

# Multivariate quantile mapping bias correction: an $N$ -dimensional probability density function transform for climate model simulations of multiple variables

Alex J. Cannon<sup>1</sup>

Received: 30 August 2016 / Accepted: 7 February 2017 / Published online: 25 March 2017  
© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** Most bias correction algorithms used in climatology, for example quantile mapping, are applied to univariate time series. They neglect the dependence between different variables. Those that are multivariate often correct only limited measures of joint dependence, such as Pearson or Spearman rank correlation. Here, an image processing technique designed to transfer colour information from one image to another—the  $N$ -dimensional probability density function transform—is adapted for use as a multivariate bias correction algorithm (MBCn) for climate model projections/predictions of multiple climate variables. MBCn is a multivariate generalization of quantile mapping that transfers all aspects of an observed continuous multivariate distribution to the corresponding multivariate distribution of variables from a climate model. When applied to climate model projections, changes in quantiles of each variable between the historical and projection period are also preserved. The MBCn algorithm is demonstrated on three case studies. First, the method is applied to an image processing example with characteristics that mimic a climate projection problem. Second, MBCn is used to correct a suite of 3-hourly surface meteorological variables from the Canadian Centre for Climate Modelling and Analysis Regional Climate Model (CanRCM4) across a North American domain. Components of the Canadian Forest Fire Weather

Index (FWI) System, a complicated set of multivariate indices that characterizes the risk of wildfire, are then calculated and verified against observed values. Third, MBCn is used to correct biases in the spatial dependence structure of CanRCM4 precipitation fields. Results are compared against a univariate quantile mapping algorithm, which neglects the dependence between variables, and two multivariate bias correction algorithms, each of which corrects a different form of inter-variable correlation structure. MBCn outperforms these alternatives, often by a large margin, particularly for annual maxima of the FWI distribution and spatiotemporal autocorrelation of precipitation fields.

**Keywords** Quantile mapping · Multivariate · Bias correction · Post-processing · Model output statistics · Climate model · Fire weather · Precipitation

## 1 Introduction

Planning for long-term climate change relies on plausible projections of the future climate. Similarly, climate-sensitive decisions on shorter time horizons rely on accurate seasonal-interannual and decadal climate forecasts. Global and regional climate models, which are based on our physical understanding of the climate system, therefore play a key role in climate impacts and adaptation and climate prediction studies. However, despite continued improvements in the representation of physical processes, systematic errors remain in climate models.

For practical reasons, users often find it necessary to remove climate model biases before outputs are incorporated into their particular applications. Methods used to post-process climate model outputs may be based on either perfect prognosis or model output statistics (MOS)

**Electronic supplementary material** The online version of this article (doi:10.1007/s00382-017-3580-6) contains supplementary material, which is available to authorized users.

✉ Alex J. Cannon  
alex.cannon@canada.ca

<sup>1</sup> Climate Research Division, Environment and Climate Change Canada, PO Box 1700 STN CSC, Victoria, BC V8W 2Y2, Canada

approaches (Maraun et al. 2010). In the former, biases are removed via a statistical model that accounts for synchronous relationships between a target variable of interest in a reference dataset and one or more observed variables that can be simulated by the climate model. In the latter, relationships (either synchronous or asynchronous) are drawn directly between the target variable and simulated climate model variables. In a climate modelling context, where free-running model simulations and observations are not synchronized in time, MOS techniques for bias correction are typically asynchronous, i.e., between distributional statistics of a variable such as the mean, variance, or quantiles. Biases are taken to refer specifically to systematic differences in such distributional properties between model simulated outputs and those estimated from the reference dataset.

Considerable effort has been expended developing these types of bias correction algorithms (Michelangeli et al. 2009; Li et al. 2010; Hempel et al. 2013), evaluating their performance (Piani et al. 2010; Gudmundsson et al. 2012; Chen et al. 2013), and determining their limitations (Ehret et al. 2012; Eden et al. 2012; Maraun 2013; Maraun and Widmann 2015; Chen et al. 2015). A recent critical review is offered by Maraun (2016). One of the most popular asynchronous bias correction methods in climatology is quantile mapping, a univariate technique that maps quantiles of a source distribution to quantiles of a target distribution. Quantile mapping (and most other bias correction methods) have typically been applied to individual variables in turn, neglecting the dependence that exists between variables (Wilcke et al. 2013). For example, if a climate model has a warm bias in high quantiles of surface temperature and a wet bias in low quantiles of precipitation, these biases would be corrected separately. Because model biases in inter-variable relationships are ignored by univariate techniques, biases in dependence structure that remain following univariate bias correction can affect subsequent analyses that make use of multiple variables (Rocheta et al. 2014). This includes, for instance, hydrological model simulations and calculations of atmospheric moisture fluxes, multivariate drought indices, and fire weather indices.

As an alternative to univariate methods, multivariate bias correction algorithms have been introduced by Bürger et al. (2011), Vrac and Friederichs (2015), Mehrotra and Sharma (2016), and Cannon (2016), among others. While these methods correct biases in multiple variables simultaneously, they either take into account only a limited measure of the full multivariate dependence structure, for example as represented by the Pearson correlation (Bürger et al. 2011; Mehrotra and Sharma 2016; Cannon 2016) or Spearman rank correlation (Cannon 2016), or they make strong stationarity assumptions about the temporal sequencing of the climate model variables (e.g., by simply replicating

observed historical rank ordering as in the empirical copulabias correction (EC-BC) method by Vrac and Friederichs 2015). In contrast to univariate techniques, many multivariate techniques are iterative, for example repeatedly applying univariate quantile mapping and multivariate transformations (Cannon 2016)—and thus the question of convergence arises. Theoretical proofs of convergence may not be available. Arguably, a direct multivariate extension of quantile mapping would map one multivariate distribution to another in its entirety, with proven convergence properties, while keeping as much of the underlying climate model's temporal sequencing intact.

Is it possible to transfer all aspects of one multivariate distribution to another in this way? In the field of image processing and computer vision, Pitié et al. (2005, 2007) developed a method, which they refer to as the *N*-dimensional probability density function transform (*N*-pdf<sub>t</sub>), for transferring colour information (e.g., red, green, and blue, RGB, colour channels) from one image to another with the goal of recolouring a target image to match the “feel” of a source image. To the best of the author's knowledge, the *N*-pdf<sub>t</sub> algorithm has not been explored outside of this context. When viewed more generally, the algorithm is a true multivariate version of quantile mapping that is proven to converge when the target distribution is multivariate normal (Pitié et al. 2007). Because the transformation is invertible, any continuous multivariate distribution can thus be mapped to another using the multivariate normal distribution as an intermediary. However, empirical evidence suggests that the use of an intermediate multivariate normal distribution is unnecessary and that direct mapping between distributions is possible via the *N*-pdf<sub>t</sub> algorithm.

In the context of climate simulations rather than image processing, if one replaces (and expands) the colour channels with climate variables, for example multiple weather elements from one or more spatial locations, then the same basic method should be an effective multivariate bias correction algorithm. In a climate modelling context, modifications are, however, necessary, especially when dealing with corrections not just between historical periods (i.e., between two images), but also to future climate projections or predictions where the range of variables may lie outside the historical range. In this case, it may be desirable to also preserve the climate change signal of the underlying climate model in the projection period, subject to bias correction of the historical period. As one example, trend-preservation is a fundamental property of the bias corrected climate model outputs provided in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) (Hempel et al. 2013). The question of whether or not trends should be preserved is discussed in depth by Maraun (2016), who concludes that “In case one has trust in the simulated change, one should employ a trend preserving bias correction.” If

the underlying simulated trends from the climate model are thought to be implausible, for example because of biases in large-scale circulation or local feedback processes, then this should be communicated explicitly along with the bias correction results. Alternatively, other methods may be required. If trends are to be preserved, univariate methods such as equidistant/equiratio quantile matching and quantile delta mapping algorithms have been proposed (Li et al. 2010; Wang and Chen 2014; Cannon et al. 2015). As pointed out by Cannon et al. (2015), the “goal then would be to avoid artificial deterioration of trends that arise simply as a statistical artifact of quantile mapping or related methods”. This feature has been incorporated into the correlation-based multivariate methods proposed by Cannon (2016). In this paper, the  $N$ -pdf algorithm is similarly extended for use with climate model projections.

Specifically, a version of the  $N$ -pdf algorithm tailored for climate models, referred to as MBCn, is introduced and illustrated using three examples spanning a range of dimensions from 3 to 25. First, MBCn is applied to a simple three dimensional problem inspired by the original image processing examples of Pitié et al. (2005, 2007). In addition to comparisons of distributional properties and error statistics, the computational demand of MBCn is compared against univariate quantile mapping. Second, MBCn is applied to seven variables from the Canadian Centre for Climate Modelling and Analysis Regional Climate Model (CanRCM4) (Scinocca et al. 2016)—3-hourly surface temperature, pressure, specific humidity, wind speed, incoming shortwave radiation, incoming longwave radiation, and precipitation—with a subset of corrected variables then used to calculate components of the Canadian Forest Fire Weather Index (FWI) system (Van Wagner and Forest 1987). The FWI has been adopted globally as a general index for the risk of wildfire. Calculations depend, in a nonlinear fashion, on current and past values of multiple weather elements. Simulated variables used to calculate the FWI are often bias corrected first, but without taking into account the dependence between variables (e.g., via quantile mapping in Lehtonen et al. 2016 or the delta method with variance inflation in Amatulli et al. 2013). Given that FWI is a multivariate index, it is possible that improvements in simulated FWI can be gained by applying a truly multivariate bias correction method like MBCn, which adjusts the full multivariate distribution of the weather elements. To determine the added value of MBCn, results are compared against the two multivariate bias correction methods from Cannon (2016), namely MBCp, which corrects Pearson correlation dependence structure, and MBCr, which corrects Spearman correlation dependence structure. In the third and final example, MBCn is used to correct biases in the spatial dependence structure of CanRCM4 precipitation fields. In this case, simulated precipitation amounts over 25 grid points are corrected simultaneously, with the ultimate

goal being the use of MBCn to both bias correct and down-scale precipitation.

## 2 $N$ -pdf algorithm

The  $N$ -pdf algorithm maps from a continuous  $N$ -dimensional multivariate source distribution to a continuous target distribution of the same dimension (Pitié et al. 2005, 2007). In the original application, the two multivariate distributions represent RGB colour channels from two images. The multivariate source distribution is corrected by iteratively applying a random orthogonal rotation to the datasets followed by quantile mapping of the rotated marginal distributions. Traditionally, quantile mapping operates on the marginal distributions of a dataset without considering the dependence between variables. In the  $N$ -pdf algorithm, the additional rotation step provides linear combinations of the original variables—rather than each original variable separately—to the univariate quantile mapping bias correction. When these two steps are combined in sequence and repeated, correction of the multivariate distribution becomes possible. In brief, the algorithm consists of three steps: (a) apply an orthogonal rotation to the source and target data; (b) correct the marginal distributions of the rotated source data via empirical quantile mapping; and (c) apply the inverse rotation to the resulting data. These three steps are repeated, iteratively, until the multivariate distribution matches the target distribution, i.e., the corrected source and target colour channel histograms—marginal and joint—are the same. The algorithm is proven to converge with random rotation matrices and a multivariate Gaussian target (Pitié et al. 2007), which means that any distribution can be mapped to another by using the multivariate Gaussian as a pivot. However, Pitié et al. (2005, 2007) show empirically that the intermediate Gaussian is unnecessary and that direct mapping between arbitrary continuous distributions is possible. Details are provided by Pitié et al. (2005, 2007).

The  $N$ -pdf algorithm starts with an  $I \times N$  matrix of source data  $\mathbf{X}_S$  and a corresponding matrix of target data  $\mathbf{X}_T$  where the  $N$  variables in both are arranged as columns. Without loss of generality, assume that these variables are each standardized to have zero mean and unit standard deviation. In step (a) of the  $j$ th iteration of the  $N$ -pdf algorithm, denoted by the superscript  $[j]$ , construct an  $N \times N$  uniformly distributed random orthogonal rotation matrix  $\mathbf{R}^{[j]}$ , for example via QR decomposition of normally distributed random values (Mezzadri 2007). Rotate the source and target data

$$\begin{aligned}\tilde{\mathbf{X}}_S^{[j]} &= \mathbf{X}_S^{[j]} \mathbf{R}^{[j]} \\ \tilde{\mathbf{X}}_T^{[j]} &= \mathbf{X}_T^{[j]} \mathbf{R}^{[j]}.\end{aligned}\quad (1)$$

In step (b), use univariate quantile mapping to map each of the  $n$  of  $N$  empirical cumulative distribution functions (CDFs)  $\tilde{F}_S^{(n)[j]}$  to the corresponding empirical CDFs  $\tilde{F}_T^{(n)[j]}$ ,

which are associated with, respectively, the vectors of source values  $\tilde{\mathbf{x}}_S^{(n)[j]}$  and target values  $\tilde{\mathbf{x}}_T^{(n)[j]}$ . Correction of  $\tilde{x}_S^{(n)[j]}(i)$ , the  $i$ th of  $I$  source values, by quantile mapping is given by

$$\hat{x}_S^{(n)[j]}(i) = \tilde{F}_T^{(n)[j]-1} \left( \tilde{F}_S^{(n)[j]} \left( \tilde{x}_S^{(n)[j]}(i) \right) \right) \quad (2)$$

where  $\hat{x}_S^{[n][j]}(i)$  is the corrected source value. In step (c) apply the inverse rotation matrix to yield source values for the next iteration  $j + 1$

$$\mathbf{X}_S^{[j+1]} = \hat{\mathbf{X}}_S^{[j]} \mathbf{R}^{[j]-1}. \quad (3)$$

The target data are carried forward to the next iteration unchanged  $\mathbf{X}_T^{[j+1]} = \mathbf{X}_T^{[j]}$ . Repeat steps (a) to (c) until the source distribution converges to the target distribution. As noted above, a proof of convergence is provided by Pitié et al. (2007).

Convergence to the target multivariate distribution is measured in the remainder of this paper using the energy distance (Székely and Rizzo 2004, 2013). The (squared) energy distance is a measure of statistical discrepancy between two multivariate distributions, defined between  $N$ -dimensional independent random vectors  $\mathbf{x}$  and  $\mathbf{y}$  with CDFs  $F$  and  $G$ , respectively, as

$$D^2(F, G) = 2E\|\mathbf{x} - \mathbf{y}\| - E\|\mathbf{x} - \mathbf{x}'\| - E\|\mathbf{y} - \mathbf{y}'\| \geq 0 \quad (4)$$

where  $E$  denotes the expected value,  $\|\cdot\|$  is the Euclidean norm, and  $\mathbf{x}'$  and  $\mathbf{y}'$  are independent and identically distributed copies of  $\mathbf{x}$  and  $\mathbf{y}$ .  $D(F, G)$  equals zero only when  $F$  equals  $G$ . Calculation details are given in Rizzo and Székely (2016).

### 3 MBCn algorithm

Three datasets are involved when bias correction algorithms are applied to climate model data: historical observations (i.e.,  $\mathbf{X}_T$ ); historical climate model simulations (i.e.,  $\mathbf{X}_S$ ); and, additionally, climate model projections/predictions  $\mathbf{X}_P$ , which will, at least in part, typically lie outside the historical time period. Bias corrected values of  $\mathbf{X}_P$  are of primary interest. In the  $N$ -pdf algorithm, when quantile mapping (Eq. 2) is applied to each modelled variable within the historical period (i.e.,  $P = S$ ), the bias corrected values will, by definition, have the same marginal distribution as the observed historical values; the quantile mapping transfer function relies exclusively on the historical CDFs. Difficulties can arise when quantile mapping is applied to projected/predicted data that lie outside the range of the historical simulations. Some method of extrapolation must be used to handle these cases. Because colour channels have fixed lower/upper bounds (zero intensity to full intensity), extrapolation is less of an issue when the  $N$ -pdf

algorithm is applied to multiple source images in an image processing context. However, the same will not be true of climate model simulations with a strong climate change signal, for instance centennial climate projections or decadal climate predictions.

As a remedy, methods like equidistant/equiratio quantile matching (Li et al. 2010; Wang and Chen 2014) make additional use of the simulated climate model CDF  $F_P$  in the projection/prediction period. Cannon et al. (2015) showed these methods to be a quantile mapping form of the “delta change method” (Olsson et al. 2009), whereby projected/predicted changes in the simulated quantiles are preserved following quantile-by-quantile bias correction. As a consequence, extrapolation is integral to the transformation. Because of this link, Cannon et al. (2015) refer to this general approach as quantile delta mapping (QDM). The QDM transfer function that preserves absolute changes in quantiles (e.g., for an interval variable such as temperature measured on the Celsius scale) is given by

$$\begin{aligned} \Delta(i) &= x_P(i) - F_S^{-1}(F_P(x_P(i))) \\ \hat{x}_P(i) &= F_T^{-1}(F_P(x_P(i))) + \Delta(i) \end{aligned} \quad (5)$$

with the corresponding transfer function that preserves relative changes in quantiles (e.g., for a ratio variable with an absolute zero such as precipitation) obtained simply by replacing the addition/subtraction operators with multiplication/division operators.

The MBCn algorithm extends the  $N$ -pdf algorithm by replacing quantile mapping with QDM and working explicitly with the  $\mathbf{X}_S$ ,  $\mathbf{X}_P$ , and  $\mathbf{X}_T$  datasets. In step (a), rotate the source, projection/prediction, and target datasets

$$\begin{aligned} \tilde{\mathbf{X}}_S^{[j]} &= \mathbf{X}_S^{[j]} \mathbf{R}^{[j]} \\ \tilde{\mathbf{X}}_P^{[j]} &= \mathbf{X}_P^{[j]} \mathbf{R}^{[j]} \\ \tilde{\mathbf{X}}_T^{[j]} &= \mathbf{X}_T^{[j]} \mathbf{R}^{[j]}. \end{aligned} \quad (6)$$

In step (b), apply the absolute change form of QDM (Eq. 5) to each variable in  $\tilde{\mathbf{X}}_S^{[j]}$  and  $\tilde{\mathbf{X}}_P^{[j]}$  using the corresponding variable in  $\tilde{\mathbf{X}}_T^{[j]}$  as the target, yielding  $\hat{\mathbf{X}}_S^{[j]}$  and  $\hat{\mathbf{X}}_P^{[j]}$ . In step (c), rotate back

$$\begin{aligned} \mathbf{X}_S^{[j+1]} &= \hat{\mathbf{X}}_S^{[j]} \mathbf{R}^{[j]-1} \\ \mathbf{X}_P^{[j+1]} &= \hat{\mathbf{X}}_P^{[j]} \mathbf{R}^{[j]-1} \\ \mathbf{X}_T^{[j+1]} &= \mathbf{X}_T^{[j]} \end{aligned} \quad (7)$$

Repeat steps (a)–(c) until the multivariate distribution of  $\mathbf{X}_S^{[j+1]}$  matches  $\mathbf{X}_T$ .

The ratio property of variables like precipitation is lost during the rotation in step (a). Hence, the ratio version of QDM cannot be used in step (b) and it is not possible to take advantage of its trend preserving property. More generally, repeated iterations may lead to corruption of the model-projected trend for both interval and ratio variables.



To restore the appropriate trends to  $\mathbf{X}_p^{[j+1]}$ , an extra step is added to the MBCn algorithm.

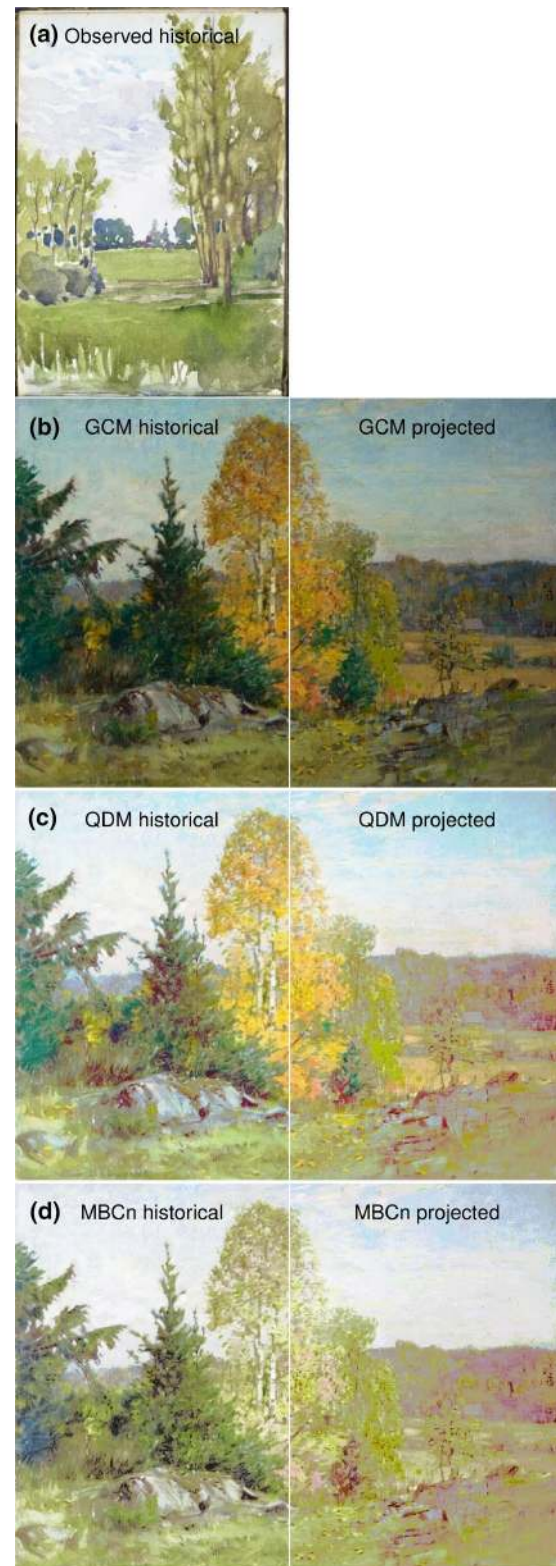
In step (d), apply the appropriate absolute/ratio version of QDM to each variable of the original  $\mathbf{X}_p$  dataset using  $\mathbf{X}_T$  and  $\mathbf{X}_S$  as historical baseline data, yielding  $\hat{\mathbf{X}}_p$ ; finally, replace quantiles of each column of the  $\mathbf{X}_p^{[j+1]}$  matrix from step (c) with those from columns of  $\hat{\mathbf{X}}_p$  obtained in the application of QDM from step (d). In short, elements of each column of  $\hat{\mathbf{X}}_p$  are ordered according to the ordinal ranks of the corresponding elements of each column of  $\mathbf{X}_p^{[j+1]}$ . (Note: the reordering operation is a fundamental part of the EC-BC method of Vrac and Friederichs (2015) and is also used in the MBCp and MBCr algorithms by Cannon (2016).) This maintains the preservation of trends in  $\hat{\mathbf{X}}_p$  by QDM and the rank dependence structure of  $\mathbf{X}_p^{[j+1]}$ .

#### 4 Image processing example

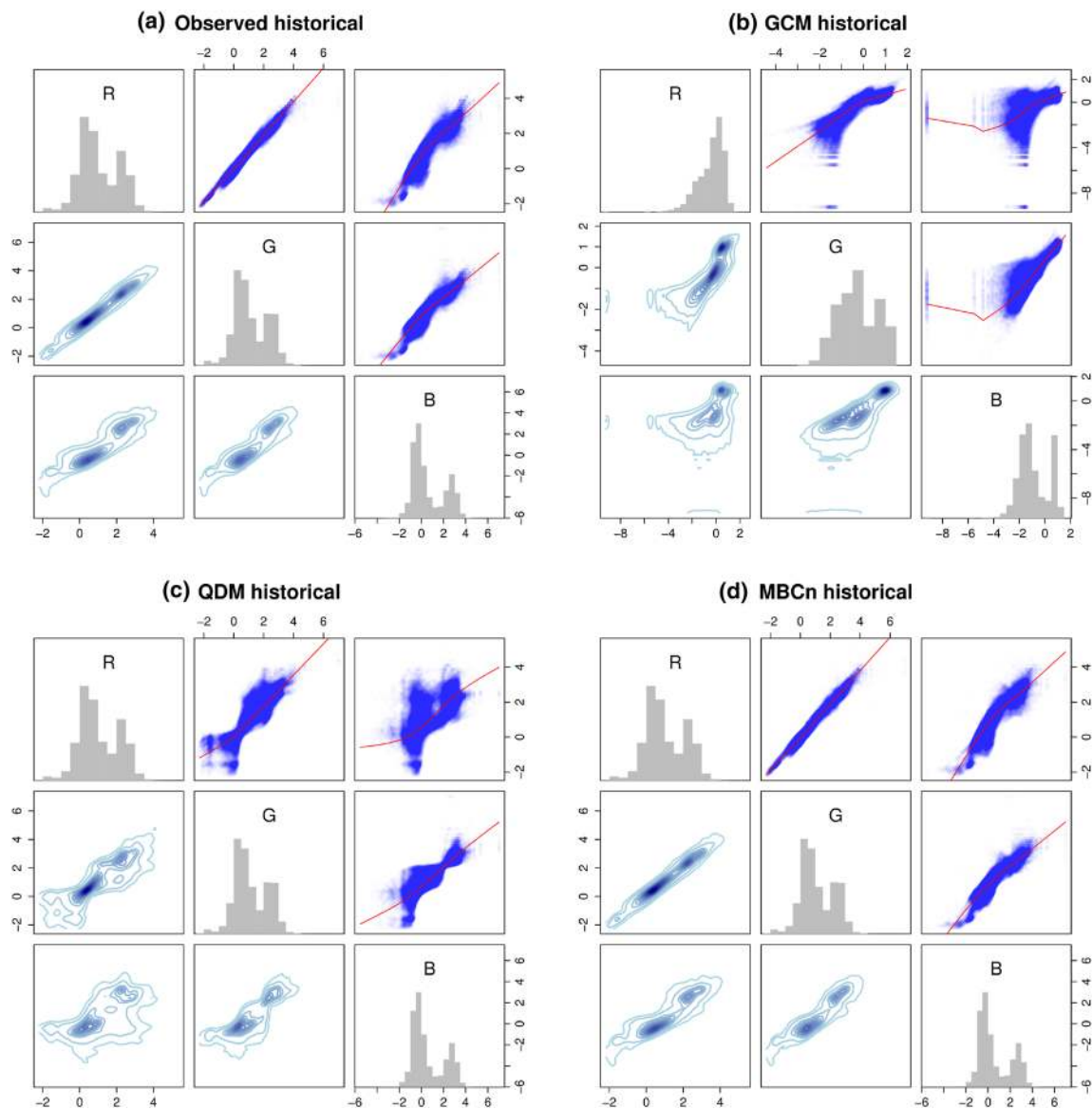
An image recolouring example is used here to demonstrate the MBCn algorithm. This is in the spirit of the image recolouring applications of the  $N$ -pdf algorithm by Pitié et al. (2005, 2007), but the experimental design is crafted to mimic characteristics of a climate projection problem. Trivariate data are RGB colour channels from two images. In this case, the  $428 \times 300$  pixel image in Fig. 1a represents historical observations  $\mathbf{X}_T$  and the two halves of the  $428 \times 600$  pixel image in Fig. 1b represent, on the left side, historical  $\mathbf{X}_S$  and, on the right side, projected  $\mathbf{X}_p$  outputs from a climate model. The total number of pixels in each image/image half is roughly the same as the number of 3-h time steps in a 50-year climate model simulation. Note that the content of  $\mathbf{X}_T$  and  $\mathbf{X}_S$  is similar—both are landscape paintings of trees—but the colour palette and brightness of the two images differ substantially.  $\mathbf{X}_p$  is broadly similar to  $\mathbf{X}_S$ , but with a different mix of foreground and background elements. The goal is to recolour (or, in a climate modelling context, bias correct)  $\mathbf{X}_S$  and  $\mathbf{X}_p$  using  $\mathbf{X}_T$  as the target image (or historical observations).

Because MBCn preserves changes in quantiles and RGB colour intensities are bounded between 0–1, it is possible that corrections to  $\mathbf{X}_p$  will lead to values outside of these bounds. Following Demidenko (2006), data on the unit interval are therefore first mapped onto the real line using the  $\text{logit}(x) = \log(\frac{x}{1-x})$  transformation. In a climate modelling context, the same transformation could be used for bounded variables like cloud cover or relative humidity. Transformed values are corrected and then inverse transformed prior to plotting.

Correcting each colour channel separately using QDM, i.e., without taking into account the dependence between colour channels, leads to the image halves shown in



**Fig. 1** Images representing **a** historical target observations; **b** global climate model (GCM) historical source (*left*) and future projections (*right*); **c** QDM historical (*left*) and projected corrections (*right*); and **d** MBCn historical (*left*) and projected corrections (*right*) following 30 iterations



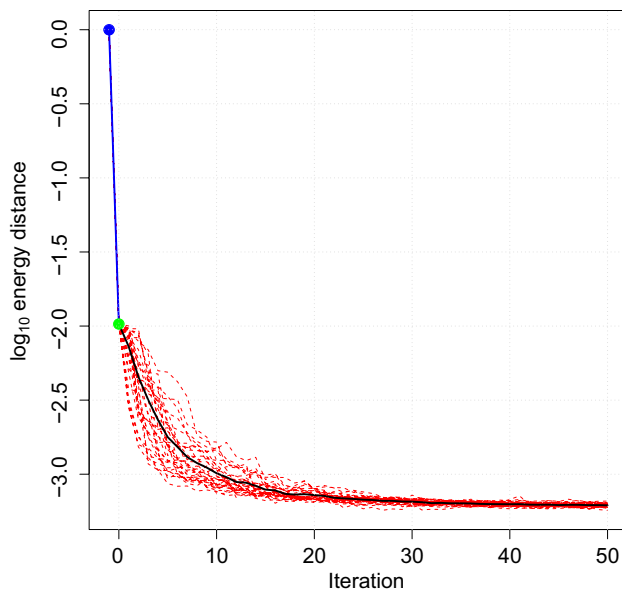
**Fig. 2** Logit transformed RGB channel marginal histograms (*diagonal panels*), pairwise bivariate histograms (*lower triangle panels*), and pairwise scatterplots with best fit local regression lines (*upper*

*triangle panels*) for the **a** observed historical target, **b** GCM historical source, **c** QDM historical, and **d** MBCn historical (30 iterations) associated with the images in Fig. 1

Fig. 1c. The overall character of the corrected image halves, notably the brightness, is now more similar to the target image, but the recolouring has not been fully successful. For instance, note the presence of yellow/orange tones in the corrected image halves that are absent from the target image (Fig. 1a). After applying the MBCn algorithm, the overall structure, brightness, and colour palette of the corrected image halves (Fig. 1d) is consistent with the target. This can be seen more clearly in Fig. 2, which shows marginal histograms and pairwise bivariate histograms of the observed historical target, GCM historical source, QDM historical, and MBCn

historical RGB channels. After correction by QDM or MBCn, the marginal distributions in the historical period are, by construction, identical to the target image marginal distributions. MBCn further corrects the historical dependence between the image channels such that the full multivariate distribution matches the target.<sup>1</sup> In the pro-

<sup>1</sup> In comparison to the EC-BC method by Vrac and Friederichs (2015), which also corrects the full multivariate distribution within the calibration sample, MBCn preserves much of the underlying content of the climate model within the projection period. EC-BC repeats the climate model's rank sequencing from the calibration period to the projection period, i.e., the left half of Fig. 1d is, aside from small changes in brightness/colour, duplicated in the right half of the image (Figure S1).



**Fig. 3** Values of the energy distance taken with respect to the target image distribution (rescaled so that the uncorrected source image has a value of one) following each iteration of 30 trials of the MBCn algorithm (red dashed lines; mean in black), as well as values for the uncorrected source image (blue circle) and QDM correction of the marginal distributions (green circle). The vertical axis is shown in  $\log_{10}$  units

jection period, MBCn also applies QDM's quantile change preserving property on the marginal distributions. The bias corrected dependence structure in the projection period—how the ordinal ranks from the climate model are adjusted—depends both on the  $N$ -pdf algorithm corrections to the historical dependence structure and also the underlying changes in the climate model dependence structure.

MBCn is an iterative algorithm whose convergence depends, in part, on the application of a sequence of random rotation matrices. Speed of convergence to the target multivariate distribution is shown here using the energy distance. Figure 3 shows energy distances with respect to the target image for 50 iterations of MBCn, as well as values for the uncorrected source image and after correction by QDM. To assess the influence of the random rotation matrices, MBCn results are shown for 30 different trials of the algorithm. Convergence occurs relatively quickly, with energy distances reduced, on average, by an order of magnitude relative to QDM (and three orders of magnitude relative to the uncorrected source image) after 10 iterations of the MBCn algorithm. The influence of the random rotations is suppressed after  $\sim 30$  iterations, with all 30 trials converging to the same energy distance after this point.

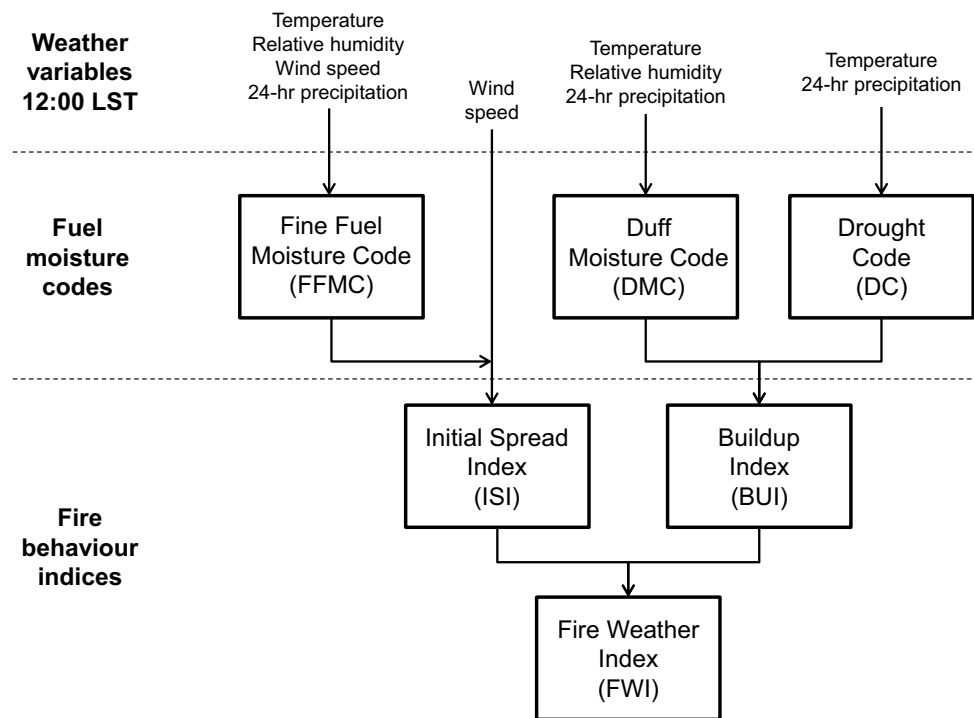
## 5 Canadian Fire Weather Index example

In this section, the MBCn algorithm is applied to a real-world climate modelling application, namely calculation of Canadian Forest Fire Weather Index (FWI) system components based on simulated outputs from the CanRCM4 regional climate model. Studies have relied on the FWI to assess future changes in fire risk based on global and regional climate model outputs (Flannigan et al. 2009; Amatulli et al. 2013; Lehtonen et al. 2016). In addition, medium range to seasonal predictions of the FWI have also been issued based on numerical weather and climate prediction models (Anderson et al. 2007; Pappenberger et al. 2013). The FWI is the ultimate index in the FWI system, which is made up of FWI and its five component indices, each representing a different aspect of fuel moisture or fire behaviour (Fig. 4). As stated by Van Wagner and Forest (1987), the Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), and Drought code (DC) represent, respectively, moisture contents of “litter and other cured fine fuels”, “loosely compacted decomposing organic matter”, and “a deep layer of compact organic matter”, while the Initial Spread Index (ISI) and Buildup Index (BUI) represent, respectively, the “rate of spread alone without the influence of variable quantities of fuel” and “the total fuel available to the spreading fire”. As shown in Fig. 4, these five indices, which are updated on a daily basis using local temperatures, relative humidities, wind speeds, and 24-h precipitation amounts (all at 12:00 LST), contribute to the final FWI. Following Amatulli et al. (2013), FWI system calculations are initialized at the start of each calendar year and are integrated forward through to the end of the year. Although FWI is of primary interest, values of the other component indices will also be used to gauge the performance of MBCn relative to three benchmark bias correction algorithms. Specifically, results are compared against univariate QDM, as well as the two multivariate methods introduced by Cannon (2016): MBCp, which corrects Pearson correlation dependence structure, and MBCr, which corrects Spearman correlation dependence structure.

As pointed out by Maraun (2016), evaluation of bias correction algorithms purely in terms of distributional properties (e.g., as corrected with asynchronous MOS techniques) is typically “not sufficient to identify artificial skill and unskillful bias correction”. For verification, the experimental setup is thus chosen so that the series of simulated and observed variables are synchronous in time, in this case by using a reanalysis-driven regional climate model. Improvements can be investigated by comparing bias corrected time series directly against observations using standard forecast verification statistics. To this end, dynamically downscaled outputs over a North American spatial domain are obtained from the CanRCM4 regional climate model (Scinocca et al.



**Fig. 4** Flowchart showing components of the Canadian Forest Fire Weather Index System after, (Van Wagner and Forest 1987)



2016). The CanRCM4 historical evaluation run, which covers the 1989–2009 period, relies on lateral boundary conditions provided by the ERA-Interim reanalysis (Dee et al. 2011) and employs interior spectral nudging to constrain large scales to respect the reanalysis driving fields.

Three hourly CanRCM4 outputs of precipitation, surface temperature, pressure, specific humidity, wind speed, incoming shortwave radiation, and incoming longwave radiation on a 0.44-deg model grid are mapped onto a regular 0.5-deg grid following the CORDEX NAM-44i domain specification (Christensen et al. 2014). Only land grid points are retained for analysis. This set of variables is sufficient to support a wide variety of climate impacts and adaptation studies, including simulating elements of the terrestrial water cycle by distributed hydrological models (Weedon et al. 2014) and calculating climate extremes indices (Zhang et al. 2011), multivariate drought indices (Vicente-Serrano et al. 2010), and fire weather indices like FWI (Van Wagner and Forest 1987). Following Haddeland et al. (2012), bias correction algorithms are applied to all seven variables, notwithstanding whether or not the index of interest, in this case FWI, requires the full set. This is consistent with a general purpose application of climate model post-processing to support multiple end users.

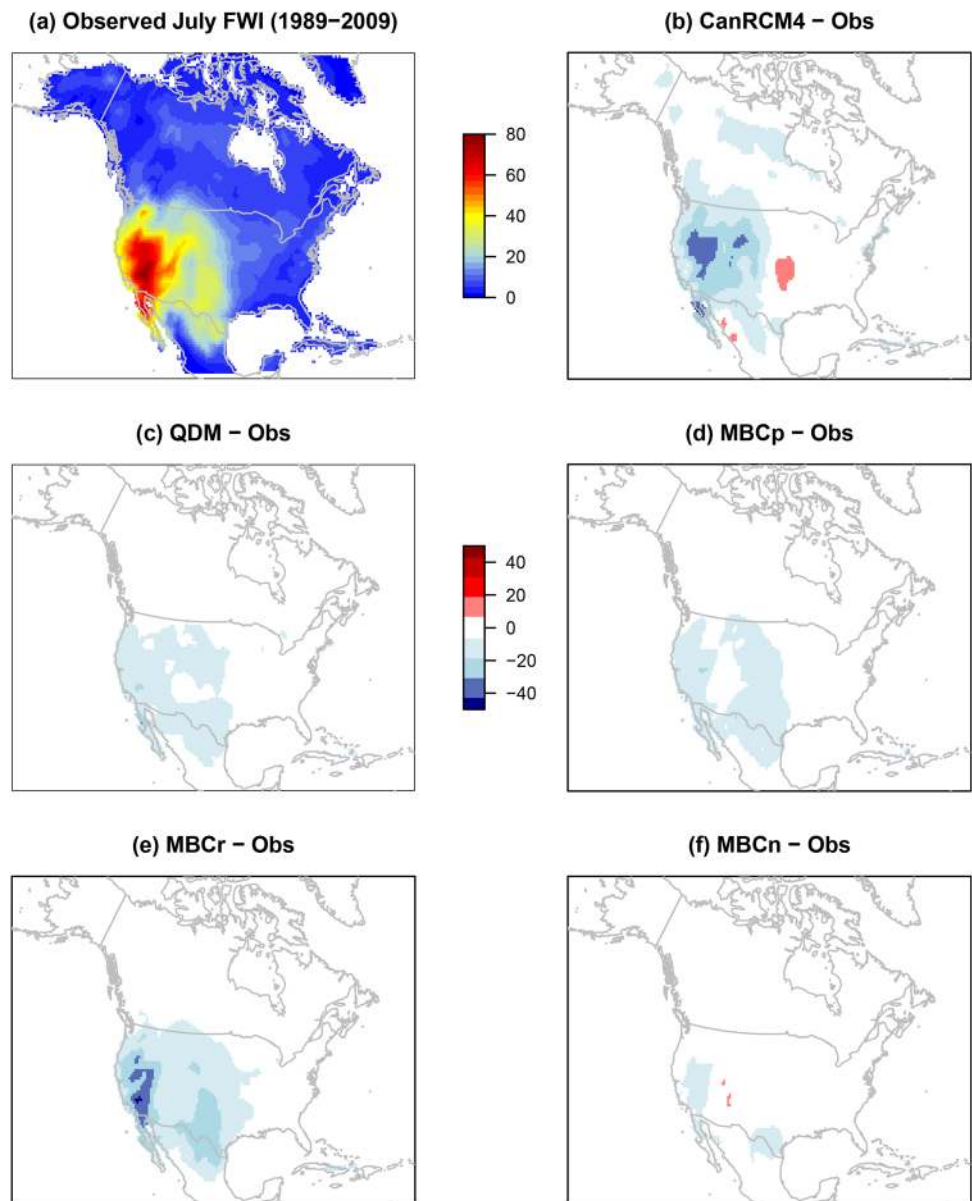
The 3-hourly 0.5-deg WATCH Forcing Dataset applied to ERA-Interim (WFDEI) (Weedon et al. 2014) is used as the observational target for the bias correction algorithms. A major assumption, which is shared by all methods, is that the underlying observational reference provides an accurate

representation of the true historical climate. Observational uncertainty is, however, a reality and no gridded observational dataset is perfect. For example, Rust et al. (2015) found discontinuities in WFDEI daily temperatures across month boundaries that resulted from a climatological correction to match Climatic Research Unit monthly temperatures. Despite such artifacts, recent work by Essou et al. (2016) concluded that the WFDEI dataset serves as a reasonable proxy to observed surface temperature and precipitation and can be reliably used as a forcing dataset for hydrological modelling purposes. Given that the focus of this study is on internal consistency of bias corrected variables relative to the WFDEI target, irrespective of its potential shortcomings, alternative reference datasets are not considered.

Bias corrections are applied separately to the central calendar month of sliding three month windows for each grid point in the NAM-44i domain. The final QDM corrections used in each algorithm treat precipitation, specific humidity, wind speed, and solar radiation components as ratio variables and temperature (in degrees Celsius) and pressure as interval variables. For precipitation, following Cannon et al. (2015) and Vrac et al. (2016), dry days are treated as censored values below a trace amount (0.05 mm). Exact zeros are replaced with non-zero uniform random values below the trace threshold prior to bias correction; values below the threshold after bias correction are set to zero. When applying MBCn, the same sequence of random rotation matrices is used at each grid point; taking into account



**Fig. 5** July mean FWI (1989–2009) based on **a** WFDEI observations; subsequent panels show differences between **b** raw CanRCM4 outputs and observed values; **c** QDM and observed values; **d** MBCp and observed values; **e** MBCr and observed values; and **f** MBCn and observed values. For **b–f**, a value of zero indicates perfect match with observations. All bias corrected results are based on cross-validation



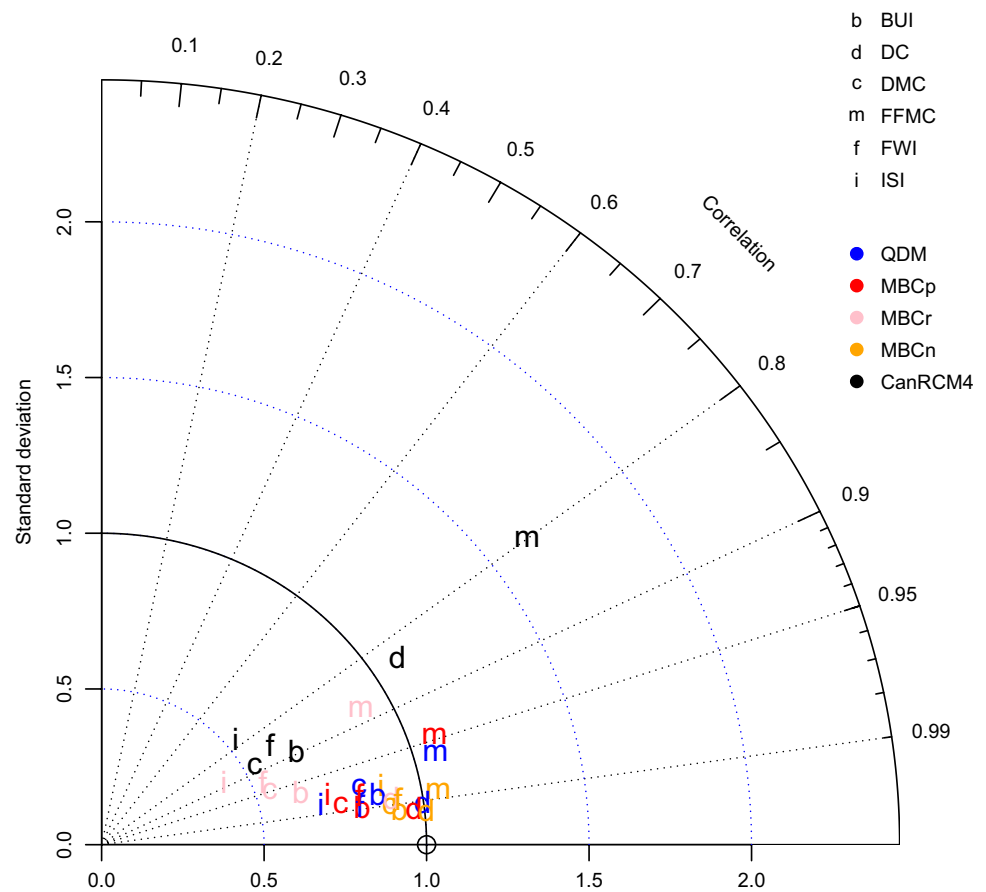
convergence properties (e.g., Fig. 3) and computational constraints, the algorithm is terminated after 20 iterations.

The 1989–1999 portion of the record is first used for calibration and the 2000–2009 portion is used as the projection period for out-of-sample verification; the two portions are then swapped, in a two-fold cross-validation design, so that each is used in turn both for calibration and for verification. The calibration sample thus provides the historical observational target  $\mathbf{X}_T$  and raw climate model simulation  $\mathbf{X}_S$  datasets, whereas the out-of-sample verification period supplies the climate model projection  $\mathbf{X}_P$  dataset. As noted earlier, the CanRCM4 simulation is forced by the ERA-Interim reanalysis, which means that WATCH Forcing observations are available in the verification period for

calculation of model performance statistics for the bias corrected projections.

FWI requires values of 24-h accumulated precipitation, surface temperature, relative humidity, and wind speed. This subset of four variables is calculated from the full set of seven (i) raw CanRCM4 outputs as well as bias corrected outputs from (ii) univariate QDM; (iii) MBCp, which additionally corrects Pearson correlation dependence structure, and (iv) MBCr, which additionally corrects Spearman rank correlation dependence structure; and finally (v) MBCn, which corrects the full multivariate distribution. Note that the marginal distributions of the seven bias corrected variables are identical for QDM, MBCp, MBCr, and MBCn; only the dependence between variables differs. In all cases,

**Fig. 6** Taylor diagram showing spatial pattern correlations and standard deviations of cross-validated climatological mean July FWI, BUI, DC, DMC, FFMFC, and ISI based on CanRCM4, QDM, MBCp, MBCr, and MBCn outputs. Observational reference values are 1989–2009 WFDEI climatologies scaled to have unit standard deviation



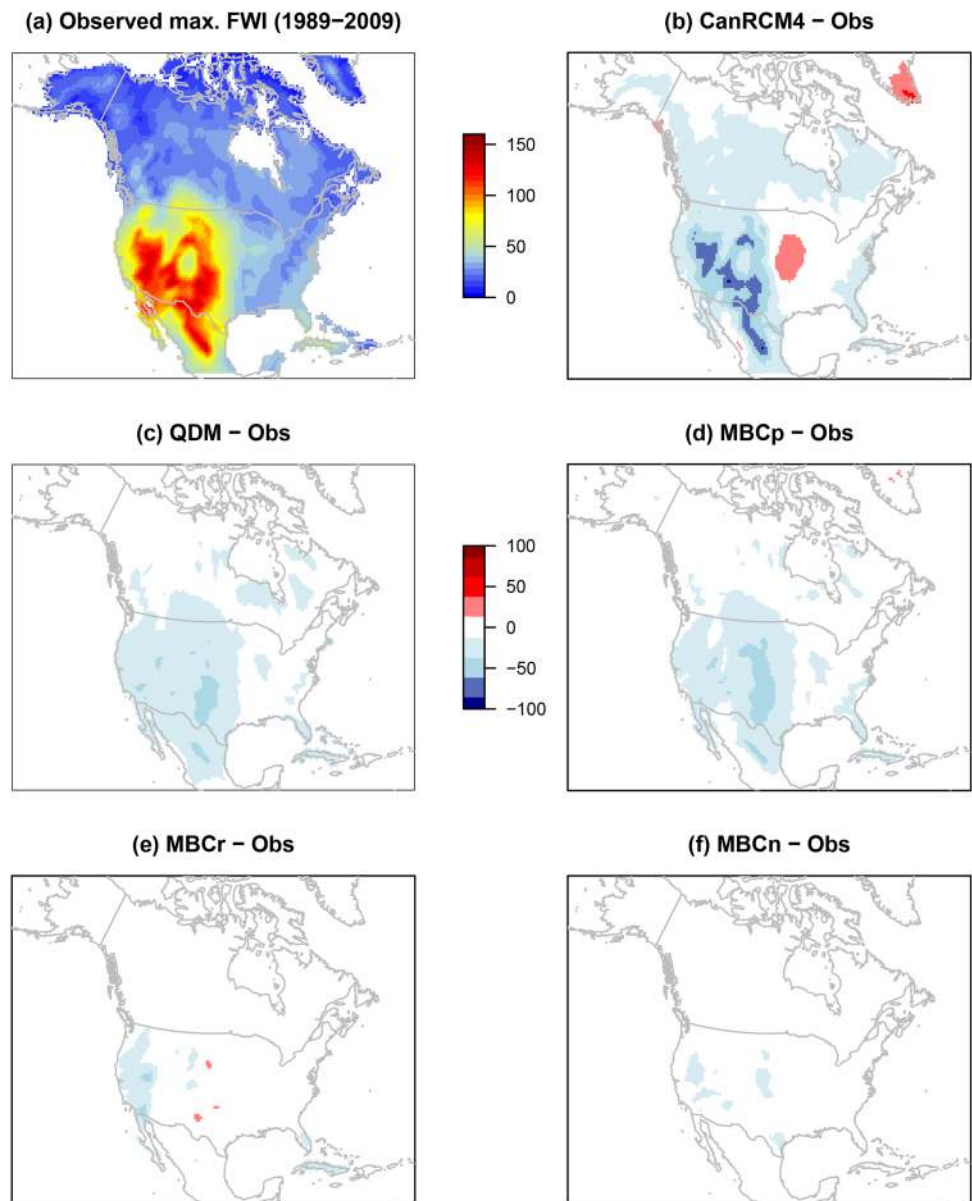
3-hourly values are interpolated to 12:00 LST and FWI is calculated using software by Wang et al. (2013).

Climatological mean values of FWI during July, when fire risk is typically highest, are shown in Fig. 5. In observations, peak FWI values are found over the southwestern United States (US), with high values extending into the central US. The spatial extent of maxima in CanRCM4 is broadly similar, but with a trough of low values between the southwestern US and central US that is absent from observations. In addition, peak values are considerably lower than observed. Spatial patterns of FWI in QDM, MBCp, MBCr, and MBCn are improved relative to CanRCM4, as are peak values in QDM, MBCp, and MBCn. The representation of peak values by MBCr, however, is worse than QDM, which does not consider inter-variable relationships, or MBCp, which considers Pearson rather than Spearman rank correlations between variables. As pointed out by Cannon (2016), the Pearson and Spearman rank correlations are not guaranteed to fully specify the dependence structure of real-world multivariate datasets; in practice, each may better describe different aspects of a given dataset. Do these results hold for the other components of the FWI system? Spatial pattern correlations and standard deviations of climatological mean July FWI and

the five FWI component indices are summarized in the Taylor diagram shown in Fig. 6. Spatial correlations with observations for CanRCM4 tend to lie between  $\sim 0.8$  and  $0.9$  with spatial standard deviations between 50 and 160% of observed values. Spatial correlations improve for the bias correction algorithms, with largest improvements (in order) for MBCr, QDM, MBCp, and finally MBCn, which exceeds 0.98 for all indices. MBCr exhibits the largest differences in spatial variability relative to observations, with smaller ranges seen for QDM and MBCp; values for MBCn are very tightly clustered around the observed value (85–105%).

Results reported in Figs. 5 and 6 are for the July monthly mean. To illustrate performance for the most extreme conditions, climatological values of the annual maximum daily FWI are shown in Fig. 7. A slightly different picture emerges for extreme fire risk. In this case, results for CanRCM4 are similar to those for July mean values, but those for QDM, MBCp, and MBCr are different. In particular, univariate QDM and MBCp, both of which exhibited close correspondence with observations for July mean conditions, now perform worse than MBCr, which was the worst method above for the July mean. On the other hand, MBCn, which performed best for

**Fig. 7** As in Fig. 5, but for annual maximum FWI (1989–2009). For **b–f**, a value of zero indicates perfect match with observations. All bias corrected results are based on cross-validation



July mean conditions, also performs best here for annual maxima. The change in performance for QDM, MBCp, and MBCr, but not MBCn, suggests that correcting the multivariate distribution of the four weather variables—rather than neglecting it (QDM) or only correcting either Pearson (MBCp) or Spearman rank (MBCr) correlation structure—may be needed to simulate the entire distribution of daily FWI values.

Results reported to this point have been for climatological values. Because the CanRCM4 evaluation run is forced at the boundaries by ERA-Interim and uses spectral nudging, direct comparison with the observed WFDEI time series is also possible. Mean absolute error (MAE) and mean absolute  $\log_{10}$  accuracy ratio (LAR) statistics with respect to observations are calculated for time series of July mean and annual

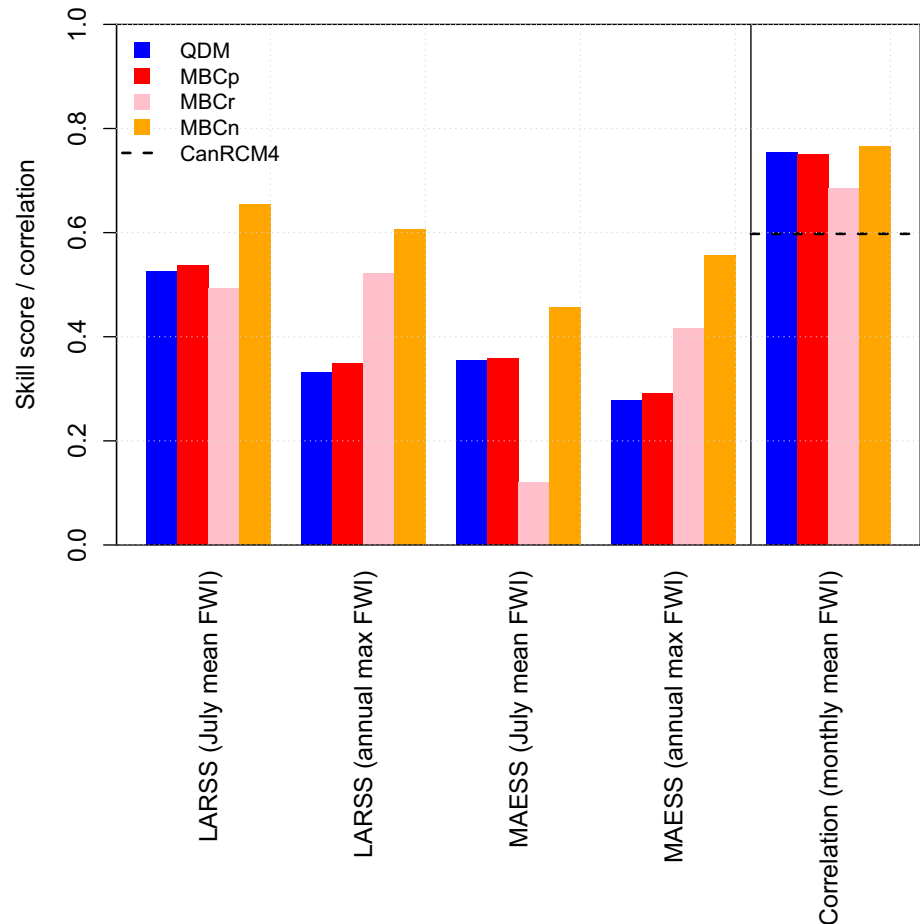
maximum daily FWI are each grid point over the North American domain

$$\text{MAE} = \frac{1}{I} \sum_{i=1}^I |\widehat{\text{FWI}}(i) - \text{FWI}(i)| \quad (8)$$

$$\text{LAR} = \frac{1}{I} \sum_{i=1}^I \left| \log_{10} \left( \frac{\widehat{\text{FWI}}(i)}{\text{FWI}(i)} \right) \right| \quad (9)$$

where  $\text{FWI}(i)$  and  $\widehat{\text{FWI}}(i)$  are, respectively, observed and modelled values of FWI in the  $i$ th of  $I$  out-of-sample verification years. The LAR statistic is a symmetric measure of relative accuracy (e.g., modelled values that differ from observations by a factor of 1/10 or 10 are both assigned

**Fig. 8** Cross-validated domain mean LAR and MAE skill scores relative to CanRCM4 for bias corrected time series of July mean and annual maximum FWI, as well as cross-validated domain mean temporal correlations between bias corrected and observed monthly mean FWI time series (the *black dashed line* shows the domain mean CanRCM4 correlation)



LAR = 1). Finally, temporal correlations between observed and modelled monthly mean FWI time series are calculated at each grid point. To summarize, MAE and LAR statistics are calculated for each bias correction method, averaged over the domain (weighted by grid cell area), and expressed as skill scores (MAESS and LARSS) relative to the raw CanRCM4 outputs; positive skill score values indicate improved performance relative to CanRCM4, with values equal to one corresponding to perfect skill (no error with respect to observations). For correlations, values are first converted to z-scores using the Fisher transformation, averaged, and then transformed back to correlation units (Silver and Dunlap 1987). Results are shown in Fig. 8. MBCn performs best for the MAE, LAR, and correlation statistics. MBCn, with a mean value over the four skill scores of +0.57 (versus +0.39 for MBCr, +0.38 for MBCp, and +0.37 for QDM), outperforms the other three bias correction methods by a substantial margin. Similarly, the domain mean correlation is highest for MBCn (+0.77), followed by QDM and MBCp (+0.75), MBCr (+0.69), and CanRCM4 (+0.60). For QDM, MBCp, and MBCr, contrasting performance statistics for July mean and annual maximum FWI, reported earlier for the climatological values, are also evident in the MAE and LAR statistics.

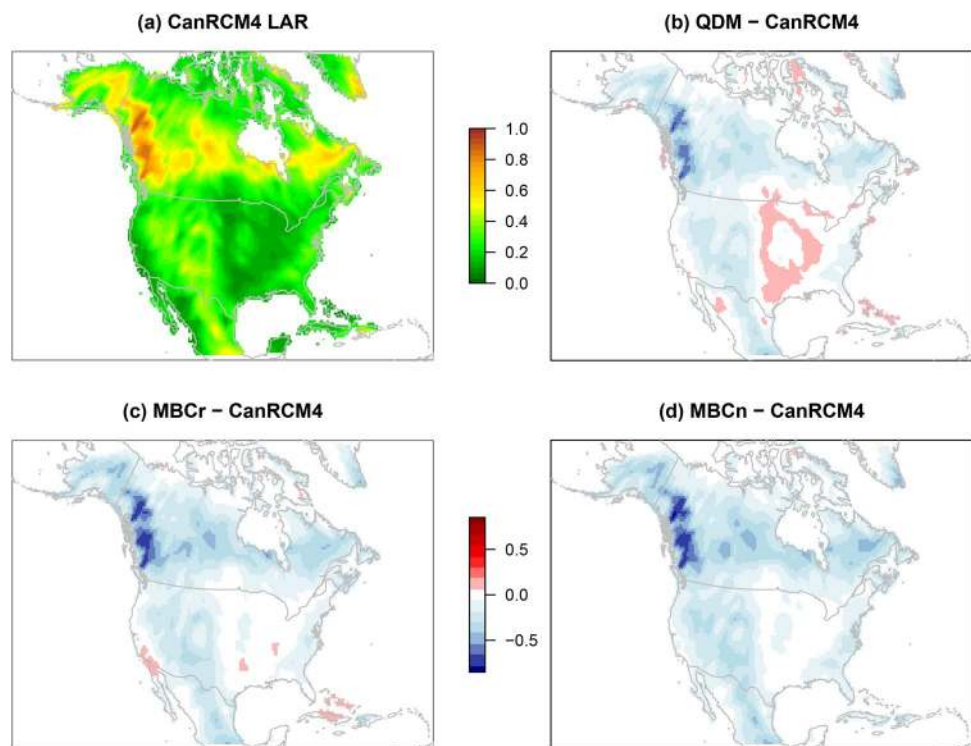
For reference, the spatial distribution of one of the performance statistics, LAR of the annual maximum daily FWI, is shown in Fig. 9 for CanRCM4, QDM, MBCr, and MBCn. Values for MBCp are similar to those for QDM and are omitted. For CanRCM4, the largest absolute relative errors are located over the North American Cordillera, in particular western Canada, with secondary maxima extending eastward across the middle of Canada. Improvements in relative error with respect to CanRCM4 for the bias correction methods follow the same spatial pattern. In terms of land area, QDM, MBCr, and MBCn perform better than CanRCM4 over 75, 87, and 96% of the domain respectively. Summaries for the full suite of performance statistics are shown in Fig. 10. MBCn performs better than CanRCM4 over more of the domain (>91% for all statistics) than any of the other methods.

## 6 Spatial precipitation example

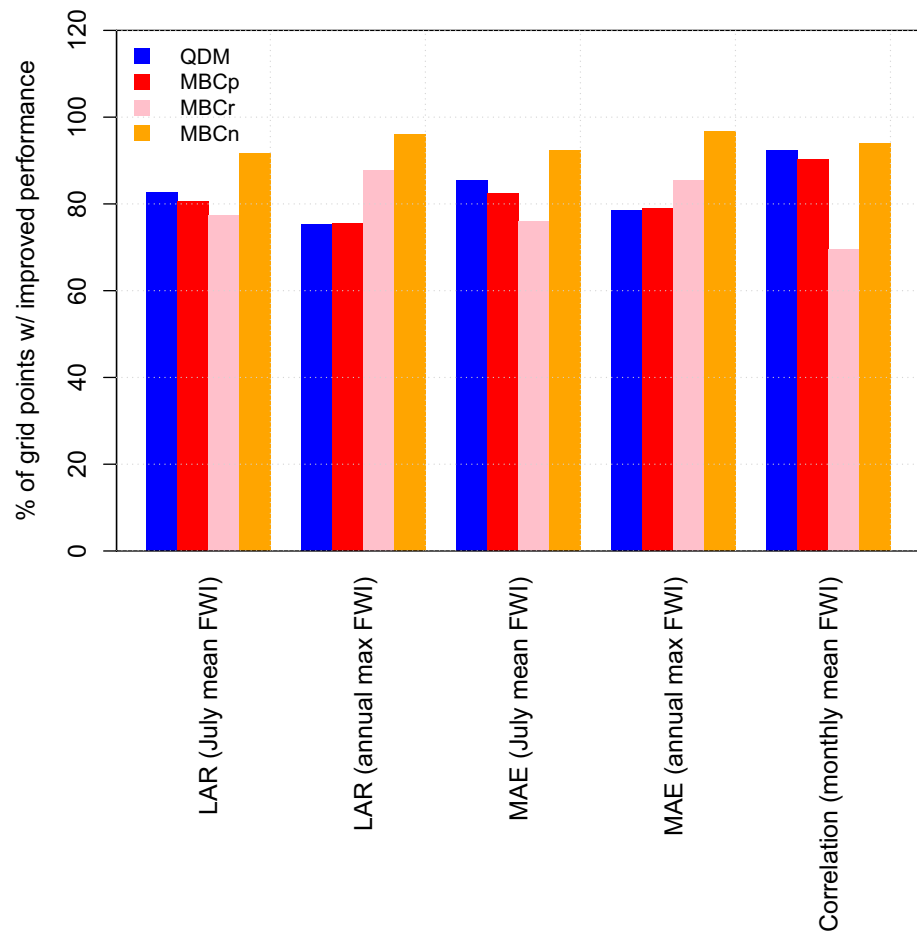
Section 5 deals with bias correction of climate model outputs, on a grid point by grid point basis, using observational data at the same resolution as the reference. Biases in spatial relationships, i.e., between different grid points, are



**Fig. 9** **a** Cross-validated LAR statistic with respect to WFDEI observations of annual maximum FWI for **a** raw CanRCM4 outputs. Differences in LAR between raw CanRCM4 outputs and **b** QDM, **c** MBCr, and **d** MBCn; negative values indicate better performance than CanRCM4



**Fig. 10** Area-weighted percentages of grid points exhibiting improved cross-validation performance with respect to CanRCM4 for bias corrected time series of July mean, annual maximum FWI, and monthly mean FWI



thus not considered. In this section, the ability of multivariate bias correction techniques to correct biases in spatial (and, as a side effect, temporal) dependence is assessed.

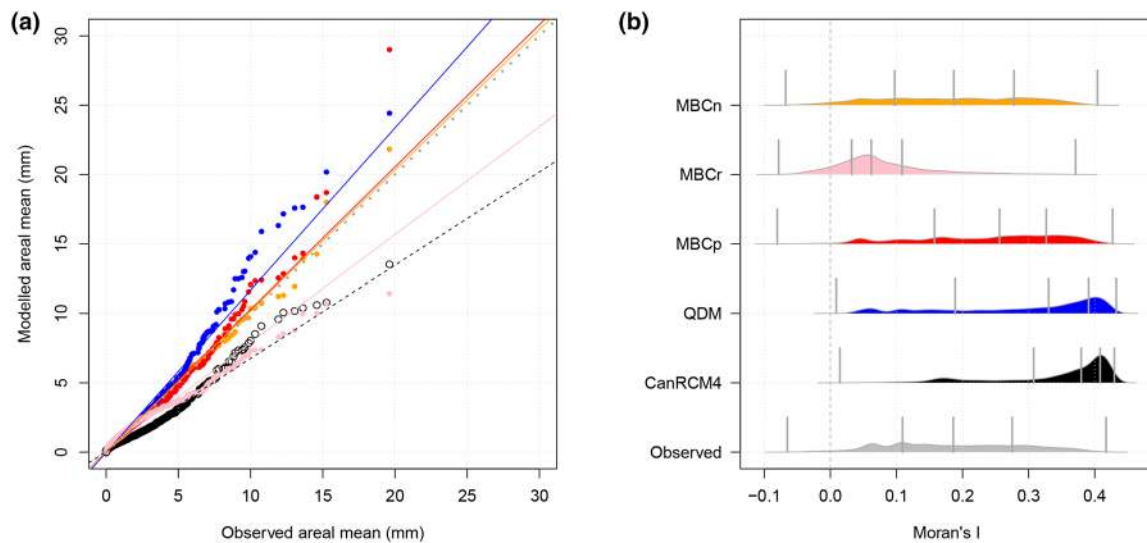
While bias correction algorithms are designed to be applied to climate model and observed fields at similar scales (Maraun 2013), algorithms have also been used to simultaneously bias correct and downscale from coarse climate model outputs to finer observed scales. For example, Abatzoglou and Brown (2012), Stoner et al. (2013), Ahmed et al. (2013) each applied quantile mapping to daily climate model data that had been interpolated to a high-resolution observational grid. However, Maraun (2013) and Gutmann et al. (2014) demonstrated that this approach—applying a univariate bias correction algorithm to interpolated climate data at individual grid points—can lead to fields with unrealistic spatial structure, especially if the variable being downscaled operates on spatial scales that are substantially finer than the climate model grid. This is particularly true for precipitation. The fine-scale bias corrected fields will inherit their spatial coherence from the coarser resolution climate model fields. For example, consider a situation where convective precipitation is triggered at a coarse-scale climate model grid point. When univariate quantile mapping is used to bias correct data that have been interpolated from the coarse-scale to fine-scale points within this grid cell, large precipitation amounts will be present over the grid cell in its entirety. In reality, observed convective precipitation would be spatially intermittent, only occurring over a fraction of the grid cell area. After univariate bias correction, marginal distributions at each fine-scale grid point will, by design, match observed distributions, but areal statistics, for example areal means or measures of spatial autocorrelation, will be biased.

One possible solution to this problem involves separation of the downscaling and bias correction operations, as done in the bias corrected constructed analogue with quantile mapping reordering (BCCAQ) algorithm by Werner and Cannon (2016), in which a constructed analogue downscaling algorithm (instead of simple interpolation) precedes the application of QDM at each grid point. Alternatively, with a multivariate bias correction algorithm, like MBCn, spatial points can be treated as additional variables. Then, by accounting for the multivariate dependence structure, joint correction of multiple grid points should lead to bias corrected fields with realistic spatial structure. In this case, a separate downscaling algorithm may be unnecessary. To test this hypothesis, the MBCn algorithm is used to simultaneously bias correct and downscale simulated precipitation fields for each meteorological season (DJF, MAM, JJA, and SON). Given the importance of correctly characterizing precipitation variability (e.g., in applications like hydrological modelling), as well as the overall difficulty in simulating spatial and temporal intermittency of

precipitation (Maraun et al. 2010), the focus here is exclusively on bias correction of 3-h precipitation outputs from CanRCM4.

Following Maraun (2013), sub-grid variability over a single coarse-scale grid cell is used for illustrative purposes. Observed WFDEI precipitation outputs are extracted over a  $5 \times 5$  window of grid points centred on  $96.75^\circ\text{W}$  and  $44.25^\circ\text{N}$ , which is close to the middle of the NAM-44i domain (near Lincoln, Nebraska). As the focus is on correction of spatial relationships (i.e., downscaling), CanRCM4 outputs over the region are degraded spatially so that they are representative of a downscaling rather than pure bias correction application. Following Dixon et al. (2016), simulated precipitation outputs are “coarsened” by spatially aggregating from the 0.5-deg NAM-44i grid to a coarser 2.5-deg grid and are then interpolated back onto the 0.5-deg grid. Spatial aggregation induces a fundamental mismatch in spatial scales between the model and observational fields. Furthermore, given the interrelation between spatial and temporal scales, this also leads to biases in temporal autocorrelation. In all cases, the first half of the record is used for calibration and the second half is used as for out-of-sample verification; all reported statistics are from the verification period. Results from MBCn are compared with those from univariate QDM (e.g., as in Abatzoglou and Brown 2012, Stoner et al. 2013; Ahmed et al. 2013), as well as MBCp, and MBCr. For the three multivariate algorithms, all 25 grid points are corrected simultaneously, whereas QDM is applied to each grid point in turn. Results are first illustrated for the summer season (JJA), when precipitation extremes are dominated by convective precipitation events with short length scales. Summary statistics for the remaining seasons are then tabulated for sake of completeness.

For summer, Fig. 11a shows quantile-quantile plots between observed areal mean precipitation amounts over the  $5 \times 5$  grid points and the corresponding areal means from CanRCM4, QDM, MBCp, MBCr, and MBCn. As expected, areally-averaged precipitation amounts during summer for the coarse-scale climate model (coarsened CanRCM4) are biased low relative to fine-scale observations, especially for the largest quantiles. Values for MBCr are similar to those from the coarsened CanRCM4 outputs; correcting Spearman correlation does not lead to improvements in sub-grid scale spatial variability. As illustrated by Maraun (2013), univariate quantile mapping (QDM) inflates areal precipitation amounts, leading to a large positive bias for the largest quantiles. This is consistent with the “convective precipitation” situation outlined above. MBCp, which corrects the Pearson correlation between grid points, leads to an improved distribution of areal means, but still shows a small positive bias. This is removed by MBCn, which exhibits the closest correspondence with observed values. Figure 11b shows distributions of



**Fig. 11** **a** Quantile-quantile plots between observed areal mean WFDEI precipitation amounts and areal mean (coarsened) CanRCM4 (black), QDM (blue), MBCp (red), MBCr (pink), and MBCn (orange) precipitation amounts. Best fit lines are shown for reference. **b** Distributions of Moran's  $I$  spatial autocorrelation coefficients for the WFDEI observations, interpolated CanRCM4 outputs, and QDM,

Moran's  $I$ , a measure of the spatial autocorrelation of a gridded field, over the series  $5 \times 5$  grid observed and modelled fields. Moran's  $I$  for a single field is given by

$$I = \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (10)$$

where  $N$  is the number of grid points,  $x$  is the variable of interest,  $\bar{x}$  is the mean of  $x$  over the grid, and  $w_{ij}$  is a binary indicator that characterizes neighbourhood structure. In this case, all grid points within the  $5 \times 5$  domain are considered to be neighbours. Values of  $I$  typically range from  $-1$  (negative autocorrelation) to  $+1$  (positive autocorrelation), with values near  $0$  indicating a random spatial pattern. Coarsened CanRCM4 outputs are more spatially coherent (median  $I = 0.38$ ) than the WFDEI observations (median  $I = 0.18$ ). Univariate QDM maintains this unrealistic spatial coherence (median  $I = 0.33$ ), which is consistent with results reported by Gutmann et al. (2014). Conversely, MBCr leads to precipitation fields that are less coherent than observed (median  $I = 0.07$ ). Both MBCp (median  $I = 0.26$ ) and MBCn (median  $I = 0.18$ ) provide a more realistic simulation of spatial dependence, but the overall correspondence is better for MBCn throughout the distribution. While biases in spatial dependence are strongest in summer, a similar pattern of results is evident in the other seasons.

Table 1 shows seasonal values of the Kolmogorov–Smirnov statistic, the maximum difference between

MBCp, MBCr, and MBCn bias correction algorithms. From left to right, vertical lines indicate 1st, 25th, 50th, 75th and 99th percentiles of each distribution. Time steps with no precipitation anywhere in the area are omitted. All statistics are based on data from summer seasons in the out-of-sample period

two CDFs, for each of the modelled distributions of  $I$  compared with the observed distribution. In three out of the four seasons, MBCn performs best and MBCp second best, with ranks reversed for autumn. Poorer performance by MBCn in autumn is likely due to sampling variability; in the calibration sample, MBCn outperforms MBCp by a similar amount—Kolmogorov–Smirnov statistic of  $0.06$  versus  $0.19$ —as MBCp outperforms MBCn in the verification period ( $0.07$  versus  $0.15$ ).

Spatial and temporal scales for a meteorological field like precipitation are intrinsically linked (Eskridge et al. 1997). Does correcting spatial coherence lead to an attendant improvement in temporal dependence? To measure discrepancies in temporal sequencing, values of autocorrelation at lags 1 to 8 (3- to 24-h) are calculated for each of the observed and modelled areal mean precipitation time series in each season. Values, summarized as mean

**Table 1** Kolmogorov–Smirnov statistics between observed and modelled distributions of Moran's  $I$  for each season in the out-of-sample verification period

Method/season	CanRCM4	QDM	MBCr	MBCp	MBCn
DJF	0.55	0.34	0.50	0.13	<u>0.07</u>
MAM	0.56	0.32	0.55	0.11	<u>0.07</u>
JJA	0.61	0.41	0.51	0.20	<u>0.08</u>
SON	0.48	0.30	0.57	<u>0.07</u>	0.15

The best performing method in each season is underlined

**Table 2** MAE between observed and modelled autocorrelation coefficients for areal mean precipitation series, averaged over lags 1 to 8, for each season in the out-of-sample verification period

Method/season	CanRCM4	QDM	MBCr	MBCp	MBCn
DJF	0.10	<u>0.06</u>	0.10	0.07	<u>0.06</u>
MAM	0.03	0.03	0.04	0.04	<u>0.02</u>
JJA	0.04	0.04	0.08	0.05	<u>0.03</u>
SON	<u>0.01</u>	0.02	0.02	0.04	0.02

The best performing method in each season is underlined

absolute errors (MAE) over all lags, are shown in Table 2. Aside from autumn, where the coarsened CanRCM4 series starts out by reproducing the observed autocorrelation function well, MBCn performs best in all seasons.

## 7 Discussion

While construction of the  $N$ -pdf algorithm by Pitié et al. (2005, 2007) was originally motivated by an  $N = 3$  dimensional image processing problem with  $I \gg N$ , the formal proof of convergence in the calibration sample is independent of  $N$ . In the examples given in Sects. 4 to 6, the number of cases  $I$  (i.e., pixels or time steps) is also substantially larger than the number of variables  $N$ . From a practical standpoint, applying the MBCn algorithm to higher dimensional problems leads to two potential problems. First, the number of iterations required to converge to the target distribution in the historical calibration period may be large, leading to unacceptably high computational cost. In general, speed of convergence will depend strongly on the characteristics of a given dataset and hence it is difficult to provide general estimates of computational demand. Second, if  $N$  is sufficiently large relative to  $I$ , then overfitting—fitting to noise in the historical calibration sample—is a very real possibility. In this case, convergence to the historical multivariate distribution may lead to spurious results in the projection period. The issue of overfitting has been raised for univariate quantile mapping, and is exacerbated when dealing with multivariate bias correction problems. As pointed out by IPCC (2015, pg. 22):

There is a trade-off between robustness and number of parameters in a [bias correction] BC method: the projections obtained from the BC data would be more credible when using simple methods (i.e., based on a parsimonious number of parameters). Results obtained using non-parametric BC methods such as quantile mapping often appear successful because of overfitting. However, when observed and simulated distributions are fundamentally different, such

BC methods may create overconfidence in the final results. 2D or higher-D corrections may do better at maintaining inter-variable links, but hinge on sufficient data availability to populate higher-dimensional histograms.

While these issues are not a specific focus of this paper, they are important and worthy of general discussion before final conclusions are given.

To address the first problem, speed of convergence, Pitié et al. (2005, 2007) suggest replacing the sequence of random orthogonal rotation matrices with a deterministic selection that instead maximizes the distance between rotation axis sets. As rotation axes are likely to be less correlated in higher dimensions, improvements gained through optimization of the rotation matrices will, however, be greatest for small  $N$ . For deterministic versus random rotations, Pitié et al. (2007) found average speed improvements of 2.15 and 1.5 times, respectively, for  $N = 2$  and  $N = 3$ . For larger  $N$ , this is unlikely to offer large gains in computational efficiency. However, because the  $N$ -pdf algorithm and, by extension, MBCn, relies on repeated applications of univariate quantile mapping, another option would be to use the most efficient form of quantile mapping during each iteration (e.g., based on interpolation between a small number of empirical quantiles). For MBCn, which applies a reordering operation after the last iteration (i.e., step d in Sect. 3), a more sophisticated univariate version of QDM could then be applied just once to get the final values per rank.

To help avoid overfitting for high dimensional problems (or more generally whenever  $N$  is large relative to  $I$ ), early stopping (Morgan and Bourlard 1990; Prechelt 1998)—an implicit form of regularization commonly used in iteratively-trained machine learning methods—may be effective for MBCn. Early stopping involves terminating calibration of a learning algorithm, in this case MBCn, prior to convergence on the calibration sample. Instead, performance is measured on a separate set of validation data and the algorithm is stopped when performance on this held-out sample, for example as measured using the energy distance (Eq. 4), is maximized. The objective thus is not to maximize calibration performance, but rather an estimate of the out-of-sample generalization performance. As a side benefit, by limiting the number of iterations, early stopping will also tend to reduce the overall computational burden of the MBCn algorithm.

## 8 Conclusion

MBCn, a modification of the  $N$ -pdf algorithm used in computer vision and image processing (Pitié et al. 2005, 2007),



is developed as a multivariate bias correction algorithm for climate model simulations of multiple variables. The result is a multivariate generalization of quantile mapping that transfers all statistical characteristics of an observed continuous multivariate distribution to the corresponding multivariate distribution of simulated variables. Unlike other multivariate bias correction algorithms, for example methods by Bürger et al. (2011), Vrac and Friederichs (2015), Mehrotra and Sharma (2016), and Cannon (2016), MBCn is not restricted to correcting a specified measure of joint dependence, such as Pearson or Spearman rank correlation, nor does it make strong stationarity assumptions about climate model temporal sequencing. The underlying  $N$ -pdf algorithm also has proven convergence properties (Pitié et al. 2007).

The method is first illustrated using an image recolouring example motivated by Pitié et al. (2005, 2007). When modified to mimic a traditional climate simulation experiment, image processing provides an effective way to visualize and gain both a qualitative feel for and quantitatively assess the performance of a multivariate bias correction algorithm. As shown in Fig. 3, one drawback of MBCn is its computational complexity, requiring several iterations to converge to the observed multivariate distribution. For image recolouring, a 3 dimensional dataset, it is  $\sim 20$ – $30$  times more expensive than quantile mapping, which is equivalent to  $\sim 2$ – $10$  times more expensive than the MBCp and MCBCr algorithms from Cannon (2016). Suggestions for improving the speed of convergence and, for higher dimension problems, avoiding overfitting, are discussed in Sect. 7.

Despite its higher computational cost, the MBCn method can be applied in a real-world climate model post-processing context. In this regard, MBCn is demonstrated by correcting biases in multiple climate variables from CanRCM4 over a North American domain and then calculating components of the Canadian Fire Weather Index (FWI) system. Results are compared with those from univariate quantile mapping and the MBCp and MBCr algorithms. Only MBCn is able to reproduce both observed annual maximum and July mean values of the FWI. Finally, MBCn is used to simultaneously bias correct and downscale CanRCM4 precipitation series. Overall, spatiotemporal statistics from MBCn match observed values more closely than the other methods. Hence, it may be feasible to use MBCn directly in downscaling applications, a practice that has been questioned for univariate quantile mapping (Maraun 2013; Gutmann et al. 2014).

One potential avenue for additional research is the explicit correction of multiple time scales. For example, Mehrotra and Sharma (2016) corrected lag 1 autocorrelation statistics at daily, monthly, quarterly, and annual time scales. With MBCn, one could decompose a given time

series into different time scales, for example using a multiresolution wavelet analysis or Kolmogorov–Zurbenko filtering (Eskridge et al. 1997), simultaneously bias correct the partitioned time series, and then reconstruct the original series. The basic approach can easily be extended to both space and time dimensions of multiple variables. This is left for future work.

Finally, an R package (R Core Team 2015) implementing the MBCn algorithm is available for download from <https://CRAN.R-project.org/package=MBC>.

**Acknowledgements** Images of paintings by Franklin Brownell (MIKAN No. 3929122) and William Redver Stark (MIKAN No. 3933129) are courtesy of Library and Archives Canada with the following terms of use: “Copyright expired. No restrictions on use for reproduction or publication”. Discussions with S. Lange about optimization speed greatly improved Sect. 7. Comments by M. Kirchmeier-Young and the anonymous reviewers are gratefully acknowledged.

**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

- Abatzoglou JT, Brown Timothy J (2012) A comparison of statistical downscaling methods suited for wildfire applications. *Int J Climatol* 32(5):772–780. doi:10.1002/joc.2312
- Ahmed KF, Wang G, Silander J, Wilson AM, Allen JM, Horton R, Anyah R (2013) Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the US northeast. *Glob Planet Change* 100:320–332. doi:10.1016/j.gloplacha.2012.11.003
- Amatulli G, Camia A, San-Miguel-Ayán J (2013) Estimating future burned areas under changing climate in the EU-Mediterranean countries. *Sci Total Environ* 450:209–222. doi:10.1016/j.scitotenv.2013.02.014
- Anderson KR, Englefield P, Carr R (2007) Predicting fire-weather severity using seasonal forecasts. In: Proceedings of the seventh symposium on fire and forest meteorology, 10.6.1–10.6.5. American Meteorological Society, Boston
- Bürger G, Schulla J, Werner AT (2011) Estimates of future flow, including extremes, of the Columbia River headwaters. *Water Resour Res* 47(10):1–18. doi:10.1029/2010WR009716
- Cannon AJ (2016) Multivariate bias correction of climate model output: matching marginal distributions and inter-variable dependence structure. *J Clim* 29(19):7045–7064. doi:10.1175/JCLI-D-15-0679.1
- Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of simulated precipitation by quantile mapping: how well do methods preserve relative changes in quantiles and extremes? *J Clim* 28(17):6938–6959. doi:10.1175/JCLI-D-14-00754.1
- Chen J, Brissette FP, Chaumont D, Braun M (2013) Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour Res* 49(7):4187–4205. doi:10.1002/wrcr.20331

- Chen J, Brissette FP, Lucas-Picher P (2015) Assessing the limits of bias-correcting climate model outputs for climate change impact studies. *J Geophys Res: Atmos* 120(3):1123–1136. doi:[10.1002/2014JD022635](https://doi.org/10.1002/2014JD022635)
- Christensen OB, Gutowski B, Nikulin G, Legutke S (2014) CORDEX Archive Design, version 3.1, 3 March 2014. [http://cordex.dmi.dk/joomla/images/CORDEX/cordex\\_archive\\_specifications.pdf](http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf)
- Dee DP, Uppala SM, Simmons AJ, Paul Berrisford P, Poli S, Kobayashi U, Andrae MA, Balmaseda G, Bauer Balsamo P (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828)
- Demidenko E (2006) Statistical comparison of color cancer cell images. *Oncol Rep* 15:1077–1079
- Dixon KW, Lanzante JR, Nath MJ, Hayhoe K, Stoner A, Radhakrishnan A, Balaji V, Gaitán CF (2016) Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Clim Change* 135(3–4):395–408. doi:[10.1007/s10584-016-1598-0](https://doi.org/10.1007/s10584-016-1598-0)
- Eden JM, Widmann M, Grawe D, Rast S (2012) Skill, correction, and downscaling of GCM-simulated precipitation. *J Clim* 25(11):3970–3984. doi:[10.1175/JCLI-D-11-00254.1](https://doi.org/10.1175/JCLI-D-11-00254.1)
- Ehret U, Zehe E, Wulfmeyer V, Warrach-Sagi K, Liebert J (2012) HESS opinions “Should we apply bias correction to global and regional climate model data?”. *Hydrol Earth Syst Sci* 16(9):3391–3404. doi:[10.5194/hess-16-3391-2012](https://doi.org/10.5194/hess-16-3391-2012)
- Eskridge RE, Ku JY, Rao ST, Porter PS, Zurbenko IG (1997) Separating different scales of motion in time series of meteorological variables. *Bull Am Meteorol Soc* 78(7):1473–1483
- Essou GRC, Sabarly F, Lucas-Picher P, Brissette F, Poulin A (2016) Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? *J Hydrometeorol*. doi:[10.1175/JHM-D-15-0138.1](https://doi.org/10.1175/JHM-D-15-0138.1)
- Flannigan MD, Krawchuk MA, de Groot WJ, Wotton BM, Gowman LM (2009) Implications of changing climate for global wildland fire. *Int J Wildland Fire* 18(5):483–507. doi:[10.1071/WF08187](https://doi.org/10.1071/WF08187)
- Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical note: downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods. *Hydrol Earth Syst Sci* 16(9):3383–3390. doi:[10.5194/hess-16-3383-2012](https://doi.org/10.5194/hess-16-3383-2012)
- Gutmann E, Pruitt T, Clark MP, Brekke L, Arnold JR, Raff DA, Rasmussen RM (2014) An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resour Res* 50(9):7167–7186. doi:[10.1002/2014WR015559](https://doi.org/10.1002/2014WR015559)
- Haddeland I, Heinke J, Voß F, Eisner S, Chen C, Hagemann S, Ludwig F (2012) Effects of climate model radiation, humidity and wind estimates on hydrological simulations. *Hydrol Earth Syst Sci* 16(2):305–318. doi:[10.5194/hess-16-305-2012](https://doi.org/10.5194/hess-16-305-2012)
- Hempel S, Frieler K, Warszawski L, Schewe J, Piontek F (2013) A trend-preserving bias correction—the ISI-MIP approach. *Earth Syst Dyn* 4(2):219–236. doi:[10.5194/esd-4-219-2013](https://doi.org/10.5194/esd-4-219-2013)
- IPCC (2015) IPCC workshop on regional climate projections and their use in impacts and risk analysis studies. In: Stocker TF, Dahe Q, Plattner G-K, Tignor M (eds) Workshop report, Bern, Switzerland. IPCC Working Group I Technical Support Unit, vol 171. ISBN: 978-92-9169-146-3
- Lehtonen I, Venäläinen A, Kämäräinen M, Peltola H, Gregow H (2016) Risk of large-scale fires in boreal forests of Finland under changing climate. *Nat Hazards Earth Syst Sci* 16(1):239–253. doi:[10.5194/nhess-16-239-2016](https://doi.org/10.5194/nhess-16-239-2016)
- Li H, Sheffield J, Wood EF (2010) Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J Geophys Res Atmos* 115(D10101):1–20. doi:[10.1029/2009JD012882](https://doi.org/10.1029/2009JD012882)
- Maraun D (2013) Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J Clim* 26(6):2137–2143. doi:[10.1175/JCLI-D-12-00821.1](https://doi.org/10.1175/JCLI-D-12-00821.1)
- Maraun D (2016) Bias correcting climate change simulations—a critical review. *Curr Clim Change Rep* 2(4):211–220. doi:[10.1007/s40641-016-0050-x](https://doi.org/10.1007/s40641-016-0050-x)
- Maraun D, Widmann M (2015) The representation of location by regional climate models in complex terrain. *Hydrol Earth Syst Sci Discuss* 12(3):3011–3028. doi:[10.5194/hessd-12-3011-2015](https://doi.org/10.5194/hessd-12-3011-2015)
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienens S, Rust HW, Sauter T, Themeßl M et al (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48(3):1–34. doi:[10.1029/2009RG000314](https://doi.org/10.1029/2009RG000314)
- Mehrotra R, Sharma A (2016) A multivariate quantile-matching bias correction approach with auto and cross dependence across multiple time scales: implications for downscaling. *J Clim* 29(10):3519–3539. doi:[10.1175/JCLI-D-15-0356.1](https://doi.org/10.1175/JCLI-D-15-0356.1)
- Mezzadri F (2007) How to generate random matrices from the classical compact groups. *Not Am Math Soc* 54:592–604
- Michelangeli PA, Vrac M, Loukos H (2009) Probabilistic downscaling approaches: application to wind cumulative distribution functions. *Geophys Res Lett* 36(11):1–6. doi:[10.1029/2009GL038401](https://doi.org/10.1029/2009GL038401)
- Morgan N, Bourlard H (1990) Generalization and parameter estimation in feedforward nets: some experiments. In: Touretzky DS (ed) *Advances in neural information processing systems*, vol 2. San Mateo, pp 630–637
- Olsson J, Berggren K, Olofsson M, Viklander M (2009) Applying climate model precipitation scenarios for urban hydrological assessment: a case study in Kalmar City, Sweden. *Atmos Res* 92(3):364–375. doi:[10.1016/j.atmosres.2009.01.015](https://doi.org/10.1016/j.atmosres.2009.01.015)
- Pappenberger F, Wetterhall F, Dutra E, Di Giuseppe F, Bogner K, Alfieri L, Cloke HL (2013) Seamless forecasting of extreme events on a global scale. In: *Climate and land surface changes in hydrology, proceedings of H01, IAHS-IAPSO-IASPEI assembly*, vol 359. IAHS Publication, Gothenburg, pp 3–10
- Piani C, Haerter JO, Coppola E (2010) 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](https://doi.org/10.1007/s00704-009-0134-9)
- Pitié F, Kokaram AC, Dahyot R (2005) N-dimensional probability density function transfer and its application to color transfer. In: *Tenth IEEE international conference on computer vision*, 2005. ICCV 2005, vol 2. IEEE, pp 1434–1439. doi:[10.1109/ICCV.2005.166](https://doi.org/10.1109/ICCV.2005.166)
- Pitié F, Kokaram AC, Dahyot R (2007) Automated colour grading using colour distribution transfer. *Comput Vis Image Underst* 107(1):123–137. doi:[10.1016/j.cviu.2006.11.011](https://doi.org/10.1016/j.cviu.2006.11.011)
- Prechelt L (1998) Automatic early stopping using cross validation: quantifying the criteria. *Neural Netw* 11(4):761–767. doi:[10.1016/S0893-6080\(98\)00010-0](https://doi.org/10.1016/S0893-6080(98)00010-0)
- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- Rizzo ML, Gábor Székely (2016) Energy distance. *WIREs Comput Stat* 8:27–38. doi:[10.1002/wics.1375](https://doi.org/10.1002/wics.1375)
- Rocheta E, Evans JP, Sharma A (2014) Assessing atmospheric bias correction for dynamical consistency using potential vorticity. *Environ Res Lett* 9(12):124010. doi:[10.1088/1748-9326/9/12/124010](https://doi.org/10.1088/1748-9326/9/12/124010)
- Rust HW, Kruschke T, Dobler A, Fischer M, Ulbrich U (2015) Discontinuous daily temperatures in the WATCH forcing datasets. *J Hydrometeorol* 16(1):465–472. doi:[10.1175/JHM-D-14-0123.1](https://doi.org/10.1175/JHM-D-14-0123.1)

- Scinocca JF, Kharin VV, Jiao Y, Qian MW, Lazare M, Solheim L, Flato GM, Biner S, Desgagne M, Dugas B (2016) Coordinated global and regional climate modeling. *J Clim* 29(1):17–35. doi:[10.1175/JCLI-D-15-0161.1](https://doi.org/10.1175/JCLI-D-15-0161.1)
- Silver NC, Dunlap WP (1987) Averaging correlation coefficients: should Fisher's  $z$  transformation be used? *J Appl Psychol* 72(1):146. doi:[10.1037/0021-9010.72.1.146](https://doi.org/10.1037/0021-9010.72.1.146)
- Stoner AMK, Hayhoe K, Yang X, Wuebbles DJ (2013) An asynchronous regional regression model for statistical downscaling of daily climate variables. *Int J Climatol* 33(11):2473–2494. doi:[10.1002/joc.3603](https://doi.org/10.1002/joc.3603)
- Székely GJ, Rizzo ML (2004) Testing for equal distributions in high dimension. *InterStat* 5:1–6. <http://interstat.statjournals.net/YEAR/2004/abstracts/0411005.php>
- Székely GJ, Rizzo ML (2013) Energy statistics: a class of statistics based on distances. *J Stat Plan Inference* 143(8):1249–1272. doi:[10.1016/j.jspi.2013.03.018](https://doi.org/10.1016/j.jspi.2013.03.018)
- Van Wagner CE, Forest P (1987) Development and structure of the Canadian forest fire weather index system. In: Forestry Technical Report 35. Canadian Forestry Service, Ottawa, ON, pp 1–37. <http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/19927.pdf>
- Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multi-scalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J Clim* 23(7):1696–1718. doi:[10.1175/2009JCLI2909.1](https://doi.org/10.1175/2009JCLI2909.1)
- Vrac M, Friederichs P (2015) Multivariate-intervariable, spatial, and temporal-bias correction. *J Clim* 28(1):218–237. doi:[10.1175/JCLI-D-14-00059.1](https://doi.org/10.1175/JCLI-D-14-00059.1)
- Vrac M, Noël T, Vautard R (2016) Bias correction of precipitation through singularity stochastic removal: because occurrences matter. *J Geophys Res Atmos*. doi:[10.1002/2015JD024511](https://doi.org/10.1002/2015JD024511)
- Wang L, Chen W (2014) Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. *Atmos Sci Lett* 15(1):1–6. doi:[10.1002/asl2.454](https://doi.org/10.1002/asl2.454)
- Wang X, Cantin A, Parisien MA, Wotton M, Anderson K, Flannigan M (2013) fwi.pbp: Fire weather index system and fire behaviour prediction system calculations. R package version 1.7. <http://CRAN.R-project.org/package=fwi.pbp>
- Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P (2014) The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour Res* 50(9):7505–7514. doi:[10.1002/2014WR015638](https://doi.org/10.1002/2014WR015638)
- Werner AT, Cannon AJ (2016) Hydrologic extremes-an intercomparison of multiple gridded statistical downscaling methods. *Hydrol Earth Syst Sci* 20(4):1483–1508. doi:[10.5194/hess-20-1483-2016](https://doi.org/10.5194/hess-20-1483-2016)
- Wilcke RAI, Mendlik T, Gobiet A (2013) Multi-variable error correction of regional climate models. *Clim Change* 120(4):871–887. doi:[10.1007/s10584-013-0845-x](https://doi.org/10.1007/s10584-013-0845-x)
- Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip Rev Clim Change* 2(6):851–870. doi:[10.1002/wcc.147](https://doi.org/10.1002/wcc.147)