Journal of Hydrology 536 (2016) 119–132

Contents lists available at ScienceDirect

Journal of Hydrology

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

Water resources climate change projections using supervised nonlinear and multivariate soft computing techniques

Ali Sarhadi [a,*], Donald H. Burn [a], Fiona Johnson [b], Raj Mehrotra [b], Ashish Sharma [b]

[a] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario N2L-3G1, Canada
[b] School of Civil and Environmental Engineering, University of New South Wales, Kensington, New South Wales 2052, Australia

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 17 February 2016
Accepted 18 February 2016
Available online 27 February 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Statistical downscaling
Dimensionality reduction
Machine-learning
Multivariate bias correction

SUMMARY

Accurate projection of global warming on the probabilistic behavior of hydro-climate variables is one of the main challenges in climate change impact assessment studies. Due to the complexity of climate-associated processes, different sources of uncertainty influence the projected behavior of hydro-climate variables in regression-based statistical downscaling procedures. The current study presents a comprehensive methodology to improve the predictive power of the procedure to provide improved projections. It does this by minimizing the uncertainty sources arising from the high-dimensionality of atmospheric predictors, the complex and nonlinear relationships between hydro-climate predictands and atmospheric predictors, as well as the biases that exist in climate model simulations. To address the impact of the high dimensional feature spaces, a supervised nonlinear dimensionality reduction algorithm is presented that is able to capture the nonlinear variability among projectors through extracting a sequence of principal components that have maximal dependency with the target hydro-climate variables. Two soft-computing nonlinear machine-learning methods, Support Vector Regression (SVR) and Relevance Vector Machine (RVM), are engaged to capture the nonlinear relationships between predictand and atmospheric predictors. To correct the spatial and temporal biases over multiple time scales in the GCM predictands, the Multivariate Recursive Nesting Bias Correction (MRNBC) approach is used. The results demonstrate that this combined approach significantly improves the downscaling procedure in terms of precipitation projection.

1. Introduction

It is now broadly accepted that global warming is impacting hydrological and climatological processes on regional and local scales. These changes are expected to increase extreme hydrological events and threaten water resources in different parts of the world. Therefore, the assessment of the climate change impacts on the availability of surface water resources is of particular interest to water resources managers and decision makers for mitigating the adverse impacts of global warming.

Future climate change information is derived from simulated large-scale atmospheric processes developed based on General Circulation Models (GCMs). GCMs simulate climate at coarse spatial scales, and are unable to provide information that can be directly used at the finer scales of interest to hydrologists for assessing how possible climate-change impacts on surface water availability may affect water supply (Bennett et al., 2012; Dingbao Wang, 2013). This inadequacy has been the reason for developing dynamical and statistical downscaling techniques to transfer large-scale global atmospheric variables (provided by GCMs) to regional and local hydro-climate information for use in climate change impact studies. One option for this is dynamical downscaling approaches, which are based on obtaining finer information from Regional Climate Models (RCMs) driven by boundary conditions simulated using GCMs (Najafi and Moradkhani, 2015). The limitation of these approaches is that they require expensive and complicated computations, and use biased lateral boundary inputs as the basis of their simulations (Rocheta et al., 2014a), inputs that cannot be easily bias corrected for use. On the other hand, most commonly used and popular regression-based statistical downscaling approaches are based on empirical and quantitative relationships developed between a local hydro-climate variable and large-scale atmospheric predictors developed by reanalysis data and GCMs. The regression-based statistical downscaling is carried out in two main steps: (i) deriving statistical relationships from historical climate information and hydro-climate variables of interest (developing a statistical model step), and (ii) using these models to project

* Corresponding author.
E-mail address: asarhadi@uwaterloo.ca (A. Sarhadi).

http://dx.doi.org/10.1016/j.jhydrol.2016.02.040
0022-1694/© 2016 Elsevier B.V. All rights reserved.

hydro-climate variables in the future, relying on the assumption that current empirical models are applicable to GCM simulations of the future (a projection step).

Bias correction has been shown to improve the quality of GCMs for use in projecting hydro-climate variables under different climate change scenarios of the future (Mehrotra and Sharma, 2012; Ojha et al., 2013). Regarding the projection step of the statistical downscaling, the accuracy of climate change simulations is influenced by the similarity in the relationship between actual atmospheric variables and observed rainfall, as compared to simulated variables and the presumed projected rainfall. This similarity is expected to be influenced by the biases that characterize the raw GCM fields. Therefore, in the statistical downscaling processes an initial post-processing correction must be carried out on GCM outputs representing the current climate, based on the statistical characteristics of observations, to remove the difference between observed and simulated large-scale atmospheric variables. The bias correction model over a historical time period is assumed to be the same in the future, and can thus be employed on future GCM simulations (Johnson and Sharma, 2015). In addition, anomalous atmospheric circulation patterns influence the hydrological cycle and large-scale atmospheric variables. Interannual and interdecadal variability in the large-scale climate modes are often not well represented in GCM simulations (Rocheta et al., 2014b), resulting in uncertainty and biases in projections of hydro-climate variables relating to the future. Thus, raw GCMs must also be corrected to capture the effect of low frequency variability of teleconnections on large-scale atmospheric variables (Mehrotra and Sharma, 2012).

It is therefore critical to identify the nature of these biases and develop methods to address these sources of uncertainty. Several bias correction approaches have been developed to quantify the difference between observed (or reanalysis) data and large-scale GCM-simulated variables and form the basis on which to correct biases in both current and future atmospheric GCM simulations. Commonly used bias correction procedures can be classified into two main categories. The first relies on delta change and scaling approaches, including quantile mapping, scaling, correction factor, and transfer functions (a detailed review of the various methods can be found in Johnson and Sharma (2012) and Fowler et al. (2007)). All the methods in this category can be applied for post-processing either on GCM variables or outputs of downscaling models. Their main drawback is that they only take into account biases in the distribution of GCM simulations rather than biases in the representation of persistence and variability in simulations. Current climate variability is thus assumed to remain the same in the future. The second category involves approaches relying on statistical bias correction. Simple techniques in this category such as Monthly Bias Correction (MBC) (Ojha et al., 2013) correct only systematic biases in the mean and variance of GCM-simulated variables or output of downscaled processes in an independent time scale, ignoring the influence of regional and global teleconnection signals. However, the impact of teleconnections on hydro-climate variable behaviors in large scales makes it important to properly represent the interannual and interdecadal fluctuation of climate in the raw GCM outputs. To do so, Johnson and Sharma (2012) developed a bias correction methodology by adding lag-1 correlation to the procedure to correct the representation of low frequency variability between GCM simulations and observed data. The approach corrects the distributional and persistence GCM biases from fine to progressively longer time series and is called Nested Bias Correction (NBC). An extension of NBC was proposed by Mehrotra and Sharma (2012) to enhance the representation of variability at multiple time series by reducing biases through repeating the nesting process several times (Recursive Nesting Bias Correction, RNBC method).

One of the criticisms of bias correction is that the statistical corrections do not maintain the physical relationships between different climate variables (Ehret et al., 2012; Haerter et al., 2011; Rocheta et al., 2014a). To overcome this problem Mehrotra and Sharma (2015) developed a bias correction method that can consider multiple variables and correct the cross correlations between them over a range of time scales. The Multivariate Recursive Nesting Bias Correction (MRNBC) extends the previous nesting bias correction approaches (Johnson and Sharma, 2012) and has been shown to be effective at correcting predictors for statistical downscaling leading to improved downscaled simulations. An alternative implementation could include using multiple locations rather than multiple variables to correct spatial as well as temporal dependence in the GCM simulations.

After correcting the biases that characterize raw GCM simulations, there remain a number of challenges in developing statistical downscaling models, due to the complexity of the climate system. Two main difficulties exist: (i) identification of the large-scale atmospheric predictors conveying relevant climate change information, and (ii) development of the right quantitative functional relationship for capturing the complex nonlinearity between target hydro-climate variables and atmospheric simulated predictors. While the first of these problems is partly due to the high dimensionality of the climate processes that lead to rainfall, the second is due to poor characterization of the functional form. This paper attempts to address both these limitations as discussed below.

To address the dimensionality problem in statistical downscaling processes, many studies have used conventional unsupervised dimensionality reduction methods, such as PCA, CCA, and clustering (Shashikanth et al., 2014; Tisseuil et al., 2010; Wójcik, 2015), exploring a limited sequence of subspaces from the high dimensional predictors to capture the maximum variability and the covariance structure of data without taking into account the target hydro-climate variable. These purely unsupervised techniques may throw away low variations having high predictive potential for the response variable, or keep high variance explanatory variables that are irrelevant for the task at hand. A few attempts have also used selective dimensionality-reduction methods, which cannot adequately capture the nonlinearity and interaction properties of predictors (Ahmadi et al., 2015; Hammami et al., 2012). A supervised dimensionality reduction method, called "Supervised Principal Component Analysis", was presented as an efficient alternative by Sarhadi et al. (2015), illustrating significant improvements in the downscaled rainfall field.

Due to the complex nonlinear relationship existing between target hydro-climate variables and large-scale atmospheric variables, standard linear methods also fail to capture the nonlinear functional relationship. Therefore, to address the second challenge in developing a statistical modeling step, considerable attention has been paid in the last few years to nonlinear-based soft computing data-driven regression models. Machine-learning methods have gained more popularity for statistical downscaling modeling. Among machine learning methods, Support Vector Regression (SVR) has been widely employed in hydrology for nonlinear stochastic modeling of different hydro-climatic variables (Chen et al., 2012, 2010; Nasseri et al., 2013). In recent years, however, a fully probabilistic Bayesian framework of the SVR known as Relevance Vector Machine (RVM) has gained more popularity in regression-based statistical modeling. Ghosh and Mujumdar (2008) compared the results obtained from the SVR and RVM models for projection of streamflow in a statistical downscaling process. They presented the advantages of the RVM over the SVR to improve the model performance. In another attempt, the authors also employed the RVM model with a fuzzy clustering method to downscale GCM outputs for monsoon streamflow projections (Mujumdar and Ghosh, 2008). Joshi et al. (2013) analyzed the