Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence Structure

ALEX J. CANNON

Environment and Climate Change Canada, Victoria, British Columbia, Canada

(Manuscript received 25 September 2015, in final form 25 April 2016)

ABSTRACT

Univariate bias correction algorithms, such as quantile mapping, are used to address systematic biases in climate model output. Intervariable dependence structure (e.g., between different quantities like temperature and precipitation or between sites) is typically ignored, which can have an impact on subsequent calculations that depend on multiple climate variables. A novel multivariate bias correction (MBC) algorithm is introduced as a multidimensional analog of univariate quantile mapping. Two variants are presented. MBCp and MBCr respectively correct Pearson correlation and Spearman rank correlation dependence structure, with marginal distributions in both constrained to match observed distributions via quantile mapping. MBC is demonstrated on two case studies: 1) bivariate bias correction of monthly temperature and precipitation output from a large ensemble of climate models and 2) multivariate correction of vertical humidity and wind profiles, including subsequent calculation of vertically integrated water vapor transport and detection of atmospheric rivers. The energy distance is recommended as an omnibus measure of performance for model selection. As expected, substantial improvements in performance relative to quantile mapping are found in each case. For reference, characteristics of the MBC algorithm are compared against existing bivariate and multivariate bias correction techniques. MBC performs competitively and fills a role as a flexible, general purpose multivariate bias correction algorithm.

1. Introduction

Climate models represent, in a necessarily simplified manner, the physical processes that make up the coupled atmosphere, ocean, sea ice, land surface, and biogeochemical system. As the ability to efficiently represent these processes improves, so too does the fidelity of climate model simulations (Reichler and Kim 2008; Ramirez-Villegas et al. 2013; Bellenger et al. 2014). Despite continued improvements, however, systematic

biases between simulated climate variables and observations remain (Cattiaux et al. 2013; Sillmann et al. 2013; Kumar et al. 2014; Mueller and Seneviratne 2014). Put simply, models may be too hot or cold, wet or dry, etc., relative to observations. Use of uncorrected model output can have serious implications for some applications. For instance, projections of variables that depend on crossing absolute thresholds (e.g., growing season length, numbers of frost days, summer days, consecutive dry days, and heavy precipitation days; Zhang et al. 2011) are often used in climate change impacts and adaptation studies and to inform policy and planning decisions (Burton et al. 2004; Liu et al. 2014). For a variety of reasons—temporal and spatial discretization (Teutschbein and Seibert 2012), unresolved or unrepresented physical processes (Stevens and Bony 2013), etc.—it is likely that biases of sufficient magnitude to affect impacts and adaptation studies will remain in climate model output for the foreseeable future (Ramirez-Villegas et al. 2013).

Development and testing of empirical–statistical bias correction techniques for climate (similarly, numerical weather prediction) model output has thus become an active area of applied research (Teutschbein and Seibert

Denotes Open Access content.

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-15-0679.s1>.

Corresponding author address: Alex J. Cannon, Climate Research Division, Science and Technology Branch, Environment and Climate Change Canada, Canadian Centre for Climate Modelling and Analysis, P.O. Box 1700 STN CSC, Victoria BC V8W 2Y2, Canada.
E-mail: alex.cannon@canada.ca

DOI: 10.1175/JCLI-D-15-0679.1

2012; Verkade et al. 2013; Maraun et al. 2015). While it is beyond the scope of this paper to discuss in detail, it bears noting that recent studies have addressed the motivations, assumptions, and general utility of such techniques (Eden et al. 2012; Ehret et al. 2012; Maraun 2012; Hempel et al. 2013; Chen et al. 2015; Cannon et al. 2015). Recent reviews of multiple methods by Teutschbein and Seibert (2012), Gudmundsson et al. (2012), and Chen et al. (2013) have recommended the use of quantile mapping bias correction algorithms, which adjust the distribution of a modeled variable so that it matches that of observations.

Systematic errors resulting from subgrid-scale parameterizations and unresolved orography can, in principle, be corrected by bias correction algorithms like quantile mapping (Eden et al. 2012). In reality, the effectiveness of bias correction methods will be limited by large-scale circulation biases that cannot be corrected. For instance, Maraun and Widmann (2015) found that climate model simulations can systematically displace precipitation features (e.g., leeward rain shadows) so that quantile mapping applied to local pairs of observed and model grid cells may not be appropriate. Bias correction methods are thus not without drawbacks and caveats (Maraun 2013; Ehret et al. 2012); however, bias correction by quantile mapping is, in practice, a common component of climate change impact studies and is also used for postprocessing weather and climate forecasts.

In quantile mapping and related methods, the distributions in question have, with few exceptions, been univariate. In other words, climate model output for a given variable and location are bias corrected, separately, to match the distributional properties of the corresponding observational series. When applied in this manner, Wilcke et al. (2013) showed that univariate quantile mapping leads to very little change in the linear dependence structure between simulated variables. However, Bürger et al. (2011) suggested that “a general distortion of the multivariate covariance structure whose patterns are represented by the set of empirical orthogonal functions” may occur. Using isentropic potential vorticity as an example, Rocheta et al. (2014) showed the damaging effect that univariate bias correction methods, applied to each of multiple variables separately, can have on the dynamical consistency of the corrected atmospheric fields. Similarly, White and Toumi (2013), in the context of bias correction of global climate model (GCM) boundary conditions used to drive a regional climate model (RCM), concluded that quantile mapping “adds spurious variability to [originally] smooth spatial gradients.” More generally, ignoring the observed dependence structure has the potential to affect any climate index or impact model whose calculation depends on more than one variable

simultaneously—for instance, multivariate drought indices (Vicente-Serrano et al. 2010), fire weather indices (Wang et al. 2015), and ecological and hydrological models (Chun et al. 2014).

As a remedy, bias correction techniques that consider the multivariate dependence structure of data have begun to appear in the climate science literature. For instance, bivariate bias correction algorithms have been developed to jointly correct temperature and precipitation at a single climate model grid point. As one example, Zhang and Georgakakos (2012) constructed a nearest-neighbor mapping between the empirical joint cumulative distribution functions of observed and modeled temperature and precipitation series. As another example, the joint bias correction (JBC) algorithm developed by Li et al. (2014) assumes Gaussian and gamma distributions for temperature and precipitation, respectively, with the joint dependence modeled using a Gaussian copula (i.e., depending on the Pearson correlation dependence structure).

Moving beyond the bivariate case, truly multivariate approaches have been proposed by Bürger et al. (2011) and Vrac and Friederichs (2015). Bürger et al. (2011) used a multivariate extension of linear mean or variance rescaling, based on Cholesky decomposition of the covariance matrix, to bias correct GCM predictor fields used as predictors in a subsequent regression-based downscaling model. As in the JBC algorithm, the Pearson correlation dependence structure is forced to match observations. Because it linearly corrects the multivariate mean and covariance only, the method is designed for variables that follow a multivariate Gaussian distribution. A more general method, empirical copula–bias correction (EC–BC), was introduced by Vrac and Friederichs (2015) for correcting the multivariate structure of many climate model output, irrespective of their underlying distribution. EC–BC combines univariate bias correction (each climate model variable is bias corrected via quantile mapping) with empirical copula coupling or Schaake shuffle (Clark et al. 2004); values are then shuffled in time to match the original ordinal ranks of the observed dataset. The marginal distributions are thus identical to those from univariate quantile mapping, and temporal sequencing is drawn entirely from the observational dataset. Because it can only reproduce temporal patterns seen in historical series, EC–BC relies very strongly on a stationarity assumption; in practice, results are very similar to the quantile delta change method of Olsson et al. (2009), which superimposes projected changes in quantiles of a simulated variable over observed historical series. Time series characteristics of the driving climate model are sacrificed in favor of the observations. As

formulated, EC–BC is also limited to correction of model datasets that are the same length as the observational dataset.

By definition, all bias correction algorithms blend information from a climate model with historical observations. In a sense, the choice of algorithm sets the state of “knobs” that control whether marginal distributions, intervariable or spatial dependence structure, and temporal sequencing are informed more by the climate model or observations. As mentioned above, univariate quantile mapping forces marginal distributions to match observed data over a calibration period, with the other characteristics derived from the climate model. On the other hand, the EC–BC algorithm draws heavily on the historical observations for all of its characteristics, aside from the projected change signal.

In this study, a multivariate bias correction (MBC) technique is introduced to occupy the midground between these two extremes. The MBC algorithm is based on an iterative application of the Cholesky decomposition bias correction of [Bürger et al. \(2011\)](#) in concert with univariate quantile mapping. In effect, MBC’s control knob for temporal sequencing is set between that of quantile mapping (i.e., informed by the climate model) and EC–BC (i.e., informed by observations). Marginal distributions are corrected by quantile mapping, with either the Pearson correlation dependence structure (MBCp version of the algorithm) or the Spearman rank correlation dependence structure (MBCr version) further pushed toward observed values. Relative to EC–BC, MBC restores some of the time series characteristics of the driving climate model while being similarly flexible in terms of its applicability to high dimensional datasets.

Importantly, the EC–BC and MBC algorithms do not conserve the modeled temporal sequences but instead modify them. Thus, they break the temporal consistency between the simulated large-scale circulation and the local corrected sequences. This is a necessary consequence of correcting the multivariate dependence structure. In the case of EC–BC this modification is extreme, as it takes the observed weather sequence and simply replaces the observed quantiles with model quantiles. For MBC, the modification is less severe. The extent to which this is true, as well as potential impacts on derived quantities, will be demonstrated in the case studies that follow.

The two MBC algorithm variants, MBCp and MBCr, are demonstrated on two case studies. First, following [Li et al. \(2014\)](#), monthly mean temperature and precipitation output from phase 5 of the Coupled Model Intercomparison Project (CMIP5) multimodel ensemble ([Taylor et al. 2012](#)) are jointly bias corrected to observed Climatic Research Unit Time Series (CRU TS;

[Harris et al. 2014](#)). Results are compared with the JBC algorithm. Second, algorithms are used for multivariate bias correction of 6-h vertical profiles of specific humidity and horizontal wind components (18 variables at each grid point) from the Second Generation Canadian Earth System Model (CanESM2; [Arora et al. 2011](#)). In this case, evaluation of algorithm performance is based on subsequent calculations of vertically integrated water vapor transport (IVT) and detection of landfalling atmospheric rivers (ARs) from the IVT fields. Here, results are compared with the EC–BC algorithm.

2. Methods

This section describes the MBC algorithm, which is a bias correction methodology for climate model simulations that combines adjustment of marginal distributions via quantile mapping, including the ability to preserve projected changes in simulated quantiles, with adjustment of dependence structure via multivariate rescaling. Two variants are presented: MBCp corrects the Pearson dependence structure, and MBCr corrects the Spearman rank correlation structure. The energy distance is used to inform the selection of the most appropriate algorithm variant for a given application.

a. Univariate bias correction by quantile mapping

In univariate quantile mapping, bias correction is applied via a transfer function that links cumulative distribution functions (CDFs) $F_{o,h}$ and $F_{m,h}$ associated with, respectively, observed values $x_{o,h}$, denoted by the subscript o , and modeled values $x_{m,h}$, denoted by the subscript m , in a historical period, denoted by the subscript h . This leads to the following transfer function for correction of $x_{m,p}(t)$:

$$\hat{x}_{m,p}(t) = F_{o,h}^{-1}\{F_{m,h}[x_{m,p}(t)]\}, \quad (1)$$

with a modeled value at time t within some projection period, denoted by the subscript p . If applied to modeled data within the historical period (i.e., $m, p = m, h$), then the bias-corrected values will have the same distribution as the observed historical values.

The quantile mapping transfer function is constructed using information from the calibration samples exclusively. Recently, methods such as equidistant and equiratio quantile matching ([Li et al. 2010](#); [Wang and Chen 2014](#)) have been proposed that take advantage of information about the CDF of the simulated climate model variable $F_{m,p}$ in the future projection period. [Cannon et al. \(2015\)](#) demonstrated that these methods are equivalent to a quantile mapping form of the “delta change method” ([Olsson et al. 2009](#)), whereby future

changes in the simulated quantiles are preserved following bias correction. Because of this link, Cannon et al. (2015) refer to this approach as quantile delta mapping (QDM). The QDM transfer function for quantile mapping that preserves absolute changes in quantiles (e.g., as might be suitable for horizontal wind components and temperatures on the Celsius scale) is given by

$$\hat{x}_{m,p}(t) = F_{o,h}^{-1}\{F_{m,p}[x_{m,p}(t)]\} + x_{m,p}(t) - F_{m,h}^{-1}\{F_{m,p}[x_{m,p}(t)]\}, \quad (2)$$

with the corresponding transfer function that preserves relative changes in quantiles obtained simply by replacing the addition (subtraction) operators with multiplication (division) operators. The QDM approach to quantile mapping is applied in the remainder of the paper, and the term QDM will be used interchangeably with quantile mapping. While distributions in quantile mapping can be estimated using parametric, semiparametric, and nonparametric methods (Gudmundsson et al. 2012), nonparametric empirical CDFs are used here for the sake of simplicity and flexibility.

b. Multivariate linear bias correction by Cholesky decomposition

Bürger et al. (2011) presented a multivariate linear bias correction algorithm for climate model output. In brief, the method is a multivariate generalization of univariate mean/variance rescaling that replaces the standard deviation with the Cholesky decomposition of the covariance matrix. In univariate rescaling, bias correction within the historical period starts by expressing the model values $x_{m,h}$ as anomalies with respect to the mean $\bar{x}_{m,h}$:

$$x'_{m,h} = x_{m,h} - \bar{x}_{m,h}. \quad (3)$$

Then, rescaling by the ratio of the observed $\sigma_{o,h}$ and model $\sigma_{m,h}$ standard deviations (square root of the variances) gives the following:

$$\hat{x}'_{m,h} = \frac{x'_{m,h}\sigma_{o,h}}{\sigma_{m,h}}. \quad (4)$$

Finally, adding the observed mean $\bar{x}_{o,h}$,

$$\hat{x}_{m,h} = \hat{x}'_{m,h} + \bar{x}_{o,h}. \quad (5)$$

Following correction, the historical modeled mean and variance will match observed values. The model projection is linearly scaled/shifted using the same factors.

The multivariate rescaling method proceeds in the same way but with multivariate generalizations of the mean and standard deviation replacing the univariate statistics. Given a matrix of observed data $\mathbf{X}_{o,h}$ (with variables in columns) and a corresponding matrix of climate model output $\mathbf{X}_{m,h}$, start, as in the univariate case, by calculating anomalies $\mathbf{X}'_{o,h}$ and $\mathbf{X}'_{m,h}$ with respect to the multivariate mean (i.e., columns of $\mathbf{X}_{o,h}$ and $\mathbf{X}_{m,h}$ are centered by subtracting matrices containing the multivariate means $\bar{\mathbf{X}}_{o,h}$ and $\bar{\mathbf{X}}_{m,h}$). Assuming that the covariance matrix of the observed anomalies is positive definite, it can be decomposed into the product of the upper triangular matrix $\mathbf{L}_{o,h}$ and its transpose via Cholesky decomposition¹:

$$\text{cov}(\mathbf{X}'_{o,h}) = \mathbf{L}_{o,h}^T \mathbf{L}_{o,h}. \quad (6)$$

The covariance matrix of the modeled anomalies can be decomposed in similar fashion:

$$\text{cov}(\mathbf{X}'_{m,h}) = \mathbf{L}_{m,h}^T \mathbf{L}_{m,h}. \quad (7)$$

The climate model anomalies can then be corrected so that they have the same covariance structure (i.e., variances and Pearson correlations) as the observations via the following:

$$\hat{\mathbf{X}}'_{m,h} = \mathbf{X}'_{m,h} \mathbf{L}_{m,h}^{-1} \mathbf{L}_{o,h}. \quad (8)$$

This is the multivariate analog of Eq. (4). As data are expressed as anomalies, values can be bias corrected to have the same mean as the observed data by adding back the observed multivariate mean [i.e., as in Eq. (5)]:

$$\hat{\mathbf{X}}_{m,h} = \hat{\mathbf{X}}'_{m,h} + \bar{\mathbf{X}}_{o,h}. \quad (9)$$

For the climate model projections $\mathbf{X}_{m,p}$, bias correction proceeds via the following:

$$\hat{\mathbf{X}}'_{m,p} = \mathbf{X}'_{m,p} \mathbf{L}_{m,h}^{-1} \mathbf{L}_{o,h} \quad \text{and} \quad (10)$$

$$\hat{\mathbf{X}}_{m,p} = \hat{\mathbf{X}}'_{m,p} + (\bar{\mathbf{X}}_{m,p} - \bar{\mathbf{X}}_{m,h}) + \bar{\mathbf{X}}_{o,h}, \quad (11)$$

where $\mathbf{X}'_{m,p}$ are modeled anomalies calculated with respect to the projected multivariate mean $\bar{\mathbf{X}}_{m,p}$ and the second term in Eq. (11) is the mean climate change signal; that is, following Bürger et al. (2011) modeled anomalies are not

¹ Under some circumstances (e.g., if variables from neighboring locations are perfectly correlated) the covariance matrices may not be positive definite. In this case, smoothing of nonpositive definite covariance matrices (e.g., following Higham 1988; Knol and ten Berge 1989) is needed before the Cholesky decomposition.

taken with respect to the historical period to avoid mixing the climate change signal into the correction.

Because it is based on a linear transformation that is specified in terms of the mean and covariance, the method is guaranteed to fully correct data that follow a multivariate Gaussian distribution. In [Bürger et al. \(2011\)](#), marginal distributions were first transformed to be univariate Gaussian via probit transformations prior to multivariate rescaling. While multivariate Gaussianity ensures that marginal distributions will be univariate Gaussian, the reverse need not be true; having the marginal distributions be Gaussian does not guarantee that the multivariate distribution will be Gaussian (e.g., [Dutta and Genton 2014](#)). Hence, it is not known a priori how closely the method will correct the full multivariate distribution in cases where data are not multivariate Gaussian. An illustration of this point is provided as supplemental material; comparisons between the multivariate linear bias correction of [Bürger et al. \(2011\)](#) and the MBCp and MBCr algorithms are given for three simple bivariate distributions.

c. MBC

Quantile mapping, as outlined in [section 2a](#), operates on the marginal distributions of a multivariate dataset, but does not attempt to correct the dependence structure. Conversely, the multivariate linear bias correction algorithm from [section 2b](#) corrects the dependence structure but, aside from the case of multivariate Gaussianity, does not fully correct the marginal distributions. Intuitively, one might expect that some combination of the two would lead to a more complete multivariate bias correction algorithm. Based on this intuition, MBC, a multivariate bias correction algorithm that combines quantile mapping and multivariate linear rescaling, is presented here. Two variants are possible: 1) MBCp, where the Pearson correlation dependence structure is pushed toward observations, and 2) MBCr, where the dependence structure is instead based on the Spearman rank correlation. In either case, marginal distributions are constrained to match observations via univariate quantile mapping. Both MBCp and MBCr are iterative in nature, successively applying multivariate linear rescaling to match Pearson or rank correlation dependence structure and univariate transformations to match marginal distributions.

The two algorithm variants are described below, and an experimental implementation in the R programming language ([R Core Team 2015](#)) is available on request.

1) MBCp

The MBCp algorithm variant is applied as follows:

- (i) Following [section 2a](#), apply univariate quantile mapping to each column of $\mathbf{X}_{m,h}$ and $\mathbf{X}_{m,p}$, using $\mathbf{X}_{o,h}$ as

observed data, to yield initial matrices of corrected output $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$.

- (ii) Following [section 2b](#), apply multivariate bias correction to the current $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$ matrices, using $\mathbf{X}_{o,h}$ as observed data, and replace the old $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$ matrices with the corrected output.
- (iii) Following [section 2a](#), apply univariate quantile mapping to each column of the current $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$ matrices, using $\mathbf{X}_{o,h}$ as observed data, to yield new $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$ matrices. If elements of the Pearson correlation matrix $\text{cor}(\hat{\mathbf{X}}_{m,h})$ are within a specified absolute tolerance of those in $\text{cor}(\mathbf{X}_{o,h})$ [see Eq. (12)] or if the rate of convergence has fallen below a specified threshold, continue from step (iv). Otherwise, repeat from step (ii).
- (iv) Return $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$ matrices from the last iteration of steps (ii) and (iii) as the final corrected output.

Depending on convergence, step (iv) may yield projected values that are slightly different than those from quantile mapping. If a strict correspondence is needed for the projected period, quantiles of the final $\hat{\mathbf{X}}_{m,p}$ matrix can be replaced with those from step (i).

Step (ii) corrects the multivariate Pearson correlation dependence structure but may modify the marginals; step (iii) restores the observed marginal distributions but may modify the Pearson correlation dependence structure. Iteratively alternating between these two steps pushes both Pearson correlation dependence structure and marginal distributions toward observed values. Convergence and rate of convergence are measured in terms of the mean absolute error between off-diagonal elements of the corrected and observed correlation matrices:

$$\text{MAE}_{\text{cor}} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k |\text{cor}(\hat{\mathbf{X}}_{m,h})_{ij} - \text{cor}(\mathbf{X}_{o,h})_{ij}|, \quad (12)$$

where k is the number of variables. In all cases reported in the remainder of the paper, convergence is said to occur when MAE_{cor} or its magnitude of change between subsequent iterations falls below 0.0001.

The iterative nature of the algorithm is motivated purely by heuristic arguments. In the case of multivariate Gaussianity, the multivariate algorithm of [Bürger et al. \(2011\)](#) guarantees perfect convergence without recourse to multiple iterations. However, this does not extend to the case of marginal Gaussianity in the absence of multivariate Gaussianity or to non-Gaussian quantities. In these cases, while a single iteration may lead to a better match between the multivariate distributions, further corrections can result in additional

improvements (see the supplemental material). While no theoretical guarantees of convergence or the uniqueness of the resulting solutions can be offered at this time, empirical results find no instances in which the algorithm diverges or stalls well before reaching the observed correlation structure. In all cases, the corrected Pearson correlation dependence structure is a closer match to observed values than if quantile mapping were to be applied to the marginal distributions without considering the multivariate distribution. Empirical evidence of convergence is addressed in more detail in the case studies presented in [section 3](#).

2) MBCr

MBCp relies on the Pearson correlation, which fully specifies intervariable relationships for a multivariate Gaussian distribution, to measure dependence. As argued by [Genest and Favre \(2007\)](#) for copula modeling, however, “inference concerning dependence structures should always be based on ranks.” This naturally leads to the Spearman rank correlation, which is simply the Pearson correlation applied to variables that have been expressed in terms of their ordinal ranks. The MBCr algorithm variant is a rank-based version of the MBCp algorithm. It proceeds as follows:

- (i) Replace elements of each column of the $\mathbf{X}_{m,h}$, $\mathbf{X}_{m,p}$, and $\mathbf{X}_{o,h}$ matrices by their ordinal ranks² to yield $\mathbf{R}_{m,h}$, $\mathbf{R}_{m,p}$, and $\mathbf{R}_{o,h}$ matrices.
- (ii) Following [section 2b](#), apply multivariate bias correction to the current $\mathbf{R}_{m,h}$ and $\mathbf{R}_{m,p}$ matrices, using $\mathbf{R}_{o,h}$ as observed data, and save the corrected output as $\hat{\mathbf{R}}_{m,h}$ and $\hat{\mathbf{R}}_{m,p}$.
- (iii) Rerank elements of each column of the current $\hat{\mathbf{R}}_{m,h}$ and $\hat{\mathbf{R}}_{m,p}$ matrices to yield new matrices of ordinal ranks $\mathbf{R}_{m,h}$ and $\mathbf{R}_{m,p}$. If elements of the Spearman rank correlation matrix $\text{cor}(\mathbf{R}_{m,h})$ are within a specified absolute tolerance of those in $\text{cor}(\mathbf{R}_{o,h})$ [see Eq. (13)] or if the rate of convergence has fallen below a specified threshold, continue from step (iv). Otherwise, repeat from step (ii).
- (iv) Following [section 2a](#), apply univariate quantile mapping to each column of $\mathbf{X}_{m,h}$ and $\mathbf{X}_{m,p}$, using $\mathbf{X}_{o,h}$ as observed data, to yield $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$.
- (v) Order elements of each column of $\hat{\mathbf{X}}_{m,h}$ and $\hat{\mathbf{X}}_{m,p}$ from step (iv) according to the ordinal ranks $\mathbf{R}_{m,h}$ and $\mathbf{R}_{m,p}$ from the last iteration of steps (ii) and (iii) and return the final corrected results.

For MBCr, convergence is measured in terms of the following:

$$\text{MAE}_{\text{cor}} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k |\text{cor}(\mathbf{R}_{m,h})_{i,j} - \text{cor}(\mathbf{R}_{o,h})_{i,j}|. \quad (13)$$

With minor modifications, steps (i), (ii), (iv), and (v) are similar to the algorithm presented by [Iman and Conover \(1982\)](#) in the context of multivariate stochastic simulation. Following MBCp, however, MBCr differs from [Iman and Conover \(1982\)](#) by iteratively alternating between the multivariate transformation in step (ii) and the univariate transformations in step (iii). Empirically, closer convergence to the desired rank correlation structure is obtained using this iterative procedure rather than a single-step algorithm, especially when working in high dimensions.

Both MBCp and MBCr operate under the assumption that quantiles and covariance matrices of both observations and climate model output are given exactly. Estimated values are, of course, subject to uncertainty. While not considered here in detail, the bootstrap can be used to provide uncertainty bands—representing sampling uncertainty alone—for the bias-corrected output. It bears noting that they do not, however, provide uncertainties in the sense of the usefulness of the bias correction methods themselves. Bootstrap-based estimates of model performance statistics are presented in [section 3](#).

Given that quantile mapping, MBCp, MBCr, and EC-BC produce bias-corrected datasets that, for all intents and purposes, share the same marginal distributions, how do they differ? As mentioned in [section 1](#), the ordering of the underlying data—the temporal sequencing—distinguishes the methods from one another. Quantile mapping takes its temporal sequencing directly from the climate model (i.e., rank correlations with the climate model series will equal one). EC-BC shuffles the quantile mapping output to match the rank ordering of the observations (i.e., rank correlations with the observed series will equal one). MBCp and MBCr simply reorder the quantile mapping output so that the resulting Pearson or Spearman dependence structure is closer to observations; rank correlations with the climate model series will be less than one, but only to a limited extent, as required for the algorithm to move toward the observed dependence structure. A quantitative analysis is provided in [section 3b](#).

Except under very limited circumstances, matching the marginal distributions and either Pearson or Spearman correlation structure of a dataset, as done by MBC, will not be sufficient to reproduce all the characteristics of a given multivariate distribution. As noted by [Iman and Conover \(1982\)](#), it is, however, “usually not

² Here, ties are broken such that ranks increase in time.

possible to obtain more rigid specifications.” The extent to which MBC is successful as a multivariate bias correction algorithm, based on these constraints, is evaluated on the case studies presented in the next section.

d. Energy distance

Because there are two algorithm variants, some means of quantitatively assessing the relative performance of MBCp and MBCr is needed for model selection purposes. Here, the energy distance, a measure of the statistical discrepancy between two (potentially high dimensional) multivariate distributions (Székely and Rizzo 2004; Baringhaus and Franz 2004; Rizzo and Székely 2016), is adopted. Given k -dimensional independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ with CDFs F and G , respectively, the squared energy distance D between the two distributions is given by

$$D^2(F, G) = 2E\|\mathbf{X} - \mathbf{Y}\| - E\|\mathbf{X} - \mathbf{X}'\| - E\|\mathbf{Y} - \mathbf{Y}'\| \geq 0, \quad (14)$$

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 $\mathbf{Y}$. The energy distance is sensitive to deviations in marginal distributions and multivariate dependence structure and is used as a proper scoring rule for multivariate probabilistic forecasts (Gneiting et al. 2008). Vrac and Friederichs (2015) evaluated the performance of EC-BC using the closely related integrated quadratic distance.

For independent samples from two k -dimensional datasets, the energy distance can be expressed via the following statistic:

$$\begin{aligned} D^2(\mathbf{X}, \mathbf{Y}) = & \frac{2}{N_x N_y} \sum_{i=1}^{N_x} \sum_{m=1}^{N_y} \|\mathbf{x}(i) - \mathbf{y}(m)\| \\ & - \frac{1}{N_x^2} \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} \|\mathbf{x}(i) - \mathbf{x}(j)\| \\ & - \frac{1}{N_y^2} \sum_{l=1}^{N_y} \sum_{m=1}^{N_y} \|\mathbf{y}(l) - \mathbf{y}(m)\|, \end{aligned} \quad (15)$$

where $\mathbf{x}(i)$ is the k -dimensional vector corresponding to the i th of N_x samples in $\mathbf{X}$ (similarly, N_y samples of $\mathbf{Y}$). If the two multivariate samples are drawn from the same distribution, then the statistic goes to zero. Because the energy distance is based on the Euclidean norm, each of the k variables should be measured on comparable scales—for example, standardized to zero mean and unit standard deviation if they are originally on different units of measure.

When comparing methods, values can instead be reported in terms of the energy distance skill score:

$$\text{ESS} = 1 - \frac{D^2(\mathbf{X}, \mathbf{Y})}{D^2(\mathbf{X}, \mathbf{Y}_{\text{ref}})}, \quad (16)$$

where $\mathbf{X}$ is the observational dataset, $\mathbf{Y}$ is the dataset under evaluation, and $\mathbf{Y}_{\text{ref}}$ is a baseline reference dataset; values of ESS can range from $-\infty$ to $+1$ (no discrepancy in distributions of the observed and evaluation data), with values greater than zero indicating improvement over the reference dataset.

3. Case studies

a. Monthly temperature and precipitation

Methods are first demonstrated on the simple bivariate bias correction problem from Li et al. (2014). In this case, monthly temperature and precipitation series from GCMs are jointly corrected to observations. Monthly observations of temperature and precipitation over global land areas from the CRU TS version 3.22 dataset are extracted for the years 1951–2005. The corresponding historical simulations from 43 GCMs (see Table S1 in the supplemental material) are obtained from the CMIP5 archive; output are from the first available run. As in Li et al. (2014), both CRU and GCM data are remapped onto a common 1° grid using bilinear interpolation.

The data record is split into two chunks: 1951–80 is used to calibrate the bias correction algorithms, and 1981–2005 is used as the out-of-sample period to evaluate performance. Monthly output from each GCM, stratified by climatological season, are bias corrected at each grid point to CRU observations via univariate QDM and both variants of the MBC algorithm. The goal is to bias correct the climate model so that observed relationships between temperature and precipitation are reproduced—for example, if hot and dry, cold and wet, etc. conditions are observed to occur together, then they should also occur together in the bias-corrected output. To illustrate, Fig. 1 shows maps of the observed Pearson correlation between temperature and precipitation at each grid point over winter months in the evaluation period; results for HadCM3 simulations, QDM bias corrections, and MBCr bias corrections are shown for comparison. In this example, HadCM3 overestimates the strength of the mostly positive observed relationship between temperature and precipitation over Eurasia. QDM, which is applied separately to temperature and precipitation at each grid point, replicates the erroneous simulated relationship, whereas MBCr successfully restores the more muted joint dependence seen in observations; results for MBCp are similar (not shown).

FIG. 1. Maps showing Pearson correlation coefficients between temperature and precipitation at each grid point in winter months (1981–2005) for (a) CRU TS version 3.22 observations, (b) HadCM3, (c) HadCM3 bias corrected using univariate QDM algorithm, and (d) HadCM3 bias corrected using multivariate MBCr algorithm.

Do the same performance characteristics hold over the full CMIP5 ensemble and all seasons? For the sake of consistency with Li et al. (2014), results for the evaluation period are summarized in terms of root-mean-square error (RMSE), calculated over all grid points, between the observed and modeled Pearson temperature and precipitation correlation (i.e., a global summary of results in Fig. 1), as well as mean of precipitation, precipitation standard deviation, mean of

temperature, and temperature standard deviation. For completeness, RMSE values for the Spearman rank correlation are also reported. Box-and-whisker plots showing the distribution of RMSE values over the 43-member CMIP5 ensemble are given in Fig. 2 for each season. While more recent versions of the CRU TS observational dataset (version 3.22 vs 3.1) and CMIP5 ensemble (43 vs 38 GCMs) are used here, preventing strict quantitative comparisons between this study and Li et al. (2014), where possible general characteristics of the JBC algorithm (cf. Fig. 3 in Li et al. 2014) are compared with those of MBCp and MBCr. As the QDM, MBCp, and MBCr methods share the same bias-corrected marginal distributions, RMSE values for the temperature and precipitation means and standard deviations are identical. Quantile mapping corrects all quantiles of the distribution, and so large improvements relative to the uncorrected CMIP5 GCMs are noted in all cases. Remnant errors can mostly be attributed to observed and simulated natural variability between the 1951–80 and 1981–2005 calibration and evaluation periods—what are referred to as “variability related apparent bias changes” by Maraun (2012). Median improvements relative to the GCM found here are roughly consistent with those reported by Li et al. (2014) for the JBC algorithm.

Looking next to corrections of the dependence structure, RMSE values for the Pearson correlation coefficients typically fall in the range 0.15–0.2 for MBCp and MBCr versus 0.3–0.4 for the uncorrected CMIP5 GCMs and univariate QDM output. Bearing in mind the differences in experimental setup, note that values reported by Li et al. (2014) for JBC typically fall in the range 0.2–0.25. Comparing RMSE values for Pearson and Spearman correlations between the MBCp and MBCr algorithms, better performance is, as expected, noted when the form of the corrected dependence structure matches the type of correlation used to measure performance. Relative to MBCp, however, MBCr exhibits more stable results, with less of a difference noted between its Pearson and Spearman RMSE values. In general, the low RMSE values for the two correlation measures are very similar, which implies that the joint dependence structure between temperature and precipitation is well represented by either measure [e.g., as might be the case if, as reported by Li et al. (2014), monthly temperature and precipitation do not deviate markedly from a bivariate Gaussian distribution]. Qualitatively, note that results from the JBC algorithm depend on the order of correction—whether temperature is bias corrected prior to precipitation or vice versa—with performance statistics for the variable that is corrected first typically being better than those for the

FIG. 2. Box-and-whisker plots of RMSE between the observed and modeled (a) Pearson (or Spearman, as marked by vertical dotted lines and S at the bottom of the panels) correlation between temperature and precipitation, (b) mean of precipitation, (c) precipitation standard deviation, (d) mean of temperature, and (e) temperature standard deviation for (from left to right) summer (JJA), fall (SON), winter (DJF), and spring (MAM) seasons for the 43-member CMIP5 ensemble (1981–2005). Within each panel, the x-axis label GCM (gray boxes) marks values for the uncorrected climate model output, QDM (blue boxes) marks quantile mapping, MBCp (red boxes) marks MBC with correction of Pearson correlation structure, and MBCr (pink boxes) marks MBC with correction of Spearman correlation structure. The horizontal line within each box shows the median RMSE, the box extends from the lower to upper quartile, and whiskers further extend to the farthest data points within 1.5 times the interquartile range. To allow direct comparisons, the box-and-whisker plot specifications follow those of [Li et al. \(2014\)](#).

FIG. 3. Scatterplots showing winter temperature and precipitation data at a grid point centered over the west coast of Alaska (59.5°N, 160°W): (top left) CRU observations, (top center) HadCM3, (top right) QDM, (middle) first three iterations of the MBCp algorithm, and (bottom) first three iterations of the MBCr algorithm. Values reported in the bottom right corner are the Pearson (r) and Spearman (s) correlations between temperature and precipitation in the 1951–80 calibration period. Lighter (darker) shaded symbols correspond to calibration (evaluation) samples.

second. MBCp and MBCr are not subject to this limitation. Overall, the two MBC algorithm variants adjust the marginal distributions in monthly temperature and precipitation to match observations while simultaneously correcting the Pearson or Spearman correlation between variables.

Because this case study involves joint correction of just two variables, scatterplots of temperature and

precipitation can be used to visually assess rates of convergence of MBCp and MBCr and their ability to reproduce the observed bivariate dependence structure. As an example, Fig. 3 compares observed, HadCM3, and bias-corrected winter temperature and precipitation data at a grid point centered over the west coast of Alaska (59.5°N, 160°W). For MBCp and MBCr, results are shown following the first three bias correction

iterations. At this location, HadCM3 simulates a relatively strong positive relationship between temperature and precipitation. The same relationship is much weaker in observations. QDM carries with it the rank ordering of the HadCM3 series and thus replicates its strong temperature–precipitation coupling. Both MBCp and MBCr more faithfully reproduce the observed dependence structure, even after a single algorithm iteration. Subsequent iterations improve the correspondence between bias-corrected and observed correlations, converging to very small absolute errors (≤ 0.01) after a few passes (Fig. 4); similar rates of convergence are found over the remainder of the GCMs and domain (not shown). Visually, however, changes to the scatterplots are almost imperceptible after the first iteration. Note, also, that differences in the character of the joint distribution are still present between MBCp, MBCr, and observations, despite convergence to observed values of the Pearson and Spearman correlation coefficients. Again, as stated in section 2c, Pearson and Spearman correlations are not sufficient to fully specify the joint dependence structure of most real-world datasets. With that said, improvements are certainly evident and results from both MBCp and MBCr are qualitatively and quantitatively better than those obtained from separate application of the QDM algorithm. In the next section, the MBC algorithm is applied to large multivariate datasets at subdaily time scales to see if this holds in higher dimensions and for less Gaussian variables.

b. The 6-h humidity and wind profiles, integrated vapor transport, and atmospheric river detection

Next, methods are evaluated on a larger multivariate bias correction task, in this case one designed to be representative of studies where GCM output are used as lateral boundary conditions for subsequent dynamical downscaling models (Colette et al. 2012; Rocheta et al. 2014). Vertical profiles of 6-h specific humidity and horizontal wind components from the historical CanESM2 simulation (Arora et al. 2011) are corrected, simultaneously, using vertical profiles from the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim; Dee et al. 2011) as the observational reference. ERA-Interim data are obtained from the ECMWF on a 2.5° grid. Output from two CanESM2 historical simulations (runs r1i1p1 and r5i1p1) on a spectral T63 ($\sim 2.8^\circ$) grid are interpolated onto the 2.5° grid using a first-order conservative remapping algorithm (Jones 1999). Reanalysis and GCM data are extracted for the 1981–2005 period.

In the spirit of Rocheta et al. (2014), who evaluated univariate bias corrections applied to wind, potential temperature, and pressure fields in terms of biases in

FIG. 4. Rates of convergence of the MBCp and MBCr algorithms for bias correction of winter temperature and precipitation data at a grid point centered over the west coast of Alaska (59.5°N , 160°W). Absolute errors with respect to observations for Pearson and Spearman correlations between temperature and precipitation are presented for MBCp and MBCr, respectively. Iteration 0 corresponds to results from the univariate QDM algorithm applied to temperature and precipitation separately.

isentropic potential vorticity, here evaluation of algorithm performance is based on subsequent calculations of IVT. As IVT is commonly used in the detection of ARs, which are relatively narrow (300–500 km) and long (around 2000 km) bands of concentrated moisture located within the warm conveyor belt of extratropical cyclones (Neiman et al. 2008), counts of AR events identified using a standard detection algorithm are used as a secondary means of assessing bias correction performance. IVT and AR counts depend, in a complicated manner, on multiple variables and their dependencies. Thus, it is expected that a successful multivariate bias correction algorithm should outperform a univariate algorithm that fails to account for the observed multivariate dependence structure.

ARs are responsible for a significant fraction of poleward atmospheric moisture transport (Zhu and Newell 1998) and can trigger heavy precipitation and flooding events when they make landfall (Lavers et al. 2013; Rutz et al. 2014; Radić et al. 2015). The AR component of this case study focuses on landfalling ARs along the west coast of North America, which tend to be most active between November and January (Rutz et al. 2014). Hence, IVT is calculated for data within this block of three months. IVT is defined here as follows:

$$\text{IVT} = g^{-1} \int_{1000\text{hPa}}^{300\text{hPa}} q \mathbf{v} dp, \quad (17)$$

where g is gravitational acceleration (m s^{-2}), q is specific humidity (kg kg^{-1}), the vector $\mathbf{v}$ represents total horizontal wind (with scalar zonal and meridional components u and v in m s^{-1}), and p is pressure (hPa). Data are defined on 1000-, 850-, 700-, 500-, 400-, and 300-hPa pressure levels, which is the same set of vertical levels (up to 500 hPa) used in the CMIP5 analysis by [Lavers et al. \(2015\)](#). Thus, 18 variables are required to calculate IVT at a given grid point. Once IVT has been calculated, ARs are identified from the IVT field using the detection algorithm described by [Rutz et al. \(2014\)](#) and [Radić et al. \(2015\)](#). If IVT at an initial grid point exceeds a $250 \text{ kg m}^{-1} \text{ s}^{-1}$ threshold, then the grid point is retained and the search moves to the adjacent upstream point with the highest IVT. If IVT at the upstream point exceeds the threshold, it is retained and the search continues; the algorithm terminates when all adjacent points fall below the threshold. An AR is detected at the initial point if the distance between the end point and initial point exceeds 2000 km.

The univariate QDM bias correction algorithm is applied, separately, to q , u , and v on the six pressure levels at each 2.5° grid point from CanESM2 for November–January. The multivariate MBCp and MBCr bias correction algorithms are applied, simultaneously, to all variables and pressure levels at each grid point. As this case study involves a relatively short observational dataset, two historical simulations from CanESM2 are bias corrected and compared with ERA-Interim data. The CanESM2 r1i1p1 run serves as the calibration sample and the r5i1p1 run as an out-of-sample evaluation dataset. In both cases, the same ERA-Interim data are used as the observational reference. However, following [Maraun \(2013\)](#) and [White and Toumi \(2013\)](#), the main focus is, by design, on the in-sample characteristics of the calibration dataset, as this allows differences in performance to be more directly attributed to characteristics of the underlying bias correction algorithms. In this case, differences between the QDM and the MBCp and MBCr algorithms are dictated by correction (or not) of the dependence structure. The r5i1p1 evaluation run serves as an added check on the robustness of the results in the presence of natural climate variability. Later it is used to demonstrate the consequences of the “time shuffling” aspect of the MBC algorithm.

Before discussing bias correction performance, a summary of convergence properties of the MBCp and MBCr algorithms is presented in [Fig. 5](#). Despite operating on a large number of variables, convergence still occurs relatively quickly for both MBCp and MBCr

FIG. 5. Box-and-whisker plots showing the number of iterations required for the MBCp and MBCr algorithm variants to converge when applied to vertical profiles of q , u , and v (18 variables) at each 2.5° grid point over the global domain. Convergence for MBCp (MBCr) occurs when the mean absolute change in off-diagonal elements of the Pearson (Spearman) correlation matrix falls below 0.0001. Note: for MBCr, the lower quartile and median number of iterations are equal.

algorithm variants. In the bivariate temperature and precipitation example, data were known to closely follow a joint Gaussian distribution, and little difference between the MBCp and MBCr algorithm variants was expected; no such expectations are carried forward here. More generally, the use of multiple bias correction algorithms on 18 variables over a global domain makes summarizing and comparing algorithm performance a difficult task. Hence, an omnibus measure of performance, the energy distance ([section 2d](#)), is used to gain a general, first impression of the data and to guide algorithm selection and subsequent reporting of results.

In this case, the energy distance is used to assess the multivariate discrepancy between the distributions of observed and CanESM2, QDM, MBCp, and MBCr IVT series over the global domain (i.e., at all grid points simultaneously). IVT values at each grid point are weighted according to the square root of the cosine of latitude prior to calculating the energy distance. For ease of comparison, values are reported in terms of the energy distance skill score (ESS). ESS results, including a 95% bootstrap sampling interval based on a seasonal block bootstrap, are shown in [Fig. 6](#) using uncorrected CanESM2 IVT output as the reference dataset. In addition, ESS values for MBCp and MBCr are reported using QDM as the reference to show improvements gained by

using a multivariate rather than univariate bias correction algorithm. The seasonal block bootstrap is conducted by sampling contiguous blocks of November–January data with replacement from the global IVT fields, calculating the ESS on the bootstrapped dataset and repeating 200 times to construct a bootstrap sampling distribution.

The bias correction algorithms lead to large, statistically significant improvements in performance relative to the uncorrected CanESM2 IVT field. All values of ESS exceed 0.9 and none of the 95% sampling intervals overlap zero. Differences in ESS between the r1i1p1 and r5i1p1 runs are much smaller than the overall magnitude of the bias correction improvement, which suggests that the GCM bias is large relative to natural variability. Using QDM as the reference in the ESS reveals that the two MBC algorithm variants provide substantial—ESS values exceed 0.35—and significant improvements relative to the univariate algorithm. The majority of the improvement in ESS relative to CanESM2 is thus due to the correction of the marginal distributions by QDM, with correction of the dependence structure by MBC leading to smaller but still significant increases in ESS. Last, ESS values for MBCp and MBCr are nearly the same, although sampling intervals are slightly narrower for MBCr. For the sake of brevity, subsequent figures will thus focus on MBCr results.

To help put the ESS values in context, Fig. 7 provides a global map of the long-term mean ERA-Interim IVT field, along with percentage biases of the uncorrected CanESM2 and bias-corrected QDM and MBCr IVT fields. For CanESM2, large biases are found over a substantial fraction of the globe. The global median (lower and upper quartile) absolute bias for CanESM2 is 27% (17% and 37%). Values over areas of the tropical Pacific Ocean are overestimated by more than +75%, with negative biases mostly confined to the extratropics. After correction of the q , u , and v fields by QDM and MBCr, the resultant absolute biases in IVT are reduced substantially. Median (lower and upper quartile) is just 2% (0.9% and 4.5%) and 0.8% (0.3% and 1.4%) for QDM and MBCr, respectively. Note, however, that the residual QDM bias field is spatially coherent, with negative anomalies primarily confined to the tropics and positive anomalies in the middle and high latitudes—roughly opposite to the pattern of CanESM2 biases—whereas the MBCr bias field does not contain noticeable spatial structure. The spatial correlation between the CanESM2 and QDM bias field is -0.43 versus -0.1 between CanESM2 and MBCr.

As described above, AR events are detected by identifying long, contiguous regions with high IVT values. Failure to correct biases in IVT at areas well upstream of the location of interest can thus have an

FIG. 6. Values of the ESS for global IVT using CanESM2 as the reference (6 points along the top of the plot with ESS > 0.8) and using QDM as the reference (4 points along the bottom of the plot with ESS < 0.4); x-axis labels show the dataset being evaluated. Error bars show 95% block bootstrap sampling intervals. Entries marked r5i1p1 correspond to values calculated using the CanESM2 r1i1p1 run as the calibration dataset and the CanESM2 r5i1p1 run as the evaluation dataset. Other entries are within-sample estimates based on the CanESM2 r1i1p1 run.

effect on the frequency of detected ARs. To illustrate, Fig. 8 shows the percentage of 6-h time steps in which an AR was detected at each grid point over a domain covering western North America, along with percentage biases for CanESM2, QDM, and MBCr. Consistent with Rutz et al. (2014), in the ERA-Interim climatology ARs are found over coastal land grid points approximately 10%–15% of November–January (NDJ) analysis times, with inland penetration occurring approximately 5%–10% of the time. The frequency of occurrence off the coast reaches 35%. Consistent with the pattern of IVT bias in Fig. 7b, CanESM2 underestimates the frequency of ARs over the ocean by -20% to -60% , with a few positive anomalies over British Columbia ($>+20\%$). Over the domain, the median (lower and upper quartile) is -20% (-38% and -12%). On the other hand, QDM overestimates the frequency of ARs over the coast and inland, with biases exceeding $+20\%$ over the majority of land grid points. Over the domain as a whole, the median (lower and upper quartile) QDM bias is $+9\%$ ($+5\%$ and $+15\%$). By comparison, MBCr is almost unbiased over both ocean and land grid points, exhibiting a domain median (lower and upper quartile) bias of $+4\%$ ($+2\%$ and $+8\%$).

FIG. 7. Global maps of (a) long-term mean ERA-Interim IVT ($\text{kg m}^{-1} \text{s}^{-1}$) over the 1981–2005 period; percentage bias in IVT between (b) CanESM2 and ERA-Interim, (c) QDM and ERA-Interim, and (d) MBCr and ERA-Interim. Bias values for (b) are capped at $\pm 75\%$ for display purposes.

While an in-depth comparison with the EC-BC method is not the focus of this case study, some general comments can be made about its characteristics with respect to those of the MBC algorithm. Both are designed as multivariate alternatives to univariate quantile mapping but differ in how this is achieved. The EC-BC algorithm of [Vrac and Friederichs \(2015\)](#) operates by shuffling quantile mapping output in time such that the ordinal ranks are the same as those from the observed series (i.e., using the Schaake shuffle or empirical copula coupling). Within the calibration sample, EC-BC thus reproduces the observed series identically and values of the ESS, bias, derived quantities, etc. will be perfect relative to observations. No trace of the climate model remains. Outside of the calibration dataset, the ordinal ranks are again taken from the observations, which, in essence, means that the climate model only provides information about changes in the marginal distribution. All other characteristics are drawn from the observed series. This precludes model-projected changes in temporal sequencing from being expressed in the corrected output. In MBC, quantile mapping output are shuffled to reproduce the observed correlation

structure, but the corrected series still retain much of the underlying climate model's temporal sequencing.

To illustrate these differences, EC-BC is applied to vertical profiles of q , u , and v at an inland grid point on the south coast of British Columbia, Canada (50°N , 120°W), a location with large positive biases in detected AR counts in the raw CanESM2 and corrected QDM output (Figs. 8b,c). Performance statistics across the 18 variables are summarized in Fig. 9. Mean absolute errors (with respect to ERA-Interim) in off-diagonal elements of the Pearson (Fig. 9a) and Spearman (Fig. 9b) intervariable correlation matrices are presented for the CanESM2 calibration (r1i1p1) and evaluation (r5i1p1) simulations. Both MBCr and EC-BC reduce absolute errors in the simulated Pearson and Spearman dependence structure of the CanESM2 output. For EC-BC, which mimics the observed series perfectly in the calibration period and, aside from small deviations due to minor differences in marginal distributions in the two runs, almost perfectly in the evaluation dataset, errors are negligible. As expected, MBCr performs well in terms of the Spearman correlation dependence structure but has higher errors for the intervariable

FIG. 8. (a) Percentage of time steps with an AR detected at each grid point in the ERA-Interim NDJ observations. Percentage bias between frequency of ARs detected in (b) CanESM2 and ERA-Interim, (c) QDM and ERA-Interim, and (d) MBCr and ERA-Interim.

Pearson correlations. Again, as noted in sections 2c and 3a, the two measures of correlation can usually only imperfectly characterize the full multivariate dependence structure of a dataset.

To assess differences in the raw and corrected marginal distributions relative to the ERA-Interim observations, the Kolmogorov–Smirnov (KS) test was applied, for each method in the evaluation dataset, to the 18 variables in turn. Distributions of the 18 KS D statistics—the maximum differences between the observed and modeled empirical CDFs—are shown in

Fig. 9c. The three bias correction algorithms share marginal distributions from QDM and therefore exhibit the same, minimal deviations from the observed distributions. Temporal sequencing is where differences between the QDM, MBCr, and EC-BC series occur. Figure 9d shows Spearman rank correlations between each simulated CanESM2 time series and the corresponding QDM, MBCr, and EC-BC series. Quantile mapping is a monotonic transformation and so the rank correlation between the QDM and CanESM2 series equals one; temporal sequencing is taken entirely from

FIG. 9. Mean absolute errors (with respect to ERA-Interim) in off-diagonal elements of the (a) Pearson and (b) Spearman intervariable correlation matrices for calibration (r1i1p1) and evaluation (r5i1p1) CanESM2 simulations and bias-corrected output at an inland grid point on the south coast of British Columbia, Canada (50°N, 120°W). (c) Box-and-whisker plots of the KS D statistic (with respect to ERA-Interim) for the evaluation dataset. The plot shows the distribution of the 18 D statistics (smaller values are better). (d) Box-and-whisker plots of Spearman correlation coefficients between each of the 18 simulated CanESM2 time series and the corresponding bias-corrected series in the evaluation dataset.

the climate model. Conversely, EC-BC takes its temporal sequencing from the historical observations, and its rank correlation with the CanESM2 simulated values is centered on zero. In order for MBCr to reproduce the observed Spearman rank correlation dependence structure, it shuffles the QDM series. The resulting temporal rank correlation with CanESM2 must also be less than one. In this case, however, the degree of change to the underlying CanESM2 series is minimal; the mean rank correlation for MBCr is 0.89. Of course, if larger corrections to the dependence structure were necessary

(e.g., at a grid point with a bigger mismatch between model and observations), larger changes to the GCM time structure would be imposed by the MBC algorithm; however, inspection of grid points along the west coast of North America find mean values in the range 0.89–0.99 (not shown). Regardless, corrections to the dependence structure cannot occur without some effect on the rank ordering.

How much does this shuffling affect temporal characteristics of the derived IVT and AR series? To illustrate, Fig. 10 shows raw and corrected CanESM2 r5i1p1

FIG. 10. (a) Raw CanESM2 r5i1p1, QDM-corrected, and MBCr-corrected time series of mean DJF IVT anomalies (1981–2005) at the 50°N, 120°W grid point; (b) corresponding 6-h IVT time series during the NDJ 2004/05 period.

time series of mean DJF IVT anomalies at the 50°N, 120°W grid point, along with 6-h IVT values during the NDJ 2004/05 period. At both time intervals (interannual and 6 h), MBCr series are highly correlated (Spearman correlation of 0.89 and 0.83, respectively, and Pearson correlation of 0.95 and 0.77, respectively) with the raw CanESM2 simulation values. Shuffling modifies the underlying climate model simulation time series, but the changes are relatively modest, especially at the interannual time scale. As mentioned earlier, the time series characteristics of the EC–BC correction are derived entirely from the ERA-Interim series (not shown); model-simulated changes in interannual variability thus cannot be analyzed using the EC–BC method.

Results shown in Figs. 9 and 10 depend only on the q , u , and v vertical profiles at one grid point. Given that ARs depend on upstream IVT conditions, persistence of an AR event depends on both spatial and temporal characteristics of the data. Hence, changes to either will influence AR event durations. Figure 11 shows the percentage of NDJ analysis times associated with events lasting different durations (6 h, 12 h, etc., up to >48 h), as well as the mean AR event duration. Rutz et al. (2014) found the mean lifetime of an AR event over the western United States to range from 12–18 h at inland locations to 23–24 h on the Oregon coast. Based on ERA-Interim, values here are consistent with these results. At the 50°N, 120°W grid point, the mean duration

FIG. 11. Stacked bar plots showing the percentage of NDJ analysis times associated with AR events at 50°N, 120°W lasting different durations for ERA-Interim observations, CanESM2 r5i1p1 simulations, QDM-corrected output, and MBCr-corrected output. Mean AR event durations are listed in the x-axis labels.

of an AR event is 17 h. In the CanESM2 evaluation dataset, the mean simulated duration is 18 h, falling to 17 and 11 h after correction by QDM and MBCr, respectively. ARs in CanESM2 and QDM are detected >35% more often than observed at this location, but their mean lifetime is consistent with observations. On the other hand, the overall frequency of events is better captured after correction by MBCr, but the longevity of events is underestimated. This underestimation is primarily due to an overestimation of 6- and 12-h events at the expense of events greater than 48 h.

4. Discussion and conclusions

A multivariate analog of quantile mapping, the MBC algorithm, is introduced as a bias correction technique for climate model output. In addition to marginal distributions, which are corrected via univariate quantile mapping, either Pearson correlation (in the MBCp variant) or Spearman rank correlation (in the MBCr variant) dependence structure is also corrected to more closely match observed values. MBC is demonstrated on two case studies: bivariate bias correction of monthly temperature and precipitation series and multivariate correction of vertical humidity and wind profiles (e.g., as required to calculate IVT and detect AR events). For reference, characteristics of the MBC algorithm are compared with the bivariate JBC algorithm of Li et al.

(2014) and the multivariate EC–BC algorithm of Vrac and Friederichs (2015).

Substantial improvements in performance relative to univariate quantile mapping, which does not take into account dependence structure, are found in each case. For the bivariate dataset, MBC results are comparable to those from the JBC algorithm but have the advantage of not depending on the order in which variables are corrected. For the second case study, MBC removes residual biases in IVT and frequencies of detected ARs that remain following correction by quantile mapping. Specific characteristics of the MBC algorithm are compared with those of the EC–BC algorithm. By definition, both MBC and EC–BC must modify the temporal sequencing of the underlying climate model to correct the multivariate dependence structure. Relative to the EC–BC algorithm, which mimics the temporal rank ordering of the observations exactly, MBC better preserves the temporal sequencing of the underlying climate model output. Over western North America, mean Spearman rank correlations between raw CanESM2 q , u , and v time series and MBCr corrected series range from 0.89 at the 50°N, 120°W grid point to 0.99, whereas values for EC–BC are centered around 0. For MBC, derived IVT time series at the 50°N, 120°W grid point are also affected, but again the degree of change with respect to CanESM2 is modest at both the 6-h (rank correlation of 0.83) and interannual (rank correlation of 0.89) time scales. However, AR event durations, which depend on both local and upstream time series, are underestimated in the MBCr output. Such trade-offs must be weighed according to the needs of a given application.

Closer adherence to some characteristics of the historical series (e.g., the climatological seasonal cycle) can be achieved in a straightforward manner by applying MBC to sliding windows around each calendar day of the year (Thrasher et al. 2012; Themeßl et al. 2012). As pointed out by Gennaretti et al. (2015), this also avoids artificial climatological discontinuities at transitions between months. While not explored in this study, operating on augmented datasets consisting of lagged copies of the data, in effect forcing MBC to correct both intervariable and lag correlation dependencies, might also help. Similarly, MBC can easily operate across spatial locations.

Aside from idealized cases (e.g., a multivariate Gaussian distribution), the Pearson and Spearman rank correlations are not guaranteed to fully specify the dependence structure of a multivariate distribution. In practice, one or the other may better describe a given dataset. For the temperature and precipitation and IVT case studies, MBCp and MBCr perform similarly well. More generally, it will not be known ahead of time which

of the two dependence measures will lead to the best performance. The energy distance (Székely and Rizzo 2004; Baringhaus and Franz 2004; Székely and Rizzo 2013) provides a succinct summary of multivariate bias correction performance and is recommended for selecting between the MBCp and MBCr algorithm variants. Overall, MBC, with its ability to correct either form of dependence structure, offers the climate community a flexible, general purpose multivariate bias correction algorithm.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and thank the climate modeling groups listed in Table S1 of the supplemental material for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Valuable comments by Y. Dibikey, M. Hofer, G. Li, and three reviewers are gratefully acknowledged. Research was conducted by the government of Canada with no external funding sources.

REFERENCES

- Arora, V., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38**, L05805, doi:10.1029/2010GL046270.
- Baringhaus, L., and C. Franz, 2004: On a new multivariate two-sample test. *J. Multivar. Anal.*, **88**, 190–206, doi:10.1016/S0047-259X(03)00079-4.
- Bellenger, H., É. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, doi:10.1007/s00382-013-1783-z.
- Bürger, G., J. Schulla, and A. Werner, 2011: Estimates of future flow, including extremes, of the Columbia River headwaters. *Water Resour. Res.*, **47**, W10520, doi:10.1029/2010WR009716.
- Burton, I., E. Malone, and S. Huq, 2004: *Adaptation Policy Frameworks for Climate Change: Developing Strategies, Policies and Measures*. Cambridge University Press, 258 pp.
- Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of simulated precipitation by quantile mapping: How well do methods preserve relative changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, doi:10.1175/JCLI-D-14-00754.1.
- Cattiaux, J., H. Douville, and Y. Peings, 2013: European temperatures in CMIP5: Origins of present-day biases and future uncertainties. *Climate Dyn.*, **41**, 2889–2907, doi:10.1007/s00382-013-1731-y.
- Chen, J., F. P. Brissette, D. Chaumont, and M. Braun, 2013: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.*, **49**, 4187–4205, doi:10.1002/wrcr.20331.
- , —, and P. Lucas-Picher, 2015: Assessing the limits of bias-correcting climate model outputs for climate change impact

- studies. *J. Geophys. Res. Atmos.*, **120**, 1123–1136, doi:[10.1002/2014JD022635](https://doi.org/10.1002/2014JD022635).
- Chun, K., H. Wheeler, and A. Barr, 2014: A multivariate comparison of the BERMS flux-tower climate observations and Canadian Coupled Global Climate Model (CGCM3) outputs. *J. Hydrol.*, **519**, 1537–1550, doi:[10.1016/j.jhydrol.2014.08.059](https://doi.org/10.1016/j.jhydrol.2014.08.059).
- Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.*, **5**, 243–262, doi:[10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2).
- Colette, A., R. Vautard, and M. Vrac, 2012: Regional climate downscaling with prior statistical correction of the global climate forcing. *Geophys. Res. Lett.*, **39**, L13707, doi:[10.1029/2012GL052258](https://doi.org/10.1029/2012GL052258).
- Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828).
- Dutta, S., and M. G. Genton, 2014: A non-Gaussian multivariate distribution with all lower-dimensional Gaussians and related families. *J. Multivar. Anal.*, **132**, 82–93, doi:[10.1016/j.jmva.2014.07.007](https://doi.org/10.1016/j.jmva.2014.07.007).
- Eden, J. M., M. Widmann, D. Grawe, and S. Rast, 2012: Skill, correction, and downscaling of GCM-simulated precipitation. *J. Climate*, **25**, 3970–3984, doi:[10.1175/JCLI-D-11-00254.1](https://doi.org/10.1175/JCLI-D-11-00254.1).
- Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, 2012: HESS opinions: “Should we apply bias correction to global and regional climate model data?” *Hydrol. Earth Syst. Sci.*, **16**, 3391–3404, doi:[10.5194/hess-16-3391-2012](https://doi.org/10.5194/hess-16-3391-2012).
- Genest, C., and A.-C. Favre, 2007: Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, **12**, 347–368, doi:[10.1061/\(ASCE\)1084-0699\(2007\)12:4\(347\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)).
- Gennaretti, F., L. Sangelantoni, and P. Grenier, 2015: Toward daily climate scenarios for Canadian arctic coastal zones with more realistic temperature–precipitation interdependence. *J. Geophys. Res. Atmos.*, **120**, 11 862–11 877, doi:[10.1002/2015JD023890](https://doi.org/10.1002/2015JD023890).
- Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, **17**, 211–235, doi:[10.1007/s11749-008-0114-x](https://doi.org/10.1007/s11749-008-0114-x).
- Gudmundsson, L., J. Bremnes, J. Haugen, and T. Engen-Skaugen, 2012: Technical note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, doi:[10.5194/hess-16-3383-2012](https://doi.org/10.5194/hess-16-3383-2012).
- Harris, I., P. Jones, T. Osborn, and D. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, doi:[10.1002/joc.3711](https://doi.org/10.1002/joc.3711).
- Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek, 2013: A trend-preserving bias correction—The ISI-MIP approach. *Earth Syst. Dyn.*, **4**, 219–236, doi:[10.5194/esd-4-219-2013](https://doi.org/10.5194/esd-4-219-2013).
- Higham, N. J., 1988: Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.*, **103**, 103–118, doi:[10.1016/0024-3795\(88\)90223-6](https://doi.org/10.1016/0024-3795(88)90223-6).
- Iman, R. L., and W. Conover, 1982: A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. Simul. Comput.*, **11**, 311–334, doi:[10.1080/03610918208812265](https://doi.org/10.1080/03610918208812265).
- Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, doi:[10.1175/1520-0493\(1999\)127<2204:FASOCR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2).
- Knol, D. L., and J. M. ten Berge, 1989: Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika*, **54**, 53–61, doi:[10.1007/BF02294448](https://doi.org/10.1007/BF02294448).
- Kumar, D., E. Kodra, and A. R. Ganguly, 2014: Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation. *Climate Dyn.*, **43**, 2491–2518, doi:[10.1007/s00382-014-2070-3](https://doi.org/10.1007/s00382-014-2070-3).
- Lavers, D. A., R. P. Allan, G. Villarini, B. Lloyd-Hughes, D. J. Brayshaw, and A. J. Wade, 2013: Future changes in atmospheric rivers and their implications for winter flooding in Britain. *Environ. Res. Lett.*, **8**, 034010, doi:[10.1088/1748-9326/8/3/034010](https://doi.org/10.1088/1748-9326/8/3/034010).
- , F. M. Ralph, D. E. Waliser, A. Gershunov, and M. D. Dettinger, 2015: Climate change intensification of horizontal water vapor transport in CMIP5. *Geophys. Res. Lett.*, **42**, 5617–5625, doi:[10.1002/2015GL064672](https://doi.org/10.1002/2015GL064672).
- Li, C., E. Sinha, D. E. Horton, N. S. Diffenbaugh, and A. M. Michalak, 2014: Joint bias correction of temperature and precipitation in climate model simulations. *J. Geophys. Res. Atmos.*, **119**, 13 153–13 162, doi:[10.1002/2014JD022514](https://doi.org/10.1002/2014JD022514).
- 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](https://doi.org/10.1029/2009JD012882).
- Liu, M., and Coauthors, 2014: What is the importance of climate model bias when projecting the impacts of climate change on land surface processes? *Biogeosciences*, **11**, 2601–2622, doi:[10.5194/bg-11-2601-2014](https://doi.org/10.5194/bg-11-2601-2014).
- Maraun, D., 2012: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophys. Res. Lett.*, **39**, L06706, doi:[10.1029/2012GL051210](https://doi.org/10.1029/2012GL051210).
- , 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, doi:[10.1175/JCLI-D-12-00821.1](https://doi.org/10.1175/JCLI-D-12-00821.1).
- , and M. Widmann, 2015: The representation of location by regional climate models in complex terrain. *Hydrol. Earth Syst. Sci.*, **19**, 3449–3456, doi:[10.5194/hess-19-3449-2015](https://doi.org/10.5194/hess-19-3449-2015).
- , and Coauthors, 2015: VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, **3**, 1–14, doi:[10.1002/2014EF000259](https://doi.org/10.1002/2014EF000259).
- Mueller, B., and S. Seneviratne, 2014: Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 128–134, doi:[10.1002/2013GL058055](https://doi.org/10.1002/2013GL058055).
- Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West Coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeorol.*, **9**, 22–47, doi:[10.1175/2007JHM855.1](https://doi.org/10.1175/2007JHM855.1).
- Olsson, J., K. Berggren, M. Olofsson, and M. Viklander, 2009: Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden. *Atmos. Res.*, **92**, 364–375, doi:[10.1016/j.atmosres.2009.01.015](https://doi.org/10.1016/j.atmosres.2009.01.015).
- Radić, V., A. J. Cannon, B. Menounos, and C. Gi, 2015: Future changes in autumn atmospheric river events in British Columbia, Canada, as projected by CMIP5 global climate models. *J. Geophys. Res. Atmos.*, **120**, 9279–9302, doi:[10.1002/2015JD023279](https://doi.org/10.1002/2015JD023279).
- Ramirez-Villegas, J., A. J. Challinor, P. K. Thornton, and A. Jarvis, 2013: Implications of regional improvement in global climate models for agricultural impact research. *Environ. Res. Lett.*, **8**, 024018, doi:[10.1088/1748-9326/8/2/024018](https://doi.org/10.1088/1748-9326/8/2/024018).

- R Core Team, 2015: R: A language and environment for statistical computing. R Foundation for Statistical Computing. [Available online at <https://www.R-project.org/>.]
- Reichler, T., and J. Kim, 2008: How well do coupled models simulate today's climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311, doi:[10.1175/BAMS-89-3-303](https://doi.org/10.1175/BAMS-89-3-303).
- Rizzo, M. L., and G. J. Székely, 2016: Energy distance. *Wiley Interdiscip. Rev.: Comput. Stat.*, **8**, 27–38, doi:[10.1002/wics.1375](https://doi.org/10.1002/wics.1375).
- Rocheta, E., J. P. Evans, and A. Sharma, 2014: Assessing atmospheric bias correction for dynamical consistency using potential vorticity. *Environ. Res. Lett.*, **9**, 124010, doi:[10.1088/1748-9326/9/12/124010](https://doi.org/10.1088/1748-9326/9/12/124010).
- Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Mon. Wea. Rev.*, **142**, 905–921, doi:[10.1175/MWR-D-13-00168.1](https://doi.org/10.1175/MWR-D-13-00168.1).
- Sillmann, J., V. Kharin, X. Zhang, F. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, doi:[10.1002/jgrd.50203](https://doi.org/10.1002/jgrd.50203).
- Stevens, B., and S. Bony, 2013: What are climate models missing? *Science*, **340**, 1053–1054, doi:[10.1126/science.1237554](https://doi.org/10.1126/science.1237554).
- Székely, G. J., and M. L. Rizzo, 2004: Testing for equal distributions in high dimension. *InterStat*, **5**, 1–6.
- , and —, 2013: Energy statistics: A class of statistics based on distances. *J. Stat. Plan. Inference*, **143**, 1249–1272, doi:[10.1016/j.jspi.2013.03.018](https://doi.org/10.1016/j.jspi.2013.03.018).
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1).
- Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, **456–457**, 12–29, doi:[10.1016/j.jhydrol.2012.05.052](https://doi.org/10.1016/j.jhydrol.2012.05.052).
- Thiemeßl, M. J., A. Gobiet, and G. Heinrich, 2012: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, **112**, 449–468, doi:[10.1007/s10584-011-0224-4](https://doi.org/10.1007/s10584-011-0224-4).
- Thrasher, B., E. P. Maurer, C. McKellar, and P. Duffy, 2012: Technical note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol. Earth Syst. Sci.*, **16**, 3309–3314, doi:[10.5194/hess-16-3309-2012](https://doi.org/10.5194/hess-16-3309-2012).
- Verkade, J., J. Brown, P. Reggiani, and A. Weerts, 2013: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.*, **501**, 73–91, doi:[10.1016/j.jhydrol.2013.07.039](https://doi.org/10.1016/j.jhydrol.2013.07.039).
- Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, doi:[10.1175/2009JCLI2909.1](https://doi.org/10.1175/2009JCLI2909.1).
- Vrac, M., and P. Friederichs, 2015: Multivariate—intervariable, spatial, and temporal—bias correction. *J. Climate*, **28**, 218–237, doi:[10.1175/JCLI-D-14-00059.1](https://doi.org/10.1175/JCLI-D-14-00059.1).
- Wang, L., and W. Chen, 2014: Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. *Atmos. Sci. Lett.*, **15**, 1–6, doi:[10.1002/asl2.454](https://doi.org/10.1002/asl2.454).
- Wang, X., D. K. Thompson, G. A. Marshall, C. Tymstra, R. Carr, and M. D. Flannigan, 2015: Increasing frequency of extreme fire weather in Canada with climate change. *Climatic Change*, **130**, 573–586, doi:[10.1007/s10584-015-1375-5](https://doi.org/10.1007/s10584-015-1375-5).
- White, R., and R. Toumi, 2013: The limitations of bias correcting regional climate model inputs. *Geophys. Res. Lett.*, **40**, 2907–2912, doi:[10.1002/grl.50612](https://doi.org/10.1002/grl.50612).
- Wilcke, R. A. I., T. Mendlik, and A. Gobiet, 2013: Multi-variable error correction of regional climate models. *Climatic Change*, **120**, 871–887, doi:[10.1007/s10584-013-0845-x](https://doi.org/10.1007/s10584-013-0845-x).
- Zhang, F., and A. P. Georgakakos, 2012: Joint variable spatial downscaling. *Climatic Change*, **111**, 945–972, doi:[10.1007/s10584-011-0167-9](https://doi.org/10.1007/s10584-011-0167-9).
- Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, doi:[10.1002/wcc.147](https://doi.org/10.1002/wcc.147).
- Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, doi:[10.1175/1520-0493\(1998\)126<0725:APAFMF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2).