed for *Ta* [e.g., *Terink et al.*, 2010; *Piani et al.*, 2010; *Haerter et al.*, 2011], while the two-parameter gamma distribution is often assumed for *Pr* [e.g., *Ines and Hansen*, 2006; *Piani et al.*, 2009] as follows:

$$f(x) = x^{k-1}\frac{\exp\left(-\dfrac{x}{\theta}\right)}{\Gamma(k)\theta^k} \tag{4}$$

where k and θ are parameters taken as positive numbers. However, this approach may leave rooms for improvement. In particular, the goodness of fit is found to be not adequate [*Piani et al.*, 2010], and the parameter estimation errors are large for daily (or sub-daily) *Pr* in some regions. Thus, some other studies [e.g., *Kiem et al.*, 2008; *Piani et al.*, 2010] have proposed to use the nonparametric techniques to estimate the cumulative probability, which is able to produce a perfect transformation function between uncorrected and observed data in the baseline period.

[18] However, *Piani et al.* [2010] mentioned that the nonparametric method may also have problems in the functional robustness, because its overlarge degree of free-dom is equal to the number of parameters. A more robust method was thus proposed by *Piani et al.* [2010] in which the exact equations that will be used for bias-correction are first obtained from the data in the baseline period. This is achieved by first sorting the data by its intensity, plotting them against observed data, and then fitting by the following equations instead of performing the quantile-based mapping:

$$x_{cor,i} = a + bx_{\mathrm{p,i}} \tag{5}$$

$$x_{cor,i} = \left(a + bx_{\mathrm{p,i}}\right)\left(1 - \exp\left\{-\frac{\left(x_{p,i} + \dfrac{a}{b}\right)}{c}\right\}\right) \tag{6}$$

Table 2. Summary of the Classification of Bias-Correction Methods

	Classification by Whether Future Statistics are Included in the Equations for Bias-Correction	
Type	Constant Type (C Type)	Variable Type (V Type)
Definition	Statistics (mean, SD, or CV) in future period are NOT included in the equations of bias-correction	Statistics in the future period is included in the equations of bias-correction
Characteristics	* Statistics (Mean, SD, or CV) in future period does NOT affect the result of bias-correction. * Any length of bias-corrected data can be obtained.	* Statistics in future period affects the result of bias-correction. * Easy to adjust the change of statistics from baseline to future period between uncorrected and corrected data. * The length of the bias-corrected data needs to be the same as that in baseline period
Example of previous studies	*Wood et al.* [2004]; *Leander and Buishand* [2007]; *Piani et al.* [2010]; *Kiem et al.* [2008]; *Ines and Hansen*, [2006]	*Diaz-Nieto and Wilby* [2005]; *Lehner et al.* [2006]; *Li et al.* [2010]
	Classification by Assumptions of Statistical Distributions	
Type	Parametric Type (P Type)	Nonparametric Type (n Type)
Definition	Parametric distributions are assumed in the simulated and/or observed data	Parametric distributions are not used in the bias-correction
Characteristics	* Unrealistic value cab be produced if the distribution is not well estimated. * Statistically robust	* Extrapolation is needed for the values which is more/less than that in the baseline period. * Since the degree of freedom is equal to the number of parameters, there is a problem in the functional robustness if non-parametric distribution is assumed
Example of previous studies	*Li et al.* [2010]; *Ines and Hansen* [2006]; *Piani et al.* [2009]	*Diaz-Nieto and Wilby* [2005]; *Leander and Buishand* [2007]; *Piani et al.* [2010]; *Kiem et al.* [2008]

where a, b, and c are parameters to be obtained during baseline period. Equation (5) is used if $a < 0$ and $1/5 < b < 5$, otherwise equation (6) is used [*Piani et al.* 2010].

[19] Another method taking into account the changes of the data statistics from the baseline to projection period was proposed by *Li et al.* [2010]. Consideration of the change of the data between two periods characterizes the difference from other available methods. This method applies the quantile-based mapping for both the data between the baseline and projection period, and between uncorrected and observed data, as in the following:

$$x_{cor,i} = x_{p.i} - \left\{F_b^{-1}\left(F_{\mathrm{p}}\left(x_{p,i}\right)\right) - F_o^{-1}\left(F_{\mathrm{p}}\left(x_{p,i}\right)\right)\right\}. \tag{7}$$

[20] However, a month when $Pr = 0$ poses a significant problem for this method. *Li et al.* [2010] used a mixed gamma distribution to overcome this problem. The CDF in equation (7) is estimated by

$$F(x) = (1-p)H(x) + \mathrm{p}\mathrm{G}(x) \tag{8}$$

where p is the percentage of the months with rain; $H(x)$ is a step function of the value 0 when there is $Pr = 0$ and 1 when $Pr \neq 0$; $G(x)$ is the gamma distribution estimated from the uncorrected data of which the amount is greater than zero.

3.2. Classification of Bias-Correction Methods

[21] The above bias-correction methods can be classified into four types based on the following two major criteria: (1) Whether the statistics (mean, SD and CV) of future simulations are used; and (2) whether the estimation of the cumulative probability of correcting variables is necessary. The results of applying methods that require the use of future statistics depend on the data in the projection period, while those which do not use are not. Thus, we define the methods that do not use future statistics as the "*CONSTANT*" type (henceforth as "*C*-type"), and methods that use future statistics as the "*VARIABLE*" type ("*V*-type"). Since the delta method (equations (1) and (2)) require the mean of the variables in the projection period and the method of *Li et al.* [2010] requires the estimation of CDF in the projection period (equation (7)), both of them fall into the V-type category, while the other methods as reviewed above belong to the *C*-type. In addition, the methods which require the estimation of the cumulative probability function are described as the "*PROBABILITY*" type ("*p*-type"), while those not as the "*NON-PROBABILITY*" type ("*n*-type").

[22] Table 2 summarizes the characteristics of the proposed classification. One of the main differences between the *V*-type and *C*-type methods is related to the differences between uncorrected and corrected data in the change of *Ta* and *Pr* from the baseline to projection period ($\Delta\mu_{\mathrm{ta}}$ and $\Delta\mu_{\mathrm{pr}}$, respectively). The $\Delta\mu_{\mathrm{ta}}$ and $\Delta\mu_{\mathrm{pr}}$ of uncorrected and corrected data are in general not the same when the *C*-type methods are used. On the other hand, $\Delta\mu_{\mathrm{ta}}$ and $\Delta\mu_{\mathrm{pr}}$ of the corrected data by using the *V*-type methods can be easily matched with that of uncorrected data because the statistics of future simulations are used in the bias-correction.

[23] Another difference is related to the data length in the projection period. When the *V*-type methods are used, the length of the data to be bias-corrected needs to be the same as that in the baseline period because the statistics of future simulation data are used. In contrast, the length of bias-corrected data can be arbitrary when the *C*-type methods are

applied, since it corrects the data using only the data in the baseline period. That is, when using the C-type methods the length of bias-corrected data is allowed to exceed that in the baseline period, and the mathematic functions between uncorrected data and observed data can be uniquely determined by only using the data in the baseline period.

[24] Depending on the proposed two criteria, the above methods can be classified into the following four types: the *Vp*, *Vn*, *Cp*, and *Cn* methods. The classification proposed in this study is useful to distinguish various methods. However, both the methods of *Leander and Buishand* [2007] and *Piani et al.* [2010] are categorized as *Cn*, but the characteristics of these two methods are largely different, thus they are differentiated by using the sub-classification such as *Cn-l* and *Cn-p*, respectively.

3.3. A New Bias-Correction Method

[25] Although $\Delta\mu_{\mathrm{ta}}$ and $\Delta\mu_{\mathrm{pr}}$ are conserved in the *V*-type methods, ΔSD and ΔCV are not. Here, for the reasons stated earlier, a new method is proposed of which both $\Delta\mu_{\mathrm{ta}}$ and ΔSD for temperature (and $\Delta\mu_{\mathrm{pr}}$ and ΔCV for precipitation) are conserved. In most of previous methods, bias-correction was applied to monthly *Ta* or *Pr* data. In contrast, the proposed method first corrects the statistical parameters for each of the baseline and projection period), and then monthly *Ta* or *Pr* are corrected by using the quantile-based mapping methods (section 3.1) with the bias-corrected statistical parameters, which are mean and SD (or CV) if a normal distribution is assumed (usually for *Ta*). The procedures for correcting the statistical parameter is as follows,

$$\mu_{cor} = \mu_o + \mu_p - \mu_b \tag{9}$$

$$\sigma_{cor} = \frac{\sigma_p \sigma_o}{\sigma_b}. \tag{10}$$

[26] After the statistical parameters are corrected, monthly *Ta* is corrected as follows:

$$x_{cor,i} = F^{-1}\left(F\left(x_{p,i}; \mu_{,p}, \sigma_p\right); \mu_{cor}, \sigma_{cor}\right) \tag{11}$$

where F is the CDF of the assumed statistical distribution.

[27] For *Pr*, the process is similar to *Ta*, but an extra process is needed because of the zeroes in *Pr*. Here we apply the two gamma-distributions defined only for positive variables. Since the months when *Pr* = 0 have to be excluded before the estimation of statistical parameters, a threshold is introduced to treat the portion of uncorrected data less than it as zero. This threshold is obtained from the data in the baseline period by calculating the percentile of uncorrected data at which the corresponding observed *Pr* exceeds zero. The threshold is zero if the number of the *Pr* = 0 months in uncorrected data exceeds that in observed data.

[28] After the exclusion of *Pr* = 0 months, the mean and CV are corrected as follows:

$$\mu_{cor} = \frac{\mu_p \mu_o}{\mu_b} \tag{12}$$

$$\mathrm{CV}_{cor} = \frac{\mathrm{CV}_p \mathrm{CV}_o}{\mathrm{CV}_b}. \tag{13}$$

[29] Then, two statistical parameters k and θ are estimated from the bias-corrected mean and CV by using the method of moments. A multiplication equation is used to avoid the occurrence of negative values because the variable of interest, *Pr*, is nonnegative. After the parameters of the statistical distribution of bias-corrected data are obtained, *Pr* is estimated from the obtained distribution as follows:

$$x_{cor,i} = F^{-1}\left(F\left(x_{p,i}; k_p, \theta_p\right); k_{cor}, \theta_{cor}\right) \tag{14}$$

where F is the CDF of the two-parameter gamma distribution.

4. Comparison of Bias-Correction Methods

4.1. Metrics of Comparison

[30] The mean, 5th percentile, and 95th percentile of bias-corrected monthly *Pr* and *Ta* data obtained by using each method were compared to that of observed data. This evaluation is conducted for the baseline and validation periods separately, based on the following spatially weighted root-mean square-error (S-RMSE):

$$\text{S-RMSE} = \sqrt{\frac{1}{A}\sum_j \frac{a_j}{\sigma_{obs,j}}\left(x^*_{cor,j} - x^*_{obs,v,j}\right)^2} \tag{15}$$

where $x^*_{cor,j}$ is the mean, 5th percentile, or 95th percentile at a grid j; with the area a_j; A is the sum of a_i; $x^*_{obs,v,j}$ is the observed data in the validation period; and $\sigma_{obs,j}$ is the standard deviation of observed data at grid *i* in the baseline period. The S-RMSE measures the difference between bias-corrected data and observed data in the validation period. The weighted difference at grid *j* is averaged over the grids in the entire target domain (covering all terrestrial areas except Antarctica and Greenland). A bias-correction method is considered to be superior to another if S-RMSE is smaller. However, grids with *Pr* less than 50 mm/mo. are excluded in the calculation of S-RMSE in order to give priority to compare the areas with higher *Pr*, because the areas with low *Pr* usually have little influences on the calculated S-RMSE.

[31] In addition to the above tasks, $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV between uncorrected and corrected data are compared for each method by the following procedures: (1) The mean and SD (or CV) of both uncorrected and corrected data are estimated for each grid and in each period; (2) the $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV were calculated; and (3) the above statistics are averaged over the target region. For *Pr*, the grids with *Pr* less than 50 mm/mo. are excluded in the averaging due to the same reason described above.

4.2. Bias-Correction for Data in the 20th Century Simulations

[32] Table 3 presents the comparison of the bias-corrected, 20th-century simulation data using different methods. The values of S-RMSE in the table are averaged over 12 GCMs. As seen, the mean difference in S-RMSE among different methods is small relative to that between uncorrected and bias-corrected data, and the same tendency is also found for all the statistical indicators considered (i.e., mean, 5th percentile and 95th percentile) and for both *Ta* and *Pr*. The statistical indicators in Table 3 are originally calculated for all twelve months; since no significant differences among

Table 3. Comparisons of the Statistics of Bias-Corrected Data Derived by Using Different Correction Methods in the 20th-Century Simulations[a]

	JAN			JUL		
	ave	5th	95th	ave	5th	95th
			S-RMSE			
Not corrected	2.54	2.70	2.57	3.42	3.42	3.62
Vn	0.50	0.62	0.58	0.57	0.71	0.64
Cp	0.47	0.63	0.67	0.53	0.66	0.73
Vp	0.50	0.69	0.66	0.57	0.77	0.77
Proposed	0.50	0.68	0.66	0.57	0.75	0.73
			SD			
Not corrected	0.312	0.293	0.327	0.447	0.248	0.686
Vn	0.007	0.008	0.004	0.011	0.016	0.003
Cp	0.009	0.007	0.006	0.007	0.006	0.005
Vp	0.007	0.009	0.004	0.011	0.008	0.010
Proposed	0.007	0.007	0.005	0.011	0.010	0.008
			S-RMSE			
Not corrected	1.15	1.07	1.76	1.22	1.15	1.83
Vn	0.38	0.38	0.88	0.37	0.36	0.77
Cp	0.39	0.41	0.97	0.36	0.38	0.83
Vp	0.38	0.46	0.90	0.36	0.41	0.77
Cn-l	0.40	0.39	1.03	0.36	0.35	0.87
Cn-p	0.38	0.56	0.93	0.35	0.45	0.81
Proposed	0.37	0.38	0.94	0.35	0.35	0.82
			SD			
Not corrected	0.22	0.14	0.19	0.19	0.14	0.14
Vn	0.03	0.03	0.06	0.04	0.04	0.06
Cp	0.03	0.03	0.07	0.02	0.04	0.05
Vp	0.03	0.07	0.05	0.02	0.05	0.04
Cn-l	0.03	0.03	0.08	0.02	0.02	0.05
Cn-p	0.02	0.06	0.04	0.02	0.03	0.04
Proposed	0.03	0.03	0.05	0.03	0.03	0.04

[a]Statistics: i.e., *S-RMSE* as defined in equation (15) and the standard deviation of *S-RMSE* among 12 GCMs. The upper table is for temperature, and the lower one for precipitation. The columns of *ave*, 5th and 95th denote the mean, 5th and 95th percentile of the 25-year data in each period.

the months are found, the comparison is presented only for January and July, from which the following findings can be summarized.

[33] For *Ta*, (1) S-RMSE of the *Vn* method for both the 5th and 95th percentiles are smaller than those of other correction methods; (2) S-RMSE of the *Cp* method for the mean is smaller than other methods; (3) S-RMSE of the proposed new method is slightly smaller than the *Vp* method, although they are nearly identical. For *Pr*, (1) S-RMSE of the *Vn* method for the 95th percentile is generally smaller than those of other methods; (2) S-RMSEs of the *Cp* and *Cn-l* methods for the 95th-percentile are higher than other methods; (3) S-RMSE of the *Vp* method is intermediate; (4) S-RMSE of the *Cn-p* method for the 5th-percentile is higher than other methods; (5) S-RMSE of the proposed method for the mean is slightly lower than other methods.

[34] The standard deviation (SD) of the S-RMSE among uncorrected data is also calculated in Table 3, which indicates the spread of bias-corrected data among different methods. Although the SD of S-RMSE is $\sim$10% for both uncorrected *Ta* and *Pr*, it is reduced to $\sim$0.1% for *Ta* and $\sim$1% for *Pr* after bias-correction. That is, the differences are also small among the bias-corrected data from different methods. Also, no significant differences in S-RMSE can be found from the comparisons among individual months.

[35] Characteristics of the bias-correction methods are not apparent when the change between two time periods is not large. The *Vn* method, of which the SD (for *Ta*) and CV (for *Pr*) of the corrected data are identical to that of observed data (because it only adjusts the mean), shows a relatively small S-RMSE because the changes of SD (or CV) between the baseline and validation period are not large. The difference of bias-correction is expected to become more significant in the future projection period because the changes of *Ta* and *Pr* from the baseline to future period are generally larger. Although the comparison in this section only focus on the relative change of *Ta* and *Pr*, the efficiency of each method and the lack of large differences among methods are indicated by this comparison.

4.3. Bias-Correction for the Data in 21st Century Simulations

[36] The differences in the statistical indicators ($\Delta\mu_{ta}$, $\Delta\mu_{pr}$, ΔSD, and ΔCV) between corrected and uncorrected data are compared among all the bias-correction methods used, and the results are presented in Figures 1 (for temperature) and 2 (for precipitation), respectively. These two figures show the absolute difference ($\Delta\mu_{ta}$) or the ratio of difference ($\Delta\mu_{pr}$, ΔSD, and ΔCV) between corrected and uncorrected data, with the latter defined as the absolute difference divided by the uncorrected mean or SD (or CV) in the projection period, respectively. These statistical indicators are calculated at each grid for each of the 12 GCMs, and then averaged over the entire target regions. The error bars in the figure represent the SD of the statistical indicators over all the grids, and the averages over 12 GCMs are also shown in the first column of the figures. Only the data in January and July are presented here, since no marked differences in tendency can be found among twelve months.

[37] Results show that for both *Ta* and *Pr* the characteristics of the differences are not identical among different methods. From Figure 1 (temperature), it can be seen that (1) for the *Vn* and *Vp* methods, $\Delta\mu_{ta}$ is zero, but ΔSD is about 10–30%; (2) for the Cp method, $\Delta\mu_{ta}$ is more than 1 *K*, but ΔSD is almost zero; (3) for the proposed new method, both $\Delta\mu_{ta}$ and ΔSD are almost zero. From Figure 2 (precipitation), it can be seen that (1) for the *Vn* method, both $\Delta\mu_{pr}$ and ΔCV are higher than other method; (2) for the *Cn-p* method, $\Delta\mu_{pr}$ is less than 10% and ΔCV is about 10–20%: the former of which is smaller than *Cn-l* method, but the latter is generally larger than *Cn-l* method; (3) for the *Cn-l* method, the result is similar to that of the *Cn-p* method, but the result of ΔCV over target region is opposite to the result of the *Cn-p* method; (4) for the *Vp* method, $\Delta\mu_{pr}$ is nearly 0, but ΔCV is higher than other methods except for the *Vn* method; (5) for the *Cp* method, $\Delta\mu_{pr}$ is higher than other methods except for the *Cn-l* method, but ΔCV is smaller than any other methods. However, the spread (SD) over the target regions is comparatively high; (6) for the proposed method, both $\Delta\mu_{pr}$ and ΔCV are lower than all other methods, and the spread over the target region is also small. The identified characteristics as summarized above are the same in other months than January and July, hence their plots are not shown here. In summary, the statistical characteristics of bias-corrected data in the projection period

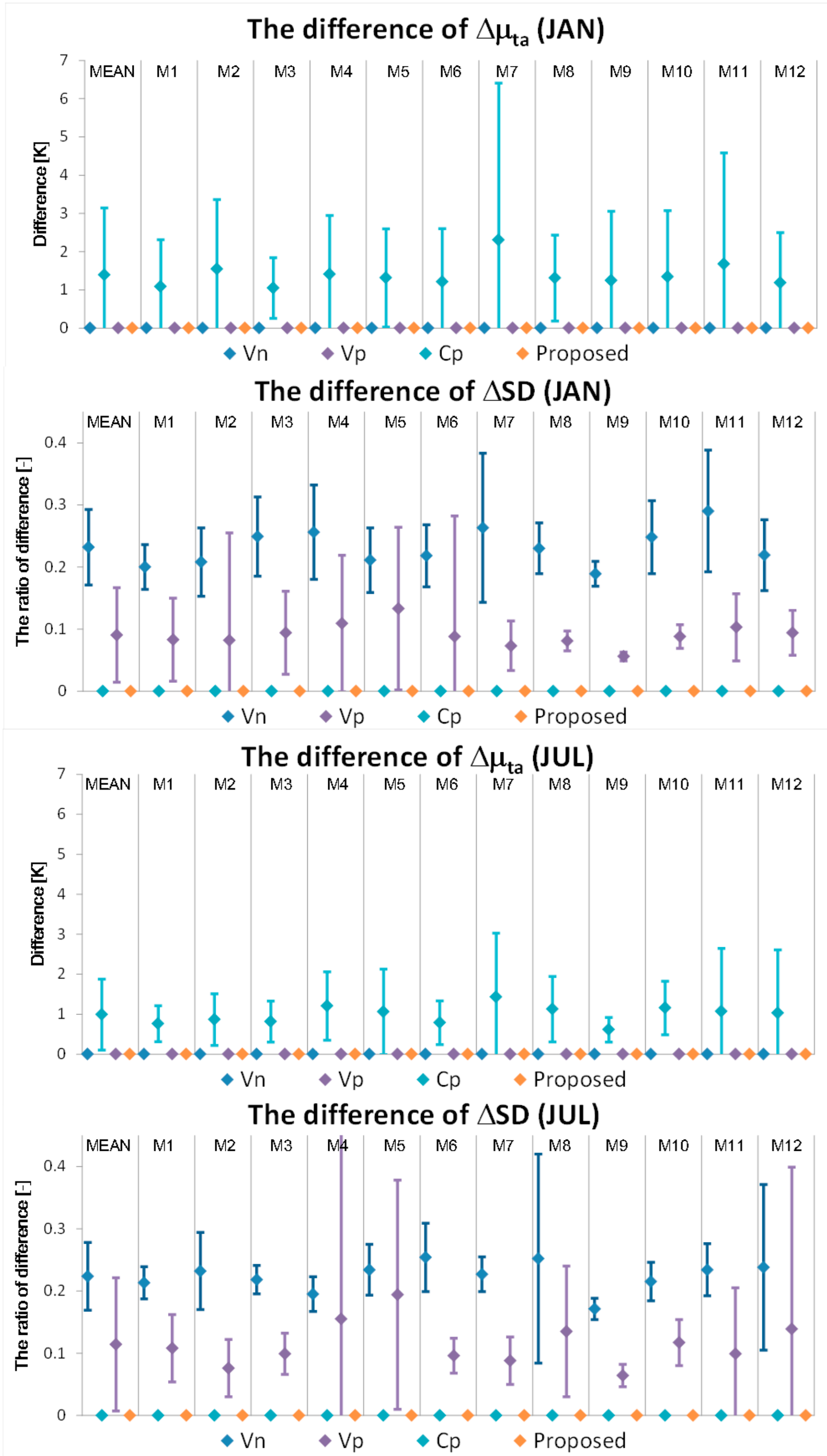

Figure 1. Differences in $\Delta\mu_{ta}$ and ΔSD (of temperature) among methods for each of 12 GCMs used (See the text for the definition of $\Delta\mu_{ta}$ and ΔSD). The first column (MEAN) is the difference averaged over 12 GCMs. The error bar indicates the standard deviation of $\Delta\mu_{ta}$ and ΔSD over all the grids in target region.

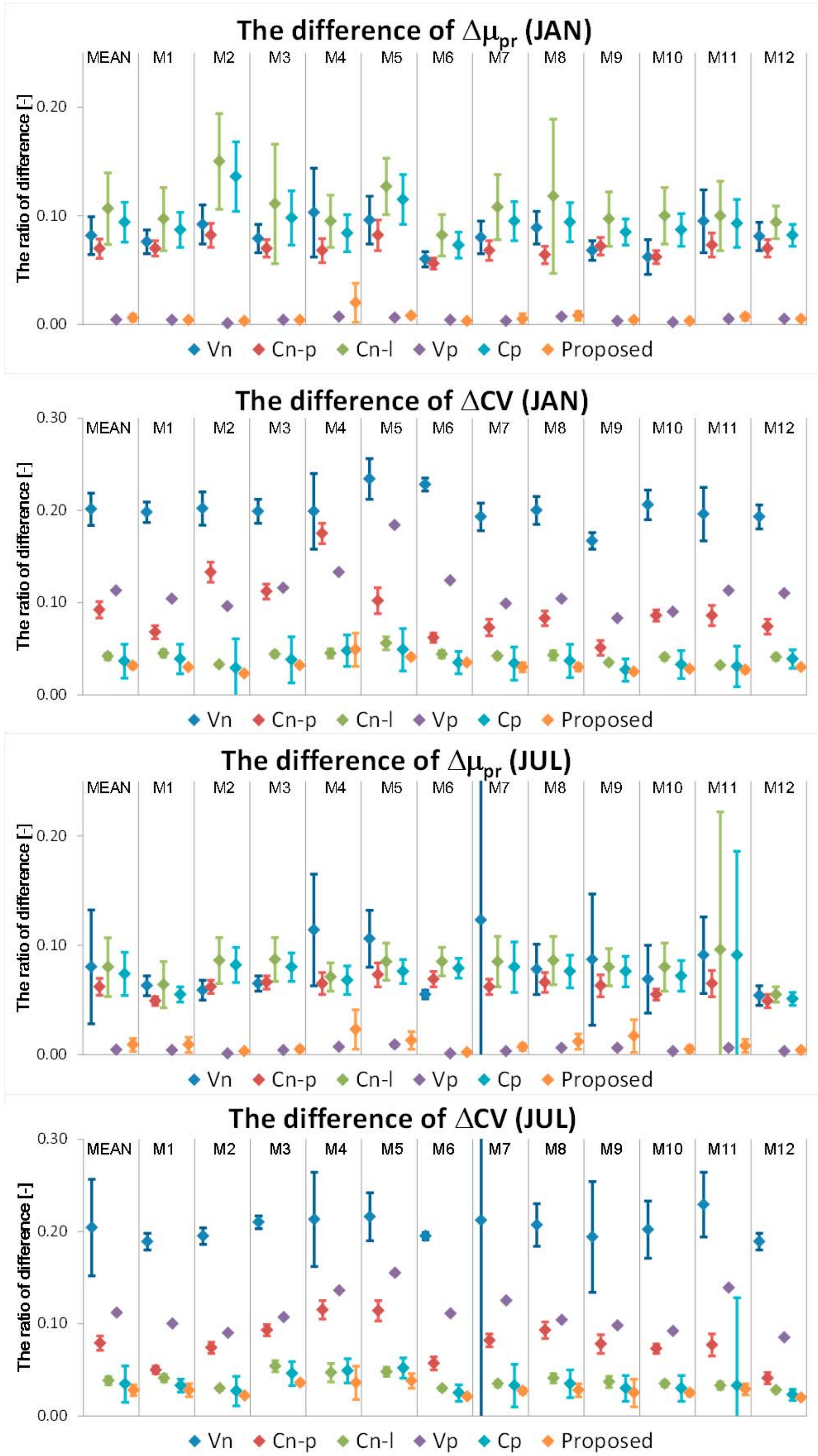

Figure 2. Same as Figure 1, but for precipitation (The difference of $\Delta\mu_{pr}$ is divided by the uncorrected μ_p in this figure).

Table 4. Summary of the Characteristics of Bias-Correction Methods[a]

Method	Vn	Cn-p	Cn-l	Vp	Cp	Proposed
The difference from observational data in validation period (*Ta*)	Several percent smaller than other methods for 5th and 95th percentile	N/A	N/A	Medium	Several percent smaller than other methods for average of 25 years	Medium
The difference from observational data in validation period (*Pr*)	Several percent smaller than other methods for 95th percentile	Several percent larger than other methods for 5th percentile	Several percent larger than other methods for 95th percentile	Several percent larger than other methods for 5th percentile	Medium	Medium
The difference of the $\Delta\mu_{ta}$ between uncorrected and corrected data	Almost zero	N/A	N/A	Almost zero	Larger than 1 [K] for most GCMs	Almost zero
The difference of the ΔSD between uncorrected and corrected data	Larger than 20% for most GCMs	N/A	N/A	Larger than 5% for most GCMs	Almost zero	Almost zero
The difference of the $\Delta\mu_{pr}$ between uncorrected and corrected data	Larger than 5%, and significantly larger than other methods for some GCMs	Larger than 5%	Larger than 5%, and SD over target region is larger than other methods in Jan.	Almost zero	Larger than 5%	Almost zero
The difference of the ΔCV between uncorrected and corrected data	Larger than other methods	Medium	Smaller than 5%	Larger than other methods, except for Vn method	Smaller than 5%, but SD over target region is larger than other methods except for the Vn method	Generally smaller than other methods

[a] *Ta*: temperature; *Pr*: precipitation; SD: standard deviation; CV: the coefficient of variation.

vary significantly depending upon the used bias-correction method, although they do not differ much in the validation period.

[38] Compared to other bias-correction methods, $\Delta\mu_{ta}$, $\Delta\mu_{pr}$, ΔSD and ΔCV of the bias-corrected data by using the proposed new method are closer to that of uncorrected data. However, note that there still remain small differences of ΔCV between corrected and uncorrected data. This is due to the difficulty in matching ΔCV of corrected data with that of uncorrected data as a result of placing a lower limit for the *Pr* = 0 months (see the discussion in section 3.3).

4.4. Summary of Method Characteristics

[39] Table 4 summarizes the major characteristics of different methods, which are discussed according to their differences in $\Delta\mu_{ta}$, $\Delta\mu_{pr}$, ΔSD and ΔCV of the bias-corrected data. For example, the proposed new method has the characteristic that the $\Delta\mu_{ta}$ and ΔSD of corrected data are closer to that of uncorrected data than other methods; and the *Vn* method has the characteristic that the ΔSD and ΔCV of corrected data are not identical to that of uncorrected data. The differences in the corrected future projection data are not small due to the differences in $\Delta\mu_{ta}$, $\Delta\mu_{pr}$, ΔSD and ΔCV caused by the method differences. However, the bias-corrected data in the validation period are in general not

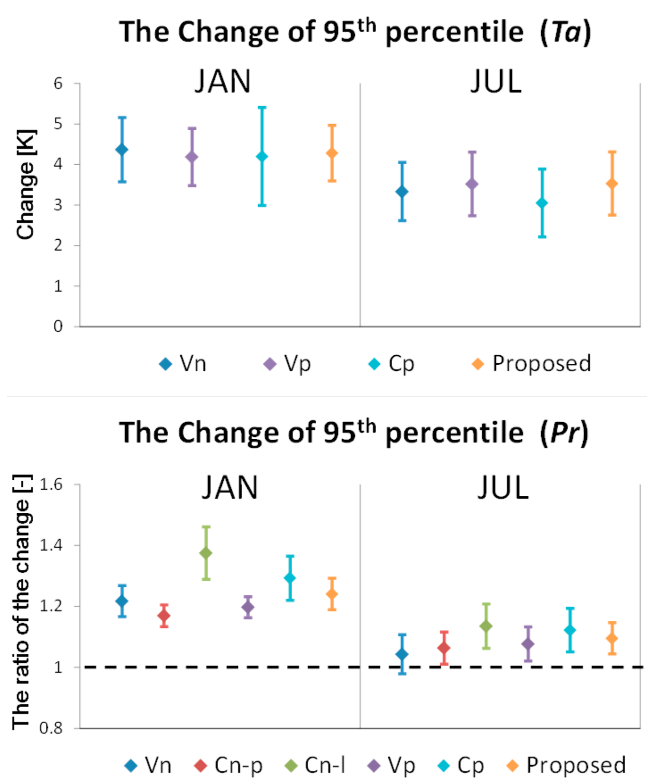

Figure 3. Comparisons of the 12-GCM averaged changes in the 95th percentile of precipitation from the baseline period (based on observed data) to project period (based on corrected data) among different methods. The change in the 95th percentile of corrected precipitation data is divided by the 95th percentile of observed precipitation data in the baseline period. The error bar shows the standard deviation among 12 GCMs.

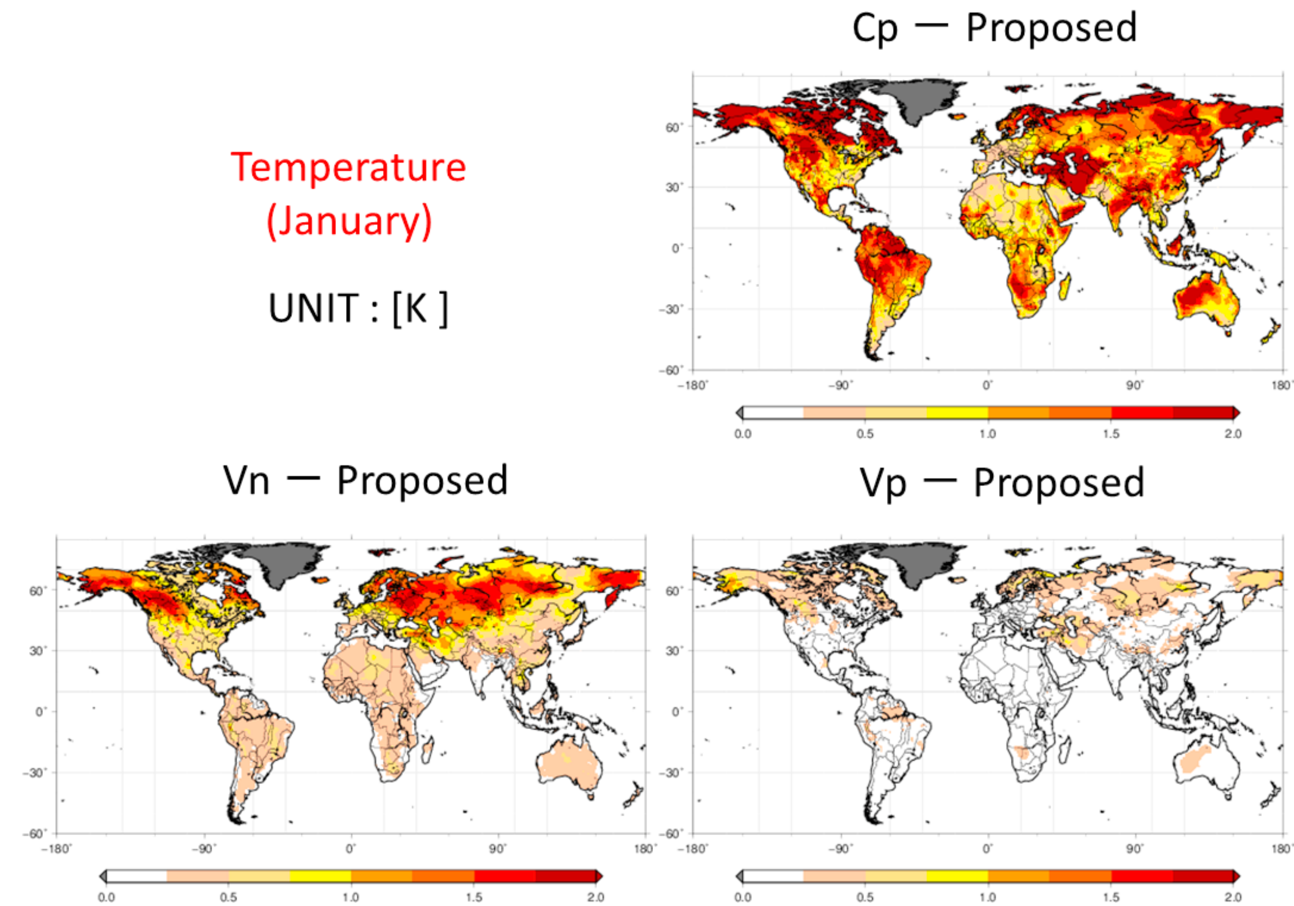

Figure 4. Differences averaged over 12 GCMs in the 95th percentile of the January temperature between the new method and other bias-correction methods. Gray areas are the grids excluded from the analysis. Difference is within 0.25 K for those grids without coloring.

significantly different among methods. Notice that the present study only focuses on the bias-correction of global-scale monthly *Ta* or *Pr* data. Different characteristics may be identified if similar intercomparison is conducted with respect to daily data and on the smaller continental or basin scales.

5. Differences of Bias-Corrected Data in 21st Century Simulations

[40] Since the extreme values of temperature and precipitation are often the focus in various impact assessment studies, it is important to investigate the difference in the extremes of bias-corrected data caused by using different correction methods. For this purpose, the globally average changes in the 95th percentiles of *Ta* and *Pr* from the baseline period (based on the observed data) to the projection period (based on the bias-corrected data) are calculated and averaged over 12 GCMSs. Figure 3 compares the differences of these changes among different methods. The change of the 95th percentile of corrected precipitation is divided by the 95th percentile of observed precipitation in the baseline period, which means that the 95th percentile precipitation is increased if the ratio plotted in Figure 3 is over 1. Note that due to the averaging over 12 GCMs, the difference in any specific GCM can be much larger than that plotted in the figure. For *Ta*, the change in the 95th percentile is similar between the *Vp* and new method, but the difference is more than 0.5 k between the *Cp* and new method. For *Pr*, the difference is more apparent. The difference obtained by using the *Cn-l* and *Cp* method are larger than that by the new method, and that by the *Vn*, *Vp* and *Cn-p* methods are lower than the new method. The range of the difference ratio in July obtained by using the *Vn* method can be less than 1, which implies that the 95th percentile of monthly *Pr* will decrease slightly in future.

[41] Figure 4 (temperature) and Figure 5 (precipitation) compare the differences in the 95th percentiles of the corrected data between the new method and other bias-correction methods. The global spatial patterns in the figures show either the absolute difference or the absolute difference divided by observed 95th percentile averaged over 12 GCMs. For *Ta*, the difference between the new method and *Cp* method is larger than 1*K* in most global regions. The difference between the new and *Vn* method is smaller, but more than 0.5 *K* in many regions, whereas the difference between the new and *Vp* method is the smallest. For *Pr*, the difference between the new and *Vn* method is

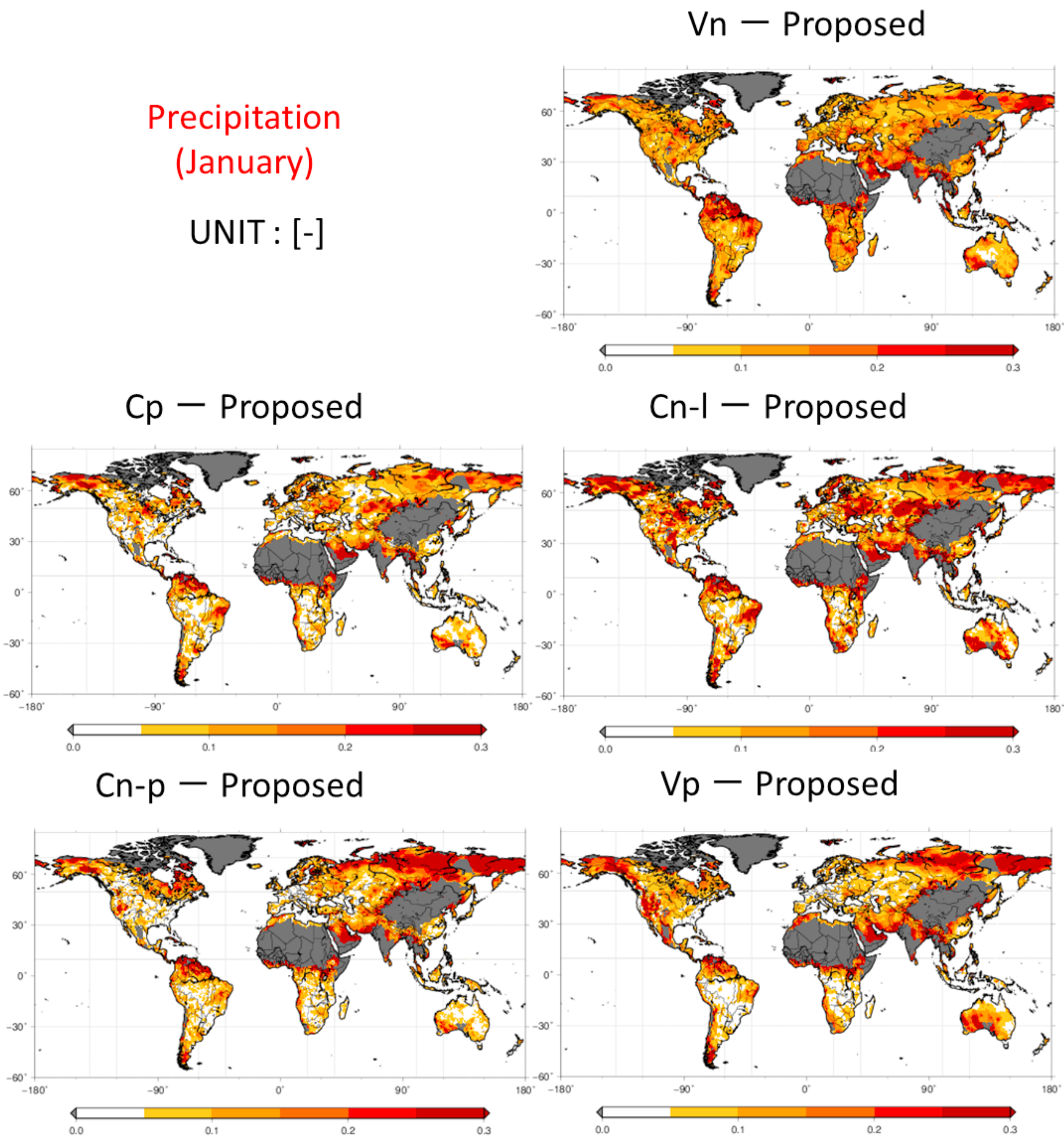

Figure 5. Same as Figure 4, but for precipitation. Each absolute difference is divided by observed 95th percentile precipitation in the baseline period. Gray areas are grids excluded from analysis. Grids for which the mean of observed data in baseline period is less than 50 mm/mo. are excluded from analysis, because of difficulties in estimating variance in dry regions.

more than 5% in most regions, but the regions with the differences >20% is less than the *Cn-l*, *Cn-p* and *Vp* method. The difference between the new and *Cp* method is smaller than other methods. The differences between the new and the *Cn-l*, *Cn-p* and *Vp* method are more than 20% in some regions, but the areas where the difference is smaller than 10% is larger than the *Vn* method. From these comparisons, it can be concluded that the choice of bias-correction methods is crucial for the impact assessment studies of climate change.

6. Summary and Conclusion

[42] Bias-correction methods for monthly mean *Ta* and total *Pr* data are intercompared in this study based on 12 GCM simulations in the Coupled Model Intercomparison

Project phase 3 (CMIP3) [*Meehl et al.*, 2007] archives. Although the CMIP phase5 (CMIP5) data set which becomes available recently is not used in this study, the presented intercomparison methodologies and the findings obtained as summarized below can also be applied and still valid to the CMIP5 data set.

[43] Before the comparisons, previous bias-correction methods are classified based on the two major criteria proposed in this study, and a new method is proposed which conserves the $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV (i.e., the changes in mean *Ta* and *Pr*, the standard deviation of *Ta*, and the coefficient of variation of *Pr*, respectively, from the baseline period 1948–1972 to the projection period 2073–2097) between uncorrected and bias-corrected data. Two types of comparisons are made to evaluate the characteristics of different methods. First, the bias-corrected data are compared with the observed data during the validation period (1974–1998), and second, the characteristics of $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV are compared among different methods.

[44] Results of the first comparison reveal that the differences between bias-corrected and observed data are not large compared to that between bias-corrected and uncorrected data. The changes from the baseline to validation period are small so that the differences among bias-corrected data are not apparent. The second comparison indicates that the $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV of bias-corrected data vary with the methods used. The differences in $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV between uncorrected and corrected data are considerably reduced by using the proposed new correction method. The characteristics of $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV of bias-corrected data can be one of the indicators to evaluate bias-correction methods. The proposed new method is useful to correct the bias under the assumption that $\Delta\mu_{\mathrm{ta}}$, $\Delta\mu_{\mathrm{pr}}$, ΔSD and ΔCV are identical between uncorrected model simulation data and bias-corrected data.

[45] The differences in the 95th percentile of corrected data are also compared, and the result indicates the large sensitivity to the choice of correction methods. Based on this finding, the use of multiple bias-correction methods and evaluation of the associated sensitivity in the specific target regions are recommended as an important future study.

[46] One possible way to take into account the large sensitivity of the differences among bias-correction methods is to apply the multiple bias-correction methods and choose one ensemble from each bias-corrected data. However, it should be kept in mind that the difference in the characteristics of each method when the final ensemble of bias-corrected data is formed from each bias-corrected data, because the weighted result can be biased if a number of similar methods are used in the final ensemble. The issues focused in this study are therefore useful to evaluate the similarity of the bias-correction methods and the appropriate weights when forming the final ensemble.

[47] Generally, the use of multiple GCM simulations is recommended to assess the impacts of climate change. Therefore, the efficiency of bias-correction methods can also be checked by using multiple GCM simulations as in this study. While much validation work remains to be done, we believe our study has strong merits for evaluating and optimizing the bias-correction methods used in the impact assessment studies.

[48] **Acknowledgments.** This work was supported by JST, CREST; Funding Program for Next Generation World-Leading Researchers (NEXT Program) from the Cabinet Office, Government of Japan; KAKENHI, Grant-in-Aid for Scientific Research (S) from Japan Society for the Promotion of Science (JSPS); and Environment Research and Technology Development Fund (S-10) from the Ministry of the Environment, Japan.

References

Arnell, N. W. (2004), Climate change and global water resources: SRES emissions and socio-economic scenarios, *Global Environ. Change*, *14*(1), 31–52, doi:10.1016/j.gloenvcha.2003.10.006.

Chen, M., P. Xie, and J. E. Janowiak (2002), Global land precipitation: A 50-yr monthly analysis based on gauge observation, *J. Hydrometeorol.*, *3*, 249–266, doi:10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2.

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, doi:10.1029/2008GL035694.

Dettinger, M. D., D. R. Cayan, M. K. Meyer, and A. E. Jeton (2004), Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American river basins, Sierra Nevada, California, 1900–2099, *Clim. Change*, *62*(1–3), 283–317, doi:10.1023/B:CLIM.0000013683.13346.4f.

Diaz-Nieto, J., and R. L. Wilby (2005), A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom, *Clim. Change*, *69*(2–3), 245–268, doi:10.1007/s10584-005-1157-6.

Fan, Y., and H. van den Dool (2008), A global monthly land surface air temperature analysis for 1948-present, *J. Geophys. Res.*, *113*, D01103, doi:10.1029/2007JD008470.

Graham, L. P., J. Andréasson, and B. Carlsson (2007), Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods—A case study on the Lule River basin, *Clim. Change*, *81*(S1), 293–307, doi:10.1007/s10584-006-9215-2.

Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani (2011), Climate model bias-correction and the role of timescales, *Hydrol. Earth Syst. Sci.*, *15*(3), 1065–1079, doi:10.5194/hess-15-1065-2011.

Hay, L. E., M. P. Clark, R. L. Wilby, W. J. Gutowski, G. H. Leavesley, Z. Pan, and R. W. Arritt (2002), Use of regional climate model output for hydrologic simulations, *J. Hydrometeorol.*, *3*(5), 571–590, doi:10.1175/1525-7541(2002)003<0571:UORCMO>2.0.CO;2.

Hirabayashi, Y., S. Kanae, K. Motoya, K. Masuda, and P. Döll (2008), A 59-year (1948–2006), global near-surface meteorological data set for land surface models. Part I: Development of daily forcing and assessment of precipitation intensity, *Hydrol. Res. Lett.*, *2*, 36–40, doi:10.3178/HRL.2.36.

Hirabayashi, Y., P. Döll, and S. Kanae (2010), Global-scale modeling of glacier mass balances for water resources assessments: Glacier mass changes between 1948 and 2006, *J. Hydrol.*, *390*, 245–256, doi:10.1016/j.jhydrol.2010.07.001

Hurkmans, R., W. Terink, U. Remko, and P. Torfs (2010), Changes in streamflow dynamics in the Rhine Basin under three high-resolution regional climate scenarios, *J. Clim.*, *23*(3), 679–699, doi:10.1175/2009JCLI3066.1.

Iizumi, T., M. Nishimori, K. Dairaku, S. A. Adachi, and M. Yokozawa (2011), Evaluation and intercomparison of downscaled daily precipitation indices over Japan in present-day climate: Strengths and weaknesses of dynamical and bias-correction-type statistical downscaling methods, *J. Geophys. Res.*, *116*, D01111, doi:10.1029/2010JD014513.

Ines, A. V. M., and J. W. Hansen (2006), Bias-correction of daily GCM rainfall for crop simulation studies, *Agric. For. Meteorol.*, *138*(1–4), 44–53, doi:16/j.agrformet.2006.03.009.

Intergovernmental Panel on Climate Change (IPCC) (2007), *IPCC Fourth Assessment Report: Climate Change 2007 Working Group II Report "Impacts, Adaptation and Vulnerability,"* Cambridge Univ. Press, Cambridge, U. K.

Kiem, A. S., H. Ishidaira, H. P. Hapuarachchi, M. C. Zhou, Y. Hirabayashi, and K. Takeuchi (2008), Future hydroclimatology of the Mekong River basin simulated using the high-resolution Japan Meteorological Agency (JMA) AGCM, *Hydrol. Processes*, *22*(9), 1382–1394, doi:10.1002/hyp.6947.

Leander, R., and T. A. Buishand (2007), Resampling of regional climate model output for the simulation of extreme river flows, *J. Hydrol.*, *332*(3–4), 487–496, doi:16/j.jhydrol.2006.08.006.

Leander, R., T. Buishand, B. Vandenhurk, and M. Dewit (2008), Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, *J. Hydrol.*, *351*(3–4), 331–343, doi:10.1016/j.jhydrol.2007.12.020.

Lehner, B., P. Döll, J. Alcamo, T. Henrichs, and F. Kaspar (2006), Estimating the impact of global change on flood and drought risks in Europe: A continental, integrated analysis, *Clim. Change*, *75*(3), 273–299, doi:10.1007/s10584-006-6338-4.

Li, H., J. Sheffield, and E. F. Wood (2010), Bias-correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, *J. Geophys. Res.*, *115*, D10101, doi:10.1029/2009JD012882.

Lopez, A., F. Fung, M. New, G. Watts, A. Weston, and R. L. Wilby (2009), From climate model ensembles to climate change impacts and adaptation: A case study of water resource management in the southwest of England, *Water Resour. Res.*, *45*, W08419, doi:10.1029/2008WR007499.

Maurer, E. P. (2007), Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios, *Clim. Change*, *82*(3–4), 309–325, doi:10.1007/s10584-006-9180-9.

Maurer, E. P., J. C. Adam, and A. W. Wood (2009), Climate model based consensus on the hydrologic impacts of climate change to the Rio Lempa basin of Central America, *Hydrol. Earth Syst. Sci.*, *13*(2), 183–194, doi:10.5194/hess-13-183-2009.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor (2007), The WCRP CMIP3 multi-model dataset: A new era in climate change research, *Bull. Am. Meteorol. Soc.*, *88*, 1383–1394, doi:10.1175/BAMS-88-9-1383.

Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*(6), 693–712, doi:10.1002/joc.1181.

Oki, T., and S. Kanae (2006), Global hydrological cycles and world water resources, *Science*, *313*(5790), 1068–1072, doi:10.1126/science.1128845.

Panofsky, H. A., and G. W. Brier (1968), *Some Applications of Statistics to Meteorology*, Pa. State Coll., University Park.

Piani, C., J. O. Haerter, and E. Coppola (2009), Statistical bias-correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, *99*(1–2), 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 (2010), Statistical bias-correction of global simulated daily precipitation and temperature for the application of hydrological models, *J. Hydrol.*, *395*(3–4), 199–215, doi:16/j.jhydrol.2010.10.024.

Quintana Seguí, P., A. Ribes, E. Martin, F. Habets, and J. Boé (2010), Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins, *J. Hydrol.*, *383*(1–2), 111–124, doi:10.1016/j.jhydrol.2009.09.050.

Seneviratne, S. I., et al. (2012), Changes in climate extremes and their impacts on the natural physical environment, in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*, edited by C. B. Field et al., pp. 109–230, Cambridge Univ. Press, Cambridge, U. K.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk, and M. F. P. Bierkens (2010), The ability of a GCM-forced hydrological model to reproduce global discharge variability, *Hydrol. Earth Syst. Sci.*, *14*(8), 1595–1621, doi:10.5194/hess-14-1595-2010.

Steele-Dunne, S., P. Lynch, R. Mcgrath, T. Semmler, S. Wang, J. Hanafin, and P. Nolan (2008), The impacts of climate change on hydrology in Ireland, *J. Hydrol.*, *356*(1–2), 28–45, doi:10.1016/j.jhydrol.2008.03.025.

Takara, K., S. Kim, Y. Tachikawa, and E. Nakakita (2009), Assessing climate change impact on water resources in the Tone River Basin, Japan, using super-high-resolution atmospheric model output, *J. Disaster Res.*, *4*, 12–23.

Terink, W., R. T. W. L. Hurkmans, P. J. J. F. Torfs, and R. Uijlenhoet (2010), Evaluation of a bias-correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin, *Hydrol. Earth Syst. Sci. Discuss.*, *7*(1), 221–267, doi:10.5194/hessd-7-221-2010.

van Pelt, S. C., P. Kabat, H. W. ter Maat, B. J. J. M. van den Hurk, and A. H. Weerts (2009), Discharge simulations performed with a hydrological model using bias corrected regional climate model input, *Hydrol. Earth Syst. Sci.*, *13*(12), 2387–2397, doi:10.5194/hess-13-2387-2009.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Clim. Change*, *62*(1–3), 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e.