Analysis of climate change impact on precipitation in Danjiangkou reservoir basin by using statistical downscaling method

HUA CHEN¹, JING GUO¹, SHENG-LIAN GUO¹, C-Y XU² & VIJAY P. SINGH³

¹State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University 430072, China
chua@whu.edu.cn
²Department of Geosciences, University of Oslo, NO-0316 Oslo, Norway
³Department of Biological and Agricultural Engineering, Texas A & M University, Scoates Hall, 2117 TAMU, College Station, Texas 77843-2117, USA

Abstract General circulation models (GCMs) are often used to assess the impact of climate change. As they operate on coarse scales, the simulation results obtained from GCMs are not particularly in accordance with comparatively smaller river basin scales. A statistical downscaling method based on a least square support vector machine (LS-SVM) is proposed for downscaling daily total precipitation series from GCMs outputs for the Danjiangkou reservoir basin in China. The downscaling method is developed and validated using large-scale predictor variables derived from the National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP/NCAR) reanalysis data set. The performance of the LS-SVM downscaling method is also compared to a well-known statistical downscaling model (SDSM). The downscaling results suggest that the LS-SVM is an efficient method for downscaling daily precipitation series in the study region. The statistical relationship obtained is used to project future precipitation from CGCM2 for IPCC SRES A2 and B2 scenarios. The downscaling results, corresponding to both scenarios, show that there is an increase in the average annual precipitation downscaled from CGCM2 by about 277.51 mm for the IPCC SRES A2 scenario and by about 157.65 mm for the IPCC SRES B2 scenario, by the 2080s.

Key words climate change; statistical downscaling; least square support vector machine (LS-SVM); Danjiangkou reservoir basin; SDSM

INTRODUCTION

New problems, arising from impact of climate changes in ungauged basins, reveal that past hydrological records may no longer be adequate to make predictions and forecasts about the future. It is desirable to know the necessary information for reliable prediction under climate change. To cope with this, it is necessary to combine downscaled General Circulation Models (GCMs) climate change scenarios with PUB to assess the potential long-term change of the water resources in ungauged basins. GCMs, which describe the atmospheric process by mathematical equations, are one of the most important tools in the study of climate variability and climate change impact on hydrology. Although GCMs are able to simulate reliably the most important mean features of the global climate at large scales, they perform poorly at smaller spatial and temporal scales relevant to the regional impact analyses (Grotch & MacCracken, 1991; Wilby et al., 2007). The main reason is that the spatial resolution of GCM grids is too coarse to resolve many important sub-grid scale processes and GCM output is often unreliable at individual grid and sub-grid scales (Wilby et al., 1999; Xu, 1999).

In recent years numerous statistical downscaling methods have been applied to deal with this problem of mismatch of spatial and temporal scales between GCMs and watershed hydrological models. Statistical downscaling concerns how to draw empirical relationships that transform large-scale features of the GCM (Predictors) to regional-scale variables (Predictands), such as precipitation and temperature (Tripathi et al., 2006; Fowler et al., 2007). A diverse range of empirical downscaling techniques have been developed over the past two decades and each method generally lies in one of the three major categories, including weather pattern schemes (Bárdossy et al., 2005), weather generators (WGs) (Kilsby et al., 2007) and regression models (Fowler et al., 2007). Among these methods, regression models which are also termed transfer functions may be the most popular approaches and are directly used to quantify a relationship
between the predictand and a set of predictors, such as multiple regression models (MLP) (Wilby et al., 1999), artificial neural networks (ANNs) (Zorita & von Storch, 1999), canonical correlation analysis (CCA) (Busuioc et al., 2001), and singular value decomposition (SVD) (Huth, 1999). One of the well-recognized statistical downscaling tools that implements a regression and stochastic method is the Statistical Downscaling Model (SDSM; Wilby et al., 2002). SDSM is used in this study as a benchmark model because it appears to be one of the most widely used models for precipitation and temperature downscaling.

Recently, support vector machines (SVM), which are novel machine learning algorithms proposed by Vapnik (1995, 1998) have been widely applied in the fields of classification and regression analysis (Tripathi et al., 2006; Ghosh & Mujumdar, 2007; Yu & Liong, 2007). Although SVM has extensive applications in various fields, it has some drawbacks in dealing with large sample data, such as slow training speed, low implementation efficiency, and inadaptability to noise and outliers. As an improved algorithm based on SVM, the least square support vector machine (LS-SVM) developed by Suykens & Vandewalle (1999), which is basically a ridge-type learning machine, has received considerable attention and has been verified to deal efficiently with large sample data sets. In this study, LS-SVM was applied to establish a statistical relationship between large-scale GCM simulations and precipitation in the Danjiangkou reservoir basin in China.

The paper is organized as follows: after the introduction, a description of the study area and the data used in this study are provided; the methods are then briefly introduced, including the LS-SVM and SDSM; in the next section the comparative results are discussed and analysed from the two downscaling methods and the impact of climate change on precipitation is assessed; finally, some concluding remarks are made in the last section.

STUDY AREA AND DATA

The study area selected in this study for the application of downscaling models is the Danjiangkou reservoir basin, which is the source of water for the middle route of the well-known South-to-North Water Diversion Project (SNWDP) in China (Fig. 1; Chen et al., 2007). The basin lies in a subtropical monsoon zone and its annual precipitation is about 700–1000 mm. Daily precipitation was selected as predictand in this study from 1961 to 2000 and its data were provided by the National Climatic Centre of China for eight National Meteorological Observatory (NMO) stations within the Danjiangkou reservoir basin as showed in Fig. 1. The altitudes and longitudes of the stations are listed in Table 1. One of the most important steps in downscaling is to select appropriate predictors or characteristics from GCMs, and there are some main factors to constrain the choice of predictors proposed (Wilby et al., 1999). Precipitation is a consequence of mean sea level pressure (MSLP), geopotential height (GH), and specific humidity (SH) (Wilby et al., 1999; Wetterhall et al., 2005; Ghosh & Mujumdar, 2007). Therefore, MSLP, surface air temperature (2 m), 500 hPa GH and SH, 850 hPa GH and SH are considered as the predictors for predicting precipitation in the present study. The downscaling methods are developed and validated using large-scale NCEP/NCAR reanalysis predictor variables which are available free on the internet (http://dss.ucar.edu/pub/reanalysis/). Climate variables corresponding to the future climate change scenario for the study area are extracted from the CGCM2. The future climate change considered in this study corresponds to the so-called business-as-usual scenario. Thus the CGCM2 output used for this study is the result of the IPCC SRES A2 and B2 scenarios. The A2 scenario envisions population growth to 15 billion by the year 2100 and rather slow economic and technological development, and the B2 scenario envisions slower population growth (10.4 billion by 2100) with a more rapidly evolving economy and more emphasis on environmental protection (IPCC, 2001).

As the Danjiangkou reservoir basin has distinct seasonal variation of the atmospheric circulation and precipitation as the result of the subtropical monsoon circulation, the downscaling models will be calibrated in four seasons: winter (December, January, and February), spring (March, April, and May), summer (June, July, and August) and autumn (September, October, and
Table 1 The latitude and longitude of the meteorological stations in the Danjiangkou reservoir basin.

<table>
<thead>
<tr>
<th>No.</th>
<th>Station name</th>
<th>Latitude(°N)</th>
<th>Longitude(°E)</th>
<th>Elevation(m a.s.l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hanzhong</td>
<td>33.04</td>
<td>107.02</td>
<td>508.4</td>
</tr>
<tr>
<td>2</td>
<td>Foping</td>
<td>33.32</td>
<td>107.59</td>
<td>1087.7</td>
</tr>
<tr>
<td>3</td>
<td>Shangzhou</td>
<td>33.52</td>
<td>109.58</td>
<td>742.2</td>
</tr>
<tr>
<td>4</td>
<td>Xixia</td>
<td>33.18</td>
<td>111.30</td>
<td>250.3</td>
</tr>
<tr>
<td>5</td>
<td>Wanyuan</td>
<td>32.06</td>
<td>108.03</td>
<td>129.2</td>
</tr>
<tr>
<td>6</td>
<td>Shiquan</td>
<td>33.03</td>
<td>108.16</td>
<td>484.9</td>
</tr>
<tr>
<td>7</td>
<td>Ankang</td>
<td>32.43</td>
<td>109.02</td>
<td>290.8</td>
</tr>
<tr>
<td>8</td>
<td>Yunxian</td>
<td>32.51</td>
<td>110.49</td>
<td>201.9</td>
</tr>
</tbody>
</table>

November), respectively. In order to evaluate the performance of each method, the entire data set was split into two parts, the training set, which is taken as 1961–1990, and the testing set as the remaining 10 years. For assessing the climate change impact on precipitation in the basin, the data series are divided into four distinct periods, namely, the current (1961–2000), the 2020s (2010–2039), the 2050s (2040–2069), and the 2080s (2070–2099) to facilitate trend analysis. Prior to downscaling the observations, NCEP/NCAR reanalysis and CGCM2 outputs data are standardized to reduce systematic biases in mean and variances. The geographical extent, 102.5°E–115°E, 27.5°N–37.5°N, is chosen to include all areas with noticeable influence on the circulation patterns that govern weather in the Danjiangkou reservoir basin. Figure 1 shows the NCEP grid points (2.5° latitude × 2.5° longitude) superposed on the map of Danjiangkou reservoir basin. CGCM2 grids (3.75° latitude × 3.75° longitude) were spatially interpolated into the NCEP grids by using the inverse distance weighting method.

There are six predictor variables at 24 NCEP grid points with a dimensionality of 144 for statistical downscaling methods, which will be a dimension disaster for personal computers.
Principal Component Analysis (PCA) was used to reduce dimensions and compress data while retaining most of the information content of the original data set. It can be observed that the first eight Principal Components have represented 90.1% of the information content of the original predictors, and therefore they are used as input to the statistical downscaling methods in this study.

METHODS

The least square support vector machine

As an improved algorithm based on SVM, the least square support vector machine (LS-SVM) developed by Suykens & Vandewalle (1999) has received considerable attention. LS-SVM is a kind of SVM under a quadratic loss function. In LS-SVM, the non-sensitive loss function is replaced by a quadratic loss function and the inequality constraints are replaced by equality constraints. The standard framework for LS-SVM estimation is based on a primal-dual formulation.

Given a training data set \( \left\{ x_i, y_i \right\}_{i=1}^{N} \), the goal is to estimate a model of the form:

\[
y = \omega^T \varphi(x) + b
\]

where \( x \in \mathbb{R}^n \), \( y \in \mathbb{R}^n \), \( \varphi(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is the mapping to a high dimensional (and possibly infinite dimensional) feature space and weight \( w \in \mathbb{R}^n \) and basis \( b \in \mathbb{R} \) are parameters to be determined.

The following optimization problem is formulated:

\[
\min_{w,b,c} \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^{N} e_i^2

s.t. \quad y_i = \omega^T \varphi(x_i) + b + e_i, i = 1, 2, \ldots, N
\]

where \( C \) is the penalized parameter which controls the trade off between the fitting accuracy and generalized ability. With the application of Mercer’s theorem on the kernel matrix \( \Omega \) as \( \Omega_{ij} = K(x_i, y_j) = \varphi(x_i)^T \varphi(y_j), (i, j = 1, 2, \ldots, N) \) it is not required to compute explicitly the nonlinear mapping \( \varphi(\cdot) \) as this is done implicitly through the use of positive definite kernel functions \( K \). There are several possibilities for the choice of a kernel function, including linear, polynomial, sigmoid, splines, and radial basis functions (RBF). An RBF kernel can map the training into a possibly infinite-dimensional space and is computationally simple. Moreover, RBF can effectively handle the nonlinear relationship between predictors and predictands. The Lagrangian was used to solve the optimization problem. The resulting LS-SVM model in the dual space becomes:

\[
y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b
\]

where \( \alpha_i \in \mathbb{R}, (i = 1, 2, \ldots, N) \) are the Lagrange multipliers.

The ridge regression formulation is presented in the cost function, and its regularization parameter \( \gamma \) avoids ill conditioning due to possible multicollinearity among the \( m \) dimensions of \( \varphi \). Usually the training of the LS-SVM model involves an optimal selection of the kernel parameter \( \sigma \) and \( \gamma \), which can be done using cross-validation techniques. A robust cross-validation score function (De Brabanter et al., 2002) is an effective method for dealing with outliers and non-Gaussian noise distribution on the data, which is very difficult for the traditional cross-validation criterion, and it is introduced to select the optimal learning parameter of LS-SVM.

Statistical downscaling model (SDSM)

The statistical downscaling model (SDSM) (Wilby et al., 2002) is a hybrid between a multilinear regression method and a stochastic weather generator. The model, which has been applied in many
catchments in the world (Wilby et al., 2002; Coulibaly et al., 2005; Wetterhall et al., 2006) uses large-scale predictors to linearly condition local-scale weather generator parameters. SDSM 4.2 decision support tool was designed for assessing local climate change impacts using a robust statistical downscaling technique (Wilby & Dawson, 2007). It facilitates the rapid development of multiple, low-cost, single-site scenarios of daily surface weather variables under present and future climate forcing, and implements bias correction and variance inflation techniques to reduce the standard error of the estimate and to increase the amount of variance explained by the models to achieve the best possible downscaling performance.

RESULTS AND DISCUSSION

To assess the accuracy of the downscaling methods in producing precipitation input to hydrological models, comparison between observed and simulated mean daily precipitation and standard deviation (SD) is shown in Table 2. It can be observed that the two methods produce small differences in the mean daily precipitation values, as shown in Table 2. In the testing period LS-SVM performs better than SDSM to simulate mean daily precipitation, except in winter, and SDSM is better in terms of SD, except in summer. However, it is also found that SD of the generated series is consistently smaller than that of the observed series for both methods. The underestimation of the observed variance of precipitation has also been found in previous studies (Srikanthan & McMahon, 2001), which is a common problem to be improved. The reason may be that regression based statistical downscaling models often cannot explain the entire variance of the downscaled variable (Wilby et al., 2004) and cannot mimic the extreme precipitation observed in the record. To check the precipitation dynamics simulated by the two methods, the monthly values, averaged over all stations in the Danjiangkou reservoir basin, are reasonably well captured by the two methods, as shown in Fig. 2. The long-term average intra-annual variation is well simulated (Fig. 3) except in autumn when simulated precipitation values are lower than observed values.

In the present study, LS-SVM method is proposed for statistical downscaling of daily precipitation from simulations of GCMs. LS-SVM is a reformulation of standard SVMs and is closely related to regularization networks and Gaussian processes but additionally emphasize and exploit primal-dual interpretations. Compared with SDSM, the proposed LS-SVM has been shown to be suitable for downscaling GCMs simulations to study the impact of climate change on hydrology in the region. The statistics of the mean and variance of precipitation are more often expected to be downscaled with sufficient accuracy when averaged over longer time periods in hydrologic modelling, because precipitation is inherently stochastic, strongly intermittent, and a nonlinear process (Deidda, 1999) that has a great impact on local climate and water balance. The presented downscaling results are typical examples that provide useful regional scale precipitation scenarios for hydrological modelling.

**Table 2** Comparison of observed and downscaled precipitation statistics for the Danjiangkou reservoir basin (mm).

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th></th>
<th>Testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs.</td>
<td>LS–SVM</td>
<td>Bias</td>
<td>SDSM</td>
</tr>
<tr>
<td></td>
<td>Values</td>
<td>Values</td>
<td></td>
<td>Values</td>
</tr>
<tr>
<td>Mean</td>
<td>2.12</td>
<td>2.22</td>
<td>0.1</td>
<td>2.09</td>
</tr>
<tr>
<td>SD</td>
<td>4.7</td>
<td>3.23</td>
<td>–1.47</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>4.52</td>
<td>4.58</td>
<td>0.06</td>
<td>4.91</td>
</tr>
<tr>
<td></td>
<td>4.72</td>
<td>4.52</td>
<td>0.06</td>
<td>4.91</td>
</tr>
<tr>
<td></td>
<td>7.95</td>
<td>5.17</td>
<td>–2.78</td>
<td>4.52</td>
</tr>
<tr>
<td></td>
<td>8.18</td>
<td>5.26</td>
<td>–2.92</td>
<td>4.49</td>
</tr>
<tr>
<td></td>
<td>2.84</td>
<td>2.84</td>
<td>0.01</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td>2.14</td>
<td>2.07</td>
<td>–0.07</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>6.17</td>
<td>3.72</td>
<td>–2.45</td>
<td>4.38</td>
</tr>
<tr>
<td></td>
<td>5.1</td>
<td>2.84</td>
<td>–2.26</td>
<td>4.03</td>
</tr>
<tr>
<td>Mean</td>
<td>0.28</td>
<td>0.29</td>
<td>0.01</td>
<td>0.29</td>
</tr>
<tr>
<td>SD</td>
<td>0.92</td>
<td>0.58</td>
<td>–0.34</td>
<td>0.51</td>
</tr>
</tbody>
</table>

\[ \text{Value}_\text{Down} = \text{Value}_\text{Obs} + \text{Bias} \]
The analysis in the above section has shown the LS-SVM method as a feasible potential alternative for climate impact studies in hydrology in the catchment. Therefore, the LS-SVM downscaling model is selected to downscale the CGCM2 output based on the IPCC SRES A2 and B2 scenarios to obtain future regional simulations of precipitation. Comparisons of observed precipitation and simulated precipitation from CGCM2 based on the IPCC SRES A2 and B2 scenarios downscaled by the LS-SVM (mm) are listed in Table 3. The results show the performance of the downscaling models on a seasonal basis. A significant increase in precipitation based on both GCMs scenarios can be observed in the 2050s and 2080s. A decrease in projected precipitation is found for the A2 scenario in all seasons and the B2 scenario except in the autumn in the 2020s.

The monthly statistics of actual observed values and the current and future CGCM2 simulations downscaled with LS-SVM are summarized and plotted in Fig. 4. This figure shows that the monthly mean values of observed precipitation are quite close to that of the LS-SVM downscaled from the NCEP/NCAR reanalysis data of the current time period (1961–2000). Figure 4(a) and (b) also show an increase in the mean daily precipitation for both scenarios between the
current and the future time periods in May, June, September, and December, while a decrease in February, March, April, and October. There is no significant change for the future downscaled precipitation for both scenarios in January and November. The results show that there is a discordant for the future downscaled precipitation from both scenarios in some months of the year. However, the seasonal variation in the future downscaled precipitation from both scenarios is concordant in almost all seasons in different future periods.

**Table 3** Comparison of observed and downscaled precipitation by LS-SVM from CGCM2 based on the IPCC SRES A2 and B2 scenario (mm).

<table>
<thead>
<tr>
<th></th>
<th>A2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Current</td>
<td>2020s</td>
<td>2050s</td>
<td>2080s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>24.27</td>
<td>26.29</td>
<td>2.02</td>
<td>20.83</td>
<td>–3.44</td>
<td>27.04</td>
<td>2.77</td>
<td>30.40</td>
<td>6.13</td>
</tr>
<tr>
<td>Spring</td>
<td>186.33</td>
<td>193.61</td>
<td>7.28</td>
<td>171.57</td>
<td>–14.75</td>
<td>208.90</td>
<td>22.58</td>
<td>241.99</td>
<td>55.67</td>
</tr>
<tr>
<td>Summer</td>
<td>415.21</td>
<td>416.85</td>
<td>1.64</td>
<td>382.28</td>
<td>–32.93</td>
<td>419.42</td>
<td>4.21</td>
<td>514.39</td>
<td>99.18</td>
</tr>
<tr>
<td>Autumn</td>
<td>242.38</td>
<td>241.14</td>
<td>–1.24</td>
<td>240.58</td>
<td>–1.80</td>
<td>286.24</td>
<td>43.86</td>
<td>358.51</td>
<td>116.13</td>
</tr>
<tr>
<td>Annual</td>
<td>868.80</td>
<td>878.54</td>
<td>9.74</td>
<td>815.96</td>
<td>–52.84</td>
<td>942.51</td>
<td>73.71</td>
<td>1146.31</td>
<td>277.51</td>
</tr>
<tr>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obs</td>
<td>Current</td>
<td>2020s</td>
<td>2050s</td>
<td>2080s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>24.27</td>
<td>26.29</td>
<td>2.02</td>
<td>21.88</td>
<td>–2.39</td>
<td>26.81</td>
<td>2.54</td>
<td>29.51</td>
<td>5.24</td>
</tr>
<tr>
<td>Spring</td>
<td>186.33</td>
<td>193.61</td>
<td>7.28</td>
<td>181.90</td>
<td>–4.43</td>
<td>203.04</td>
<td>16.71</td>
<td>241.15</td>
<td>54.82</td>
</tr>
<tr>
<td>Summer</td>
<td>415.21</td>
<td>416.85</td>
<td>1.64</td>
<td>384.93</td>
<td>–30.29</td>
<td>433.02</td>
<td>17.81</td>
<td>455.55</td>
<td>40.34</td>
</tr>
<tr>
<td>Autumn</td>
<td>242.38</td>
<td>241.14</td>
<td>–1.24</td>
<td>271.63</td>
<td>29.25</td>
<td>298.55</td>
<td>56.17</td>
<td>299.25</td>
<td>56.87</td>
</tr>
<tr>
<td>Annual</td>
<td>868.80</td>
<td>878.54</td>
<td>9.74</td>
<td>861.07</td>
<td>–7.73</td>
<td>962.32</td>
<td>93.52</td>
<td>1026.45</td>
<td>157.65</td>
</tr>
</tbody>
</table>

**Fig. 4** Observed and downscaled precipitation with LS-SVM from the CGCM2 based on the IPCC SRES A2 (a) and B2 scenarios (b).
CONCLUSIONS

This study investigates the applicability of the LS-SVM as the statistical downscaling method for generating daily precipitation at the Danjiangkou reservoir basin in China, and its capability in downscaling daily precipitation amount and other characteristics were compared with SDSM, which is the most widely used regression/based statistical downscaling method. LS-SVM is also used to investigate the possible impact of climate change on precipitation. The results will be helpful to cope with new problems arising from climate changes in ungauged basins.

The study results show that LS-SVM is effective as a statistical downscaling technique as compared to the commonly used statistical downscaling method. The main advantage of this downscaling method is its ability to solve problems in nonlinear function estimation and is closely related to regularization networks and Gaussian processes. The downscaling results, corresponding to both scenarios, show that LS-SVM estimates an increase in the average annual precipitation downscaled from CGCM2 for the IPCC SRES A2 scenario by about 277.51 mm by the 2080s and an increase by about 157.65 mm for the IPCC SRES B2 scenario during the same time period. The results also demonstrate that emphasis should be given in identifying appropriate downscaling tools for impact studies by showing how different future climate scenarios could result in significantly different hydrological impact simulation results for the same watershed.

Acknowledgements The study is financially supported by the National Natural Science Fund of China (50679063; 50809049), International Cooperation Research Fund of China (2005DFA20520) and Research Fund for the Doctoral Program of Higher Education (200804861062). The authors are very grateful to the Chinese National Climate Centre, NOAA Earth System Research Laboratory and Canadian Centre for Climate Modelling and Analysis for providing valuable climatic data.

REFERENCES


Analysis of climate change impact on precipitation in Danjiangkou reservoir basin


