Environmental Modelling & Software 104 (2018) 130–152

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A software toolkit for correcting systematic biases in climate model simulations

R. Mehrotra*, F. Johnson, A. Sharma

School of Civil and Environmental Engineering, University of New South Wales, Australia

ARTICLE INFO

Article history:
Received 13 September 2017
Received in revised form
1 February 2018
Accepted 11 February 2018

Keywords:
Bias correction
R package
Persistence
Multivariate
Distribution based quantile matching
Nested bias correction

ABSTRACT

Simulations from climate models require bias correction prior to use in impact assessments or for statistical or dynamic downscaling to finer scales. There are a number of different approaches to bias correction, although most of these focus on a single variable for a particular location. Another limitation is that often corrections are only applied for one time scale of interest, for example daily or monthly aggregated simulations despite evidence of different bias structures existing at different time scales. Recent works have sought to address each of these limitations and have led to the development of the Multivariate Recursive Nesting Bias Correction (MRNBC) and Multivariate Recursive Quantile-matching Nested Bias Correction (MRQNBC) methods. An open-source software toolkit in the R statistical computing environment has been developed to provide access to these methods. Several applications of the software are demonstrated in this paper along with information about the capabilities of the software.

1. Introduction

General circulation models (GCMs) are becoming increasingly sophisticated with improvements in resolution and the range of processes that are represented. As a result, in many cases GCMs are now more accurately referred to as Earth System Models (ESMs) because of the number of processes that can be simulated. Despite these improvements and overall confidence in the representation of large scale responses such as the global temperature sensitivity, there remain a number of biases in GCM simulations, particularly with respect to the hydrological cycle. Dynamic downscaling using regional climate models (RCMs) can improve some of these biases because their finer resolutions allow topography to be more accurately represented and at the finest resolutions, these models are now considered convection-permitting. However in many cases significant biases can persist either from the driving GCM or the RCM itself. When GCM or RCM simulations are used in statistical downscaling approaches or directly for impact assessments, bias correction of the variables of interest is required (Mehrotra and Sharma, 2006, 2010). There is also an increasing interest in the need to correct GCM biases in the lateral boundary conditions used to downscale to finer resolutions using appropriately chosen RCMs (Rocheta et al., 2017).

Traditionally bias correction has focussed on correcting the representation of individual variables over a single time-scale of interest (e.g., daily or monthly data). The underlying idea behind any bias correction approach is to identify the bias (in a statistic or quantile) for the current climate and correct the future climate under the assumption that the bias does not change over time. Daily or monthly standardization forms the most basic bias correction and is used to correct for systematic biases in the mean and variances of GCM simulations (Wilby et al., 2004). Nonparametric bias correction approaches include quantile matching, correction factors and transfer functions based approaches (e.g., Arnell and Reynard, 1996; Chen et al., 2013; Chiew and McMahon, 2002; Teutschbein and Seibert, 2013; Mpelasoka and Chiew, 2009; Ines and Hansen, 2006; Li et al., 2010; Piani et al., 2010; Wood et al., 2004). These approaches address biases in the overall distribution of GCM simulations (e.g., Cayan et al., 2008; Li et al., 2010; Teutschbein and Seibert, 2013; Maurer and Hidalgo 2008). A variation of quantile matching, named equidistant quantile matching (EQM), has been proposed by Li et al. (2010). Analogous approaches have also been proposed to correct biases in the frequency

* Corresponding author. Water Research Centre, School of Civil and Environmental Engineering, University of New South Wales, Sydney, 2052, Australia.
E-mail address: raj.mehrotra@unsw.edu.au (R. Mehrotra).

https://doi.org/10.1016/j.envsoft.2018.02.010
1364-8152/

spectrum of variables of interest (Nguyen et al., 2016, 2017).

Commonly used bias correction approaches generally consider a single time scale (e.g. day, month or year) and do not consider the biases in persistence attributes. When the bias corrected variables are aggregated/averaged to longer time scales (for example, daily to monthly/seasonal or annual), observed and bias corrected statistics can be quite different. Johnson and Sharma (2012) proposed the idea of nesting multiple time scales including a persistence correction in the standard bias correction procedure. This was named Nested Bias Correction (NBC). As the nesting was found to create artifacts in some of the statistics of the bias corrected series, Mehrotra and Sharma (2012) proposed multiple repeats of the nested bias correction procedure to minimise the biases at all time scales. This modification was termed Recursive Nested Bias Correction (RNBC).

One of the criticisms of bias correction is that it is generally applied to each variable separately (Mehrotra and Sharma, 2015, 2016; Vrac and Friederichs, 2015; Li et al., 2014). As a result, although it improves the statistics of each variable, the physical dependencies between different variables are overlooked (Colette et al., 2012; Maraun, 2013). For water resources impact assessments, bias corrected time series of a number of different variables is often needed in catchment modelling (for example precipitation and temperature, potential evapotranspiration etc.) and statistical downscaling (requires a number of bias corrected upper air variables). A related problem can arise with poor representation of spatial correlations if variables are corrected separately for different locations (Hnilica et al., 2016; Hanel et al., 2017).

To address these problems, multivariate bias correction approaches have been proposed. Piani and Haerter (2012) proposed a bias correction approach to simultaneously correct temperature and precipitation. This was achieved by correcting one time series (e.g., precipitation) conditionally to the bias-corrected values of the other variable's time series (e.g., temperature). Copula-based methods have also been proposed to consider the joint dependence between variables or the spatial dependence across grids (Mao et al., 2015; Vrac and Friederichs, 2015). Mehrotra and Sharma (2015) proposed a parametric multivariate extension, whilst a multivariate and multi-timescale extension of quantile matching based nonparametric bias correction alternatives was suggested by Mehrotra and Sharma (2016). The latter approach corrects biases in probability space as well as the more routine distribution corrections. The bias corrected simulations are shown to have the correct dependence between variables or locations as well as improved persistence structures and distributions over multiple time-scales.

The mathematical relationships used in bias correction are developed based on historical and current climate observations and are applied in a future climate under the assumption of stationarity over time (Salvi et al., 2016). The stationary bias assumption is questionable (Nahar et al., 2017; Buser et al., 2009; Ehret et al., 2012) but efforts to improve on the assumption still need further development. Different researchers have recognised this issue and have suggested possible solutions. Grillakis et al. (2016) provide a review of a few of these approaches in the context of bias correction.

While multivariate bias correction approach is attractive, the multivariate setup requires estimation of additional parameters, extremely large matrices and complex mathematical formulations, making it inaccessible to practitioners wishing to use such methods for climate change impact assessments. Keeping in view these aspects, a Multivariate Bias Correction (MBC) software package has been developed in the R statistical computing environment. The package includes both Multivariate Recursive Nesting Bias Correction (MRNBC) and Multivariate Quantile-matching Recursive Nesting Bias Correction (MRQNBC) approaches (Mehrotra and Sharma, 2015, 2016) and makes it simple to implement both these approaches in a fairly simple manner. This paper describes the software package and provides simple examples of its applications.

2. Multivariate bias correction

The multivariate modelling of Mehrotra and Sharma (2015, 2016) corrects the raw GCM simulations at pre-defined time-scales to match the observed distributional and persistence attributes at each of these time-scales. While we do not claim that the proposed multivariate modelling will keep the physical relationship among the climate variable intact, it is certainly a better choice than the univariate bias correction option, especially when dependence biases (between the multiple variables of interest) are present. Future GCM simulations have the same corrections applied, which allows for changes in the statistical properties over time but corrects for biases, assuming that the biases are stationary and smaller than the magnitude of changes that are projected (Chen et al., 2015). The approach first applies a univariate bias correction at each time-scale to match the observed statistical/distributional attributes. These univariate bias corrected time series are subsequently adjusted to reproduce the observed auto and cross dependence attributes at each time-scale. More details on the structure of the multivariate bias correction models are discussed in Salas (1980) and Mehrotra and Sharma (2015, 2016) and only a few key points related to multivariate and multi-timescale aspects

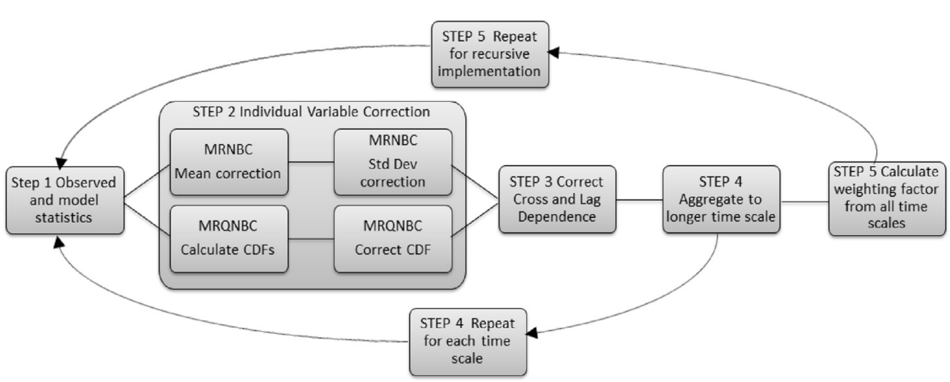

Fig. 1. Correction flow chart of MBC.