Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin

ARTICLE *in* ADVANCES IN ATMOSPHERIC SCIENCES · JANUARY 2010

Impact Factor: 1.48 · DOI: 10.1007/s00376-009-8071-1

Hua Chen
Wuhan University

Jing Guo
Hydrochina Corporation

Shenglian Guo
Wuhan University

Chong-Yu Xu
University of Oslo

ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 27, NO. 2, 2010, 274–284

Downscaling GCMs Using the Smooth Support Vector Machine Method to Predict Daily Precipitation in the Hanjiang Basin

CHEN Hua[1] (陈 华), GUO Jing[1] (郭 靖), XIONG Wei[3] (熊 伟),
GUO Shenglian[1] (郭生练), and Chong-Yu XU*[2]

[1] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan* 430072

[2] *Department of Geosciences, University of Oslo, PO Box* 1047 *Blindern,* NO-0316 *Oslo, Norway*

[3] *Radar and Avionics Institute of Aviation Industry Corporation of China, Wuxi* 214063

(Received 5 May 2008; revised 18 May 2009)

ABSTRACT

General circulation models (GCMs) are often used in assessing the impact of climate change at global and continental scales. However, the climatic factors simulated by GCMs are inconsistent at comparatively smaller scales, such as individual river basins. In this study, a statistical downscaling approach based on the Smooth Support Vector Machine (SSVM) method was constructed to predict daily precipitation of the changed climate in the Hanjiang Basin. NCEP/NCAR reanalysis data were used to establish the statistical relationship between the larger scale climate predictors and observed precipitation. The relationship obtained was used to project future precipitation from two GCMs (CGCM2 and HadCM3) for the A2 emission scenario. The results obtained using SSVM were compared with those from an artificial neural network (ANN). The comparisons showed that SSVM is suitable for conducting climate impact studies as a statistical downscaling tool in this region. The temporal trends projected by SSVM based on the A2 emission scenario for CGCM2 and HadCM3 were for rainfall to decrease during the period 2011–2040 in the upper basin and to increase after 2071 in the whole of Hanjiang Basin.

Key words: SSVM, GCM, statistical downscaling, precipitation, Hanjiang Basin

Citation: Chen, H., J. Guo, W. Xiong, S. L. Guo, and C.-Y. Xu, 2010: Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin. *Adv. Atmos. Sci.*, **27**(2), 274–284, doi: 10.1007/s00376-009-8071-1.

1. Introduction

The issue of global climate change has been the subject of much discussion in the literature because of the potentially serious impacts upon the Earth's environment. In 2007 the United Nations Climate Change Conference stated that it would be vital for the world's development to resolve global climate change problems and the climate change issue must be placed on the top of member states' political agendas in order to achieve the goals of sustainable development. In January 2008, 20 out of 30 provinces of China suffered from heavy snow which caused enormous economic loss—approximately 13 billion US dollars. As climate, water resources, biophysical and socioeconomic systems are interconnected in complex ways, it follows that a change in any one of these can induce a change in any other. Water-related issues are critical in determining key regional and sectoral vulnerabilities, and therefore the relationship between climate change and water resources is of primary concern to human society and also has implications for all living species.

It is well known that GCMs, which are numerical coupled models and describe the atmospheric processes through mathematical equations, have been one of the most important tools for studying climate

*Corresponding author: Chongyu XU, chongyu.xu@geo.uio.no

change. GCMs represent various earth systems including the atmosphere, oceans, land surface and sea ice and offer considerable potential for studying climate change. At large scales, GCMs which have been steadily evolving over several decades are able to simulate reliably the most important features of global climate. However, these same models perform poorly at smaller spatial and temporal scales relevant to regional impact analyses (Grotch and MacCracken, 1991; Wilby and Wigley, 1997; Wilby et al., 2007). The reason is that the spatial resolution of GCM grids is too coarse to resolve many important sub-grid scale processes and GCM outputs are therefore often unreliable at individual grid and sub-grid box scales (Wilby et al., 1999; Xu, 1999).

To deal with this issue several downscaling methodologies, such as dynamic downscaling and statistical downscaling have been developed. Dynamic downscaling refers to the use of regional climate models (RCMs), or limited-area models (LAMs) which use large-scale and lateral boundary conditions from GCMs to produce higher resolution outputs (Fowler et al., 2007). Statistical downscaling methods seek to draw empirical relationships that transform large-scale features of GCM (predictors) to regional-scale variables (predictands), such as precipitation and temperature (Tripathi et al., 2006). Sophisticated statistical downscaling methods are generally classified into three groups: weather pattern schemes (Conway et al., 1996; Fowler et al., 2000; Bárdossy et al., 2005), weather generators (WGs) (Mason, 2004; Dubrovsky et al., 2004; Kilsby et al., 2007) and regression models (Wilby et al., 1999; Zorita and Storch, 1999; Tripathi et al., 2006; Ghosh and Mujumdar, 2007). Among the statistical downscaling methods, regression models, which are used to directly quantify a relationship between the predictand and a set of predictor variables, are possibly the most popular; examples include multiple regression models (MRMs) (Wilby et al., 1999), artificial neural networks (ANNs) (Zorita and Storch, 1999; Olsson et al., 2004; Tatli et al., 2004; Coulibaly et al., 2005), canonical correlation analysis (CCA) (Karl et al., 1990; von Storch et al., 1993; Busuioc et al., 2001) and singular value decomposition (SVD) (Huth, 1999). MRMs and ANNs have been applied widely owing to their powerful ability in regression analysis and forecasting. However, high dimension problems for MRMs, and, for ANNs, getting trapped in local minima, subjectively choosing model architecture, and over-learning have hampered their more frequent and wider application. Vapnik (1995, 1998) proposed the Support Vector Machine (SVM)—a novel machine learning algorithm—and provided an elegant solution to the above problems. Recently, SVM has been widely applied in the fields of classification and regression analysis (Tripathi et al., 2006; Ghosh and Mujumdar, 2007; Yu and Liong, 2007). Tripathi et al. (2006) proposed a SVM approach for statistical downscaling of monthly precipitation and showed that the method provides a promising alternative to ANNs. Yu and Liong (2007) applied SVM to predict streamflow, comparing the results with those of a local model in a chaotic time series analysis. Significantly better prediction accuracy and faster processing speed were obtained from the SVM scheme. Although SVM has extensive applications in various fields, it has some drawbacks in dealing with large data samples, such as slow training speed, low implementation efficiency and inadaptability to noise and outliers. To overcome these limitations for large data samples there have been many improved algorithms developed (Joachims, 1999; Mangasarian and Musicant, 1999; Platt, 1998; Lee et al., 2005). Lee et al. (2005) proposed a new smoothing strategy for solving the regression of large-scale training data called the Smooth Support Vector Machine (SSVM), which has been verified as being very efficient in his study.

This study investigates the potential use of SSVM in downscaling GCM simulations and assesses the impact of climate change on precipitation in the Hanjiang Basin, a tributary of the Yangtze River in China. More specifically, the following objectives have been set for this paper: (1) to establish the statistical relationship between large-scale circulation (using NCEP/NCAR reanalysis data) and precipitation in the Hanjiang Basin by using two statistical downscaling methods; and (2) to apply the established statistical relationship to predict future precipitation in the Hanjiang Basin by using outputs from CGCM2 and HadCM3 run for the A2 emission scenarios as inputs.

2. Study region and data

2.1 *Study region*

The Hanjiang River is the source of water for the middle route of the well known South-to-North Water Diversion Project (SNWDP) in China, as shown in Fig.1. The basin has a subtropical monsoon climate and the whole Hanjiang Basin is divided into three regions: the Danjiangkou reservoir sub-basin (upper sub-basin), the middle sub-basin, and the lower sub-basin (Chen et al., 2007). The basin's annual precipitation is around 700–1000 mm, which gradually increases from the upper to the lower basin and decreases from south to north in the upper basin. As some of its water is transferred via the SNWDP, this has an impact on socioeconomic development and on the en-

Table 1. Latitude and longitude of the meteorological stations in the Hanjiang Basin.

No.	Station name	Latitude (°N)	Longitude (°E)	Elevation (m)
1	Hanzhong	33.04	107.02	508.4
2	Foping	33.32	107.59	1087.7
3	Shangzhou	33.52	109.58	742.2
4	Xixia	33.18	111.30	250.3
5	Wanyuan	32.06	108.03	129.2
6	Shiquan	33.03	108.16	484.9
7	Ankang	32.43	109.02	290.8
8	Zhenan	33.43	109.15	693.7
9	Fangxian	32.03	110.44	434.4
10	Laohekou	32.23	111.40	90.0
11	Zaoyang	32.09	112.45	125.5
12	Nanyang	33.03	112.58	129.2
13	Tianmen	30.40	113.10	34.1
14	Wuhan	30.37	114.08	23.3
15	Jiayu	29.98	113.92	22.2

vironment in the middle and lower sub-basins. The Hanjiang Basin and the middle route of the SNWDP have been the subject of studies looking into available water resources in the Hanjiang River under the impact of climate change (Guo et al., 2002; Chen et al., 2007). Guo et al. (2002) studied the impact of climate change on water resources in the Hanjiang Basin based on a semi-distributed monthly water balance model and the results showed that precipitation change is the main factor affecting changes in runoff. Chen et al. (2007) investigated spatial and temporal trends of observed annual and seasonal precipitation and temperature from 1951 to 2003 in the Hanjiang Basin by using the Mann-Kendall test and assessed the impact of climate change on runoff in the Danjiangkou reservoir basin. It is important for the management of the Hanjiang River and the middle route of the SNWDP to predict future precipitation, and this can be done through the application of statistical downscaling methods, as discussed above.

2.2 *Predictands and predictors*

The predictand in this study is daily precipitation from 1961 to 2000. Observed data for this were provided by the National Climatic Centre of China, covering 15 National Meteorological Observatory (NMO) stations in the Hanjiang Basin. The location of these stations are shown in Fig. 1 and their altitudes and coordinates are listed in Table 1.

One of the most important steps in a downscaling exercise is to select appropriate predictors, or characteristics from GCMs. Wilby et al. (1999) proposed that there are three main factors constraining the choice of predictors: (1) whether the predictors were reliably simulated by the GCM in the first place; (2) how readily available the GCM output data are; and (3) the correlation strength with the surface variables of interest. Wilby et al. (1999) predicted future precipitation by mean sea level pressure (MSLP), geopotential height (GH) and specific humidity (SH) in the San Juan River Basin. Having considered both the similarities and differences of the Hanjing Basin to the San Juan River Basin used in Wilby et al.'s (1999) study, the predictors for precipitation selected for the present study were: MSLP, surface air temperature (2 m), 500 hPa GH and SH, and 850 hPa GH and SH. NCEP/NCAR daily reanalysis data (Kalnay et al., 1996) were used for training the downscaling model and the daily outputs of CGCM2 and HadCM3 run for the A2 emission scenario were used for projecting future precipitation with the trained model. Owing to projections of climate change depending heavily upon future human activity, climate models such as CGCM2 and HadCM3 are run against scenarios which make different assumptions for changes in future greenhouse

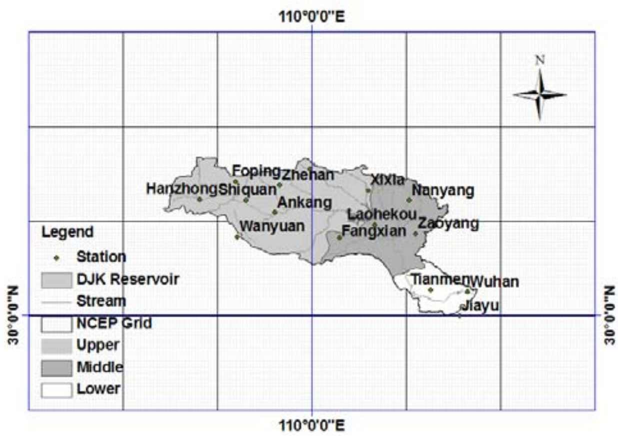

Fig. 1. Location of the Hanjiang Basin and the middle route of the SNWDP in China.

gas levels, land use, and other driving forces. Compared to other plausible scenarios, in the A2 emission scenario global population is expected to increase at a high rate, which assumes regional resiliency and adaptation, with economic development being moderate and focused within regions. For the two GCMs in this study, CGCM2, which is based on the earlier CGCM1 (Flato and Boer, 2001) extends from January 1948 to December 2100, and HadCM3, which is a coupled atmosphere–ocean GCM developed at the Hadley Centre (Gordon et al., 2000) extends from January 1961 to December 2099. For the statistical downscaling approaches, the geographical extent should be chosen to include all areas with noticeable influence on the circulation patterns in the Hanjiang Basin. Figure 1 shows the NCEP/NCAR grid points ($2.5°\times2.5°$) superimposed on a map of the Hanjiang Basin. CGCM2 grids ($3.75°\times 3.75°$) and HadCM3 grids ($2.5°\times3.75°$) are interpolated spatially into the NCEP grids by using the inverse distance weighting method. The following section will present the statistical downscaling methods (SSVM and the ANN) with training and testing steps.

3. Methods and model development

3.1 *SSVM*

SVM is a new machine study method in the field of statistical learning theory and stresses to study statistical learning rules under small samples (Vapnik, 1998). Via structural risk minimization principle to enhance generalized ability, SVM preferably solves many practical problems, such as small sample, nonlinear, high dimension number and global minimum points. The architecture of SVM is shown in Fig. 2. However, SVM cannot deal efficiently with large data samples, as in the case of this study. As an improved algorithm based on SVM, SSVM, developed by Lee et al. (2005), is better able to handle the cases of classification and nonlinear regression with a larger dataset, and has received considerable attention. In SSVM, smoothing techniques are applied to solve important mathematical programming problems and the $\varepsilon-$ insensitive loss function is replaced by the squares of 2-norm $\varepsilon-$ insensitive loss function. In addition, the term $\frac{1}{2}b^2$ is added in the objective function to induce strong convexity and to guarantee that the problem has a unique global optimal solution. The standard framework for SSVM for nonlinear regression consists of the following steps.

A training dataset was given:

$$\boldsymbol{S} = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_i, y_i), \cdots, (\boldsymbol{x}_m, y_m)\} \subseteq \boldsymbol{R}^n \times \boldsymbol{R}, i = 1, \cdots, m$$

where $\boldsymbol{x}_i \in \boldsymbol{R}^n$ represents the input data and $y_i \in \boldsymbol{R}$ is called the observation. The training dataset $\boldsymbol{S}$ is expressed in Eq. (1) which consists of m points in n-dimensional real space $\boldsymbol{R}^n$ represented by the matrix $\boldsymbol{A} \in \boldsymbol{R}^{m\times n}$ and m observations of real value associated with each point. $\boldsymbol{A}_i$ is the ith row of a row vector $\boldsymbol{A}$ in $\boldsymbol{R}^n$. A column vector of ones of arbitrary dimension will be denoted by $\boldsymbol{L}$.

$$\boldsymbol{S} = \{(\boldsymbol{A}_i, y_i) \,|\, \boldsymbol{A}_i \in \boldsymbol{R}^n, y_i \in \boldsymbol{R}, i = 1, 2, \cdots, m\} \quad (1)$$

The goal of the nonlinear support vector regression is to estimate a model of the form:

$$\boldsymbol{y} \approx \boldsymbol{A}\boldsymbol{w} + \boldsymbol{L}b \approx \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} + \boldsymbol{L}b \quad (2)$$

where $\boldsymbol{w} \in \boldsymbol{R}^n$and $b \in \boldsymbol{R}$ are parameters of SVM and $\boldsymbol{w}$ can be represented by $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}$ for some $\boldsymbol{u} \in \boldsymbol{R}^m$. The kernel technique is used to simply replace the $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ in Eq. (2) by a nonlinear kernel matrix $K(\boldsymbol{A}, \boldsymbol{A}^{\mathrm{T}})$, where $K(\boldsymbol{A}, \boldsymbol{A}^{\mathrm{T}})_{i,j} = K(\boldsymbol{A}_i, \boldsymbol{A}_j^{\mathrm{T}})$. The following unconstrained optimization problem is formulated by:

$$\min_{(\boldsymbol{u},b)\in \boldsymbol{R}^{m+1}} \frac{1}{2}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}+b^2)+\frac{C}{2}\sum_{i=1}^{m}\left|K(\boldsymbol{A}_i, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + b - y_i\right|_{\varepsilon}^{2} \quad (3)$$

where $K(\boldsymbol{A}_i, \boldsymbol{A}^{\mathrm{T}})$ is a kernel map from $\boldsymbol{R}^{1\times n} \times \boldsymbol{R}^{n\times m}$ to $\boldsymbol{R}^{1\times m}$. There are several possible functions for the

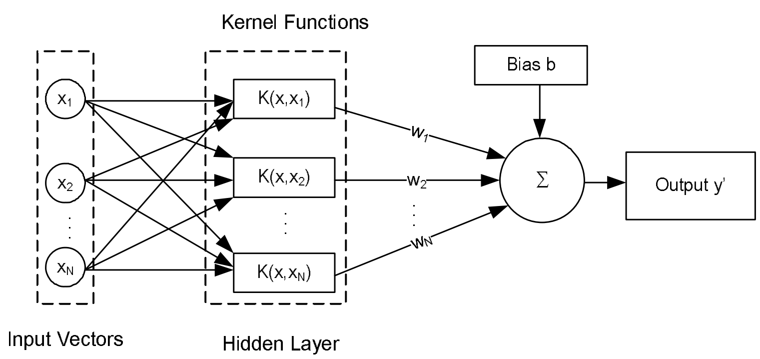

Fig. 2. Architecture of the Support Vector Machine. The coefficients w and b are the adjustable model parameters.

choice of kernel function, including linear, polynomial, sigmoid, splines and Radial Basis Function (RBF). RBF kernel can map the training set into a possibly infinite-dimensional space and is computationally simple. Moreover, the RBF can effectively handle the situation when the relationship between predictors and predictand is nonlinear. The RBF is given by:

$$K(\boldsymbol{A}_i, \boldsymbol{A}_j^{\mathrm{T}}) = \exp\left(-\mu \left\|\boldsymbol{A}_i - \boldsymbol{A}_j\right\|_2^2\right) , \quad i, j = 1, 2, \cdots, m \tag{4}$$

where μ is the width of RBF, which can be adjusted to control the expressivity of RBF.

The squares of 2-norm $\varepsilon-$ insensitive loss function in the above formulation can be accurately approximated by a smooth function which is infinitively differentiable and defined as below:

$$\begin{aligned} & p_\varepsilon^2\left[K(\boldsymbol{A}_i, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + b - y_i, \alpha\right] \\ &= \left\{p[K(A_i, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + b - y_i - \varepsilon, \alpha]\right\}^2 + \\ &\quad \left\{p[-[K(\boldsymbol{A}_i, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + b - y_i] - \varepsilon, \alpha]\right\}^2 \end{aligned} \tag{5}$$

where

$$p(x, \alpha) = x + \frac{1}{\alpha}\log(1 + e^{-\alpha x}), \quad \alpha > 0$$

which is a smooth $p-$function. The following equations can be obtained:

$$\begin{aligned} \min_{(\boldsymbol{u},b)\in \boldsymbol{R}^{m+1}} \Phi_{\varepsilon,\alpha}(\boldsymbol{u}, b) &= \frac{1}{2}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} + b^2) + \frac{C}{2}\sum_{i=1}^{m} p_\varepsilon^2 \times \\ &\quad \left[K(\boldsymbol{A}_i, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + b - y_i, \alpha\right] = \\ &= \min_{(\boldsymbol{u},b)\in \boldsymbol{R}^{m+1}} \frac{1}{2}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} + b^2) + \frac{C}{2}\boldsymbol{L}^{\mathrm{T}} \\ &\quad p_\varepsilon^2[K(\boldsymbol{A}, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + \boldsymbol{L}b - \boldsymbol{y}, \alpha] \end{aligned} \tag{6}$$

where $p_\varepsilon^2[K(\boldsymbol{A}, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + \boldsymbol{L}b - \boldsymbol{y}, \alpha]$ is defined by:

$$\begin{aligned} & p_\varepsilon^2\left[K(\boldsymbol{A}, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + \boldsymbol{L}b - \boldsymbol{y}, \alpha\right]_i \\ &= p_\varepsilon^2\left[K(\boldsymbol{A}_i, \boldsymbol{A}^{\mathrm{T}})\boldsymbol{u} + b - y_i, \alpha\right] \end{aligned} \tag{7}$$

This problem retains the strong convexity and differentiability properties for any arbitrary kernel. Newton-Armijo Algorithm (Lee et al., 2005) is adopted directly to solve Eq. (6), which is the unconstrained minimal problem for $\boldsymbol{u}$ and b.

3.2 *ANN method*

Many ANN structures have been proposed and explored for tasks such as recognition, learning, forecasting and controlling. Among these different structures, the multilayer feed forward networks have the best performance in the context of input-output function approximation (Haykin, 1994). As a matter of fact, almost all ANNs explored in rainfall-runoff modelling are multilayer feed forward networks (Campolo et al., 1999). Among the algorithms used to perform supervised training, the backpropagation method (Rumelhart et al., 1986) has emerged as the most widely used and successful algorithm for the design of the multilayer feed forward neural networks (Haykin, 1994). The backpropagation method has already been used in hydrology (French et al., 1992; Gautam et al., 2000; Wilby et al., 2003; Pang et al., 2007). Pang et al. (2007) applied an ANN to develop a nonlinear perturbation model (NLPM-ANN) for improving rainfall–runoff forecasting efficiency and accuracy. A detailed description of the algorithm, which has been implemented by Pang et al. (2007), is used in this approach.

3.3 *Method evaluation*

To evaluate the performance of SSVM and ANN in downscaling GCM simulations there are subjective decisions that must be made to get a good simulation. The choice of objective functions is governed by the purpose of study (Obled et al., 2002; Wetterhall et al., 2005). If, for example, the purpose is to evaluate extreme events, evaluation parameters that capture those characteristics are selected. The precipitation to be evaluated is also time-dependent, so the objective function must be sensitive to temporal properties. The main purpose of the present study is to predict daily precipitation scenarios that are to be used as an input to hydrological models to simulate future water resources scenarios for the Hanjiang Basin. For such a hydrological application, the differences in the mean and standard deviation between observed and simulated daily precipitation are considered to be most important and are therefore used as criteria in evaluating the downscaling model.

3.4 *Model development*

The Hanjiang Basin has distinct seasonal variation in atmospheric circulation and precipitation as a result of the subtropical monsoon circulation. The analysis is divided into two seasons: one wet season stretching from May to October and accounting for 70%–80% of total annual precipitation; and one dry season from November to April. In this study the downscaling models will be calibrated in wet and dry seasons respectively. In order to evaluate the performance of each method, we split the entire dataset into two parts: the training set which is taken as 1961–1990, and the testing set as the remaining 10 years (i.e. 1991–2000). The training data is used to establish the regression function; the testing set, which is not involved in the training procedure, is used to evaluate the prediction

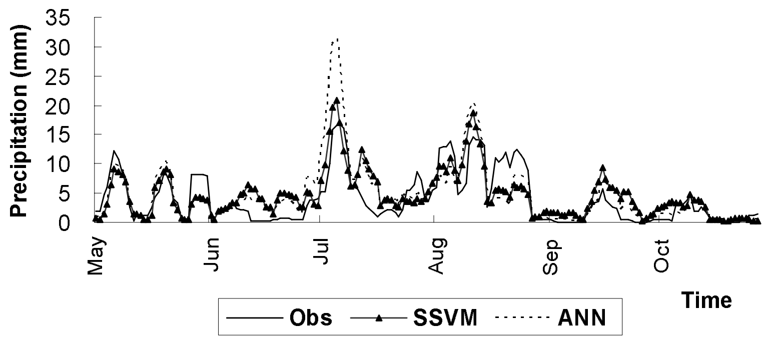

Fig. 3. Five-day running-average daily precipitation of observed (Obs) and simulated (by SSVM and ANN) for the upper Hanjiang Basin during the wet season in 1998.

ability of the resulting regression function.

Prior to downscaling, NCEP/NCAR reanalysis data and GCM data are standardized to reduce systematic biases in the mean and variances of GCM outputs. There are six predictor variables at 24 NCEP grid points with a dimensionality of 144 for statistical downscaling models, and so the predictors' dimensionality may have been too high for a personal computer to handle. Therefore, PCA, which has been widely used to reduce dimension and compress data while keeping most of the information content of the original dataset, is performed to transform the set of correlated 144-dimensional predictors matrix into another set of N-dimensional uncorrelated vectors—called principal components (PCs)—by linear combination, such that most of the information content of the original data set is stored in the first few dimensions of the new set. In this study, it can be observed from Table 2 that the first 8 PCs represent 90% of the information variance of the original predictor matrix; therefore, they are used as an input for the downscaling models to establish relationships with daily precipitation.

In this study, SSVM has two parameters to be determined: the width of RBF, μ and the penalty factor, C. A tuning procedure, which can automatically optimize parameters, is applied to select them (Lee et al., 2005). The ANN model is trained by using the back-propagation algorithm (Pang et al., 2007). The generalization performance of the SSVM and ANN downscaling models is measured on the testing subset, which is different from the training subset.

Table 2. Percentage variances and their cumulative values for the first eight PCs by using PCA in processing NCEP predictors' dataset.

PCs No.	Percent variance	Cumulative percent
1	54%	54%
2	18%	71%
3	7%	79%
4	4%	83%
5	2%	85%
6	2%	88%
7	1%	89%
8	1%	90%

4. Results

To assess the accuracy of the downscaling methods in producing rainfall inputs for hydrological models, a comparison of the mean and standard deviation between observed and simulated daily precipitation is shown in Table 3. It can be seen that there is a small difference between the simulated and observed mean daily precipitation in all three regions. It is also evident that SSVM performs very well in simulating the mean values; however, ANN has slightly better results for standard deviation than SSVM. In addition, the standard deviation of the downscaled series is consistently smaller than that of the observed series for both models. The underestimation of the observed variance of precipitation has also been found in previous studies (e.g. Srikanthan McMahon, 2001), and this defect of stochastic precipitation models will need to be remedied (Wilks, 1989; Gregory et al., 1993). The reason may be that regression-based statistical downscaling models often cannot explain entire variance of the downscaled variable (Wilby et al., 2004) and cannot mimic extreme precipitation observed in the record. Exploration of a wider range of predictor variables and a much longer validation phase could possibly provide more insight into this problem (Tripathi et al., 2006).

To check the precipitation dynamics simulated by the two models, a comparison between simulated and observed values for two techniques after smoothing with a five-day-moving-average filter for the upper Hanjiang Basin is shown in Fig. 3. This shows that simulation by SSVM is more consistent with observed

Table 3. Simulation results for the SSVM and ANN downscaling methods in the Hanjiang Basin (mm).

			Training					Testing				
				SSVM		ANN			SSVM		ANN	
			Obs.	Values	Bias	Values	Bias	Obs.	Values	Bias	Values	Bias
Upper	wet	Mean	4.01	4.02	0.01	3.70	−0.31	3.59	3.39	−0.20	3.32	−0.27
		Stdev	7.41	5.35	−2.06	5.52	−1.89	7.08	5.11	−1.97	5.59	−1.49
	dry	Mean	0.86	0.86	0.00	0.81	−0.05	0.73	0.76	0.03	0.61	−0.12
		Stdev	2.62	1.99	−0.63	2.02	−0.60	2.21	1.87	−0.34	1.90	−0.31
Middle	wet	Mean	3.48	3.48	0.00	2.89	−0.59	3.30	3.56	0.26	2.82	−0.48
		Stdev	8.25	5.06	−3.19	5.39	−2.86	7.81	5.38	−2.43	5.28	−2.53
	dry	Mean	1.18	1.19	0.01	1.10	−0.08	1.04	1.07	0.03	1.15	0.11
		Stdev	3.77	3.07	−0.70	2.69	−1.08	3.21	3.38	0.17	3.28	0.07
Lower	wet	Mean	4.28	4.29	0.01	3.93	−0.35	4.57	4.38	−0.19	3.88	−0.69
		Stdev	10.37	6.70	−3.67	6.68	−3.69	10.94	7.04	−3.90	7.12	−3.82
	dry	Mean	2.12	2.90	0.78	2.01	−0.11	2.17	2.41	0.24	2.07	−0.10
		Stdev	5.30	4.40	−0.90	3.48	−1.82	5.81	4.08	−1.73	4.11	−1.70

data than by ANN. The monthly values, averaged over all stations in different parts of the Hanjiang Basin, are reasonably well captured by the two methods (Fig. 4), especially in the upper basin. The long-term average intra-annual variation is well simulated (Fig. 5a) and the simulation by SSVM is closer to the observed precipitation than that of ANN, except for August and September when it is overestimated.

Through the above analysis, the SSVM method has been confirmed as a feasible potential alternative to the ANN method for climate impact studies in hydrology. Therefore, a SSVM downscaling model is used to downscale CGCM2 and HadCM3 simulations for the A2 emission scenario to obtain simulations of future regional precipitation. Annual precipitation totals are calculated from the simulated future scenarios, which are divided into three periods: 2011–2040, 2041–2070, and 2071–2099. This is done to determine the trend in projected values of precipitation. The results are presented in Table 4, from which it can be observed that there is a decreasing trend of precipitation in both wet and dry seasons during the 2011–2040 period; however, there is a discordant change in wet season during the 2041–2070 period in the upper Hanjiang Basin for both GCMs. In the middle basin the two GCMs show a decrease in precipitation in the wet season during the 2011–2040 period, and an increase in wet and dry seasons during the 2041–2070 period. Opposite results are found for the two GCMs in the dry season during the 2011–2040 period. The precipitation in the lower basin has an upward trends in both GCMs in the wet season during the 2041–2070 period; however, it has different patterns of change in the 2011–2040 period in the wet and dry seasons for the two GCMs. After 2070 both GCMs show precipitation will increase in the wet and dry seasons in all parts of the Hanjiang Basin.

5. Discussion

This study has shown that, compared with the ANN statistical downscaling method, the proposed

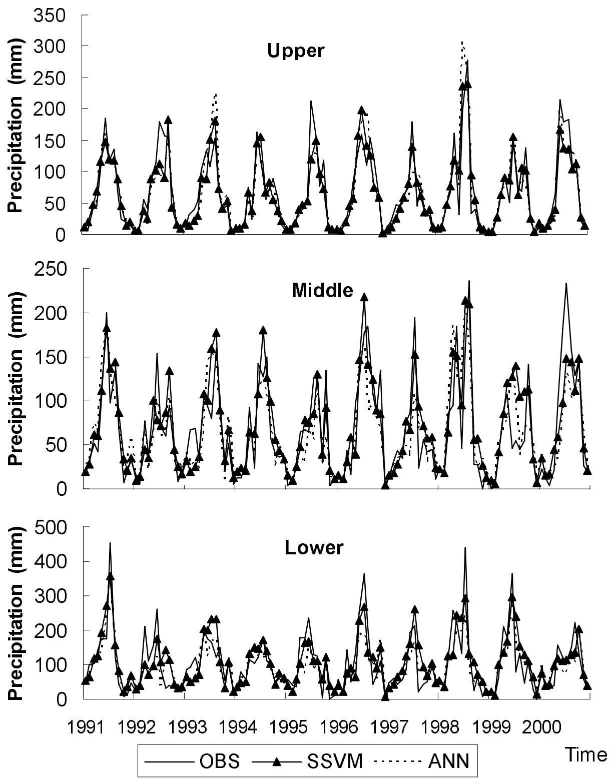

Fig. 4. Monthly observed (Obs) and downscaled (by SSVM and ANN) precipitation averaged over all stations in different parts of the Hanjiang Basin for the period 1991–2000.

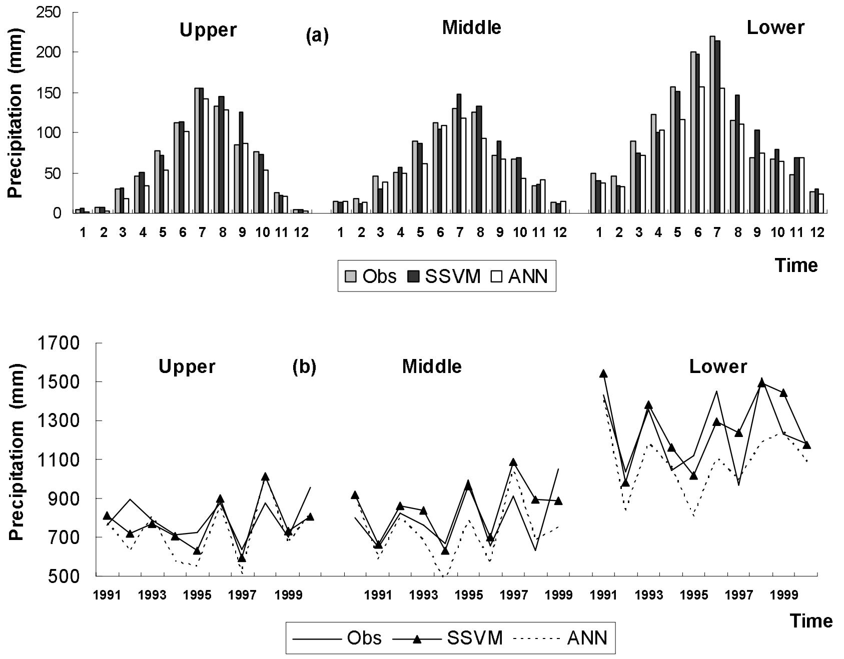

Fig. 5. (a) Mean monthly observed (Obs) and downscaled precipitation (by SSVM and ANN) and (b) annual total precipitation during 1991–2000 averaged over all stations in the upper, middle and lower parts of the Hanjiang Basin.

SSVM method is suitable for downscaling GCM simulations to study the impact of climate change on hydrology in the Hanjiang Basin.

Precipitation is an inherently stochastic, strongly intermittent, and nonlinear process (Deidda, 1999), and has great impact on local climate and water balance (floods and droughts); the statistics of mean and variance are more often expected to be downscaled with a sufficient accuracy when averaged over longer time periods. The presented downscaling results are typical examples that provided useful regional scale precipitation scenarios for hydrological modeling. Table 3 demonstrates that SSVM performs well in capturing the daily mean values and less well in capturing the variability as measured by standard deviation, which is a common problem needing to be addressed. It is necessary for statistical downscaling techniques to propose an integral and strict evaluation method with adjustable weights among various objective measures.

There have been some studies to detect statistically significant trends in precipitation in the Hanjiang Basin. Chen et al. (2007) showed that when entering the 1990s there was a very dry period in the Hanjiang Basin, and they argued that if this situation does not alter in the 21st century it would have serious impacts on agriculture, industry and drinking water supply in the middle China region; in particular, on the viability of the middle route of the SNWDP. Decreasing trends of precipitation in the wet and dry seasons during the 2011–2040 period are found over the upper basin, which may cause a critical situation for the Hanjiang Basin in meeting future irrigation demands,

Table 4. Changes in mean precipitation in the different periods of the 21st century downscaled by SSVM from the observed precipitation in baseline periods (1961–1990) (%).

		Upper		Middle		Lower	
	Period	CGCM2	HadCM3	CGCM2	HadCM3	CGCM2	HadCM3
Wet	2011–2040	−14.16	−13.83	−3.13	−7.98	3.92	−0.82
	2041–2070	−4.42	0.64	10.99	6.01	16.17	4.50
	2071–2100	21.54	23.95	27.77	26.41	25.87	8.07
Dry	2011–2040	−1.88	−8.99	−10.58	14.87	−10.19	12.90
	2041–2070	3.44	9.80	0.90	33.40	−1.31	25.51
	2071–2100	18.28	29.02	25.76	53.17	14.08	37.52

hydropower generation, and especially water supply for the SNWDP. Increasing trends are found over the whole basin after 2070. For the period 2041–2070 different patterns of change in regional precipitation are found for the two GCMs. It is worth mentioning that the future projections of predictands provided by a downscaling model for a given climate change scenario depend on the capability of GCMs to simulate future climate (Tripathi et al., 2006). It is necessary to use more than one GCM on a given climate change scenario to test the robustness of the result projected by the downscaling model, as any one GCM could generate efficient or inaccurate values of predictors. In this study, CGCM2 and HadCM3 for the A2 emission scenarios were selected to project future change in precipitation and the results predicted from the two models are discordant in some parts of the Hanjiang Basin for 2011–2040 and 2041–2070. From the results, it can be concluded that the choice of GCM could have a significant impact on the timing and extent of adaptation responses. Wilby et al. (2006) compared daily precipitation and potential evaporation series arising from three GCMs (HadCM3, CGCM2 and CSIRO) under two emission scenarios (SRES A2 and B2). The study showed that scenarios downscaled from CGCM2 suggested slight increases in deployable abstraction, pointing to a more favorable resource situation. HadCM3 suggested little change in the interannual variability of deployable yield, whereas CGCM2 indicated greater stability in the future. Although it might be better to use more GCMs for the different scenarios to predict precipitation in the Hanjiang Basin, the main purpose of this study was to introduce SSVM as a statistical downscaling tool and, therefore, only two GCMs were chosen to project future climate.

6. Conclusion

Downscaling of GCM outputs to daily precipitation was performed by using SSVM and ANN methods. It has been shown that SSVM is an effective statistical downscaling technique and useful for assessing the impact of climate change in the Hanjiang Basin. The proposed method is capable of producing satisfactory results in terms of daily and monthly mean precipitation in the testing periods. However, it was found that the SSVM method is less skillful in reproducing extreme daily precipitation and standard deviation. Future changes in precipitation were projected using CGCM2 and HadCM3 for the A2 emission scenario and similar conclusions can be drawn from the two models in most of the studied periods in different parts of the Hanjiang Basin. The proposed method could be used in future research to downscale other predictands, such as temperature and evaporation, in order to further assess the impacts of climate change.

Acknowledgements. NCEP/NCAR daily reanalysis data, and CGCM2 and HadCM3 daily data were downloaded freely from the Internet: http://dss.ucar.edu/pub/reanalysis/, www.cccma.ec.gc.ca, and www.ccsn.ca, respectively. The study was supported financially by the National Natural Science Foundation of China (Grant Nos. 50679063 and 50809049), the International Cooperation Research Fund of China (2005DFA20520), and the Research Fund for the Doctoral Program of Higher Education (200804861062).

REFERENCES

Bárdossy, A., I. Bogardi, and I. Matyasovszky, 2005: Fuzzy rule-based downscaling of precipitation. *Theor. Appl. Climatol.*, **82**, 119–129.

Busuioc, A., D. Chen, and C. Hellström, 2001: Performance of statistical downscaling models in GCM validation and regional climate change estimates: Application for Swedish precipitation. *International Journal of Climatology*, **21**, 557–578.

Campolo, M., P. Andreussi, and A. Soldati, 1999: River flow forecasting with a neural network model. *Water Resour. Res.*, **35**, 1191–1197.

Chen, H., S. L. Guo, C-Y. Xu, and V. P. Singh, 2007: Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin. *J. Hydrol.*, **334**, 171–184.

Conway, D., R. L. Wilby, and P. D. Jones, 1996: Precipitation and air flow indices over the British Isles. *Climate Research*, **7**(2), 169–183.

Coulibaly, P., Y. B. Dibike, and F. Anctil, 2005: Downscaling Precipitation and Temperature with Temporal Neural Networks. *Journal of Hydrometeorology*, **6**(4), 483–496.

Deidda, R., 1999: Multifractal analysis and simulation of rainfall fields in space. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, **24**(1–2), 73–78.

Dubrovsky, M., J. Buchtele, and Z. Zalud, 2004: High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling. *Climatic Change*, **63**, 145–179.

Flato, G. M., and G. J. Boer, 2001: Warming asymmetry in climate change simulations. *Geophys. Res. Lett.*, **28**, 195–198.

Fowler, H. J., C. G. Kilsby, and P. E. O'Connell, 2000: A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions. *Hydrology and Earth System Sciences*, **4**, 261–280.

Fowler, H. J., S. Blenkinsopa, and C. Tebaldib, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, **27**, 1547–1578.

French, M. N., W. F. Krajewski, and R. R. Cuykendall, 1992: Rainfall forecasting in space and time using a neural network. *J. Hydrol.*, **137**, 1–31.

Gautam, M. R., K. Watanabe, and H. Saegusa, 2000: Runoff analysis in humid forest catchment with artificial neural network. *J. Hydrol.*, **235**, 117–136.

Ghosh, S., and P. P. Mujumdar, 2007: Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*, **31**(1), 132–146, doi: 10.1016/j.advwatres.2007.07.005.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **16**, 147－168.

Gregory, J. M., T. M. L. Wigley, and P. D. Jones, 1993: Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. *Climate Dyn.*, **8**, 299–310.

Grotch, S. L., and M. C. MacCracken, 1991: The use of general circulation models to predict regional climatic change. *J. Climate*, **4**, 286–303.

Guo, S. L, J. X. Wang, L. H. Xiong, and A. W. Ying, 2002: A macro-scale and semi-distributed monthly water balance model to predict climate change impacts in China. *J. Hydrol.*, **268**, 1–15.

Haykin, S., 1994: *Neural Networks.* MacMillan College Publishing Company, New York, USA, 372pp.

Huth, R., 1999: Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research*, **13**, 91–101.

Joachims, T., 1999: Making large-scale support vector machine learning practical. *Advances in Kernel Methods-Support Vector Learning*, Schölkopf et al., Eds., MIT Press, Cambridge, MA, 169–184.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**(3), 437–471.

Karl, T. R., W. C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman, 1990: A method of relating general circulation model simulated climate to observed local climate. Part I: seasonal statistics. *J. Climate*, **3**, 1053–1079.

Kilsby, C. G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, and R. L. Wilby, 2007: A daily weather generator for use in climate change studies. *Environmental Modelling and Software*, **22**, 1705–1719.

Lee, Y., W. Hsieh, and C. Huang, 2005: SSVR: A smooth support vector machine for insensitive regression. *IEEE Transactions on Knowledge and Data Engineering*, **17**, 678–685.

Mangasarian, O. L., and D. R. Musicant, 1999: Successive overrelaxation for support vector machines. *IEEE Transactions on Neural Networks*, **10**(5), 1032–1037.

Mason, S. J., 2004: Simulating climate over Western North America using stochastic weather generators. *Climatic Change*, **62**, 155–187.

Obled, C., G. Bontron, and R. Garcon, 2002: Quantitative precipitation forecasts: A statistical adaptation of model outputs through an analogues sorting approach. *Atmospheric Research*, **63**, 303–324.

Olsson, J., C. B. Uvo, K. Jinno, A. Kawamura, K. Nishiyama, N. Koreeda, T. Nakashima, and O. Morita, 2004: Neural networks for rainfall forecasting by atmospheric downscaling. *Journal of Hydrologic Engineering*, **9**(1), 1–12.

Pang, B., S. L. Guo, L. H. Xiong, and C. Q. Li, 2007: A nonlinear perturbation model based on artificial neural network. *J. Hydrol.*, **333**, 504–516.

Platt, J., 1998: Fast training of support vector machines using sequential minimal optimization. Schölkopf et al., Eds., *Advances in Kernel Methods-Support Vector Learning*, MIT Press, Cambridge, MA, 185–208.

Rumelhart, D. E., E. Hinton, and J. Williams, 1986: Learning internal representation by error propagation. Volume 1, *Parallel Distributed Processing*, Rumelhart et al., Eds., MIT Press, Cambridge, MA, 318-362.

Srikanthan, R., and T. A. McMahon, 2001: Stochastic generation of annual, monthly and daily climate data: a review. *Hydrology and Earth System Sciences*, **5**(4), 653–670.

Tatli, H., H. N. Dalfes, and S. Mentes, 2004: A statistical downscaling method for monthly total precipitation over Turkey. *International Journal of Climatology*, **24**(2), 161–180.

Tripathi, S., V. V. Srinivas, and R. S. Nanjundiah, 2006: Downscaling of precipitation for climate change scenarios: A support vector machine approach. *J. Hydrol.*, **330**(3–4), 621–640.

Vapnik, V. N., 1995: *The Nature of Statistical Learning Theory.* Springer Verlag, New York, 314pp.

Vapnik, V. N., 1998: *Statistical Learning Theory.* Wiley, New York, 732pp.

von Storch, H., E. Zorita, and U. Cubasch, 1993: Downscaling of global climate change estimates to regional scales: An application to Iberian Rainfall in wintertime. *J. Climate*, **6**, 1161–1171.

Wetterhall, F., S. Halldin, and C.-Y. Xu, 2005: Statistical precipitation downscaling in Central Sweden with the analogue method. *J. Hydrol.*, **306**, 174–90.

Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling General Circulation Model output: a review of methods and limitations. *Progress in Physical Geography*, **21**, 530–548.

Wilby, R. L., L. E. Hay, and G. H. Leavesley, 1999: A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. *J. Hydrol.*, **225**(1–2), 67–91.

Wilby, R. L., R. J. Abrahart, and C. W. Dawson, 2003: Detection of conceptual model rainfall-runoff processes inside an artificial neural network. *Hydrological Science Journal*, **48**, 163–181.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns, 2004: The guidelines for use of climate scenarios developed from statistical downscaling methods. *The Procedures for the Preparation, Review, Acceptance, Adoption, Approval,and Publication of IPCC Reports*, 27pp.

Wilby, R. L., P. G. Whitehead, A. J. Wade, D. Butterfield, R.J. Davis and G. Watts, 2006: Integrated modelling of climate change impacts on water resources and quality in a lowland catchment: River Kennet, UK. *J. Hydrol.*, 330(1–2), 204−220.

Wilby, R. L., K. J. Beven, and N. Reynolds, 2007: Climate change and fluvial flood risk in the UK: More of the same? *Hydrological Processes,* **22**(14), 2511–2523.

Wilks, D. S., 1989: Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resour. Res.*, **25**, 1429–1439.

Xu, C.-Y., 1999: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography*, **23**, 229–249.

Yu, X. Y., and S. Y. Liong, 2007: Forecasting of hydrologic time series with ridge regression in feature space. *J. Hydrol.*, **332**, 290–302.

Zorita, E., and H. von. Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489.