Water Resour Manage (2016) 30:1515–1531
DOI 10.1007/s11269-016-1237-6

A Comparison of Nonlinear Stochastic Self-Exciting Threshold Autoregressive and Chaotic *k*-Nearest Neighbour Models in Daily Streamflow Forecasting

Hakan Tongal[1] · **Martijn J. Booij**[2]

Received: 29 January 2015 / Accepted: 14 January 2016 /
Published online: 27 January 2016

Abstract A nonlinear stochastic self-exciting threshold autoregressive (SETAR) model and a chaotic *k*-nearest neighbour (*k*-nn) model, for the first time, were compared in one and multi-step ahead daily flow forecasting for nine rivers with low, medium, and high flows in the western United States. The embedding dimension and the number of nearest neighbours of the *k*-nn model and the parameters of the SETAR model were identified by a trial-and-error process and a least mean square error estimation method, respectively. Employing the recursive forecasting strategy for the first time in multi-step forecasting of SETAR and *k*-nn, the results indicated that SETAR is superior to *k*-nn by means of performance indices. SETAR models were found to be more efficient in forecasting flows in one and multi-step forecasting. SETAR is less sensitive to the propagated error variances than the *k*-nn model, particularly for larger lead times (i.e., 5 days). The *k*-nn model should carefully be used in multi-step ahead forecasting where peak flow forecasting is important by considering the risk of error propagation.

Keywords Multi-step flow forecasting · Nonlinear stochastic self-exciting threshold autoregressive (SETAR) model · *k*-nearest neighbour (*k*-nn) model · Western United States

✉ Hakan Tongal
hakantongal@sdu.edu.tr

Martijn J. Booij
m.j.booij@utwente.nl

[1] Engineering Faculty, Department of Civil Engineering, Süleyman Demirel University, Isparta, Turkey

[2] Faculty of Engineering Technology, Department of Water Engineering and Management, University of Twente, Enschede, The Netherlands

1 Introduction

Streamflow forecasting is one of the most complicated tasks in hydrology owing to nonlinear dynamics in various physical mechanisms acting at a wide range of temporal and spatial scales. For several decades, a common belief among hydrologists has been that the tremendous variability observed in streamflow dynamics (and in other hydrological dynamics) resulted from a large number of dominant variables (Sivakumar 2003). Consequently, numerous works have been published regarding the application of stochastic theory in streamflow modelling and forecasting (Dutta et al. 2012; Latt and Wittenberg 2014; Patel and Ramachandran 2015). Although a plethora of applications of linear stochastic models is available for hydrologic modelling, their applications for forecasting of complex nonlinear phenomena are rather limited since they require the assumption of stationarity of time series which is not appropriate for operational forecasting and uncertainty assessment of nonlinear and non-stationary heteroskedastic processes (Komorník et al. 2006). In recent years, advancements in nonlinear time series modelling have brought us closer to understanding the nonlinear nature of hydrologic time series.

An example of a stochastic nonlinear model is the self-exciting threshold autoregressive (SETAR) model. SETAR allows us to model system dynamics by consideration of different sub-spaces (i.e., regimes) using threshold variables. In the literature, applications of the SETAR model to streamflow forecasting are rather limited. Its applicability to streamflow forecasting has to be proven and thoroughly tested, case by case, to deepen our understanding of nonlinear streamflow dynamics. Recently, Komorník et al. (2006) and Chen et al. (2008) applied SETAR to forecast streamflows. However, both studies do not investigate the forecasting capability of SETAR in daily forecasting of streamflows. While Komorník et al. (2006) employed SETAR to a monthly data set, Chen et al. (2008) handled 10-day streamflow data set by moving averaging of daily flows. Thus, utilization of SETAR in forecasting daily streamflows is an important task and need further investigation.

In recent years, with the advent of chaos theory it has become increasingly realized that irregular random behaviour could be the outcome of simple interdependent deterministic systems influenced by a few nonlinear variables. The applications of chaos theory to streamflow dynamics enhance our understanding towards underlying dynamics of generating mechanisms of streamflows. It is possible to determine the number of dominant variables if the underlying dynamics are low dimensional chaos (e.g., Ng et al. 2007; Tongal et al. 2013). Further, studies have shown that low-dimensional chaotic approaches give relatively accurate predictions of runoff series (Liu et al. 1998; Porporato and Ridolfi 1997; Sivakumar 2003). However, there are few studies on the comparison of forecasting performances of stochastic and chaotic approaches. Among others, Jayawardena and Gurung (2000) predicted three synthetic data sets and four real data sets with chaotic and stochastic approaches and concluded that the chaotic model is superior to the stochastic approach. Lisi and Villi (2001) compared the forecasting performance of chaotic and linear ARIMA models and showed that the chaotic approach improved the forecasting performances of ARIMA. In the literature, to the authors' knowledge, Yu et al. (2004) is the only study that concludes a conventional chaotic approach is worse than ARMA(1,1) and ARIMA(1,1,1). Furthermore, the authors claim that a Naïve forecasting method in which the forecast at time $t+1$ is assumed to be equal to the forecast at time t, performed better than the chaotic prediction method. These results gave inspiration to this study. Firstly, from a pragmatic engineering point of view, accurate forecasting is usually the primary motivation for developing engineering models. Secondly,

less computationally expensive models are more preferred than expensive ones. From the above results, one might wonder that (1) if simple ARMA(1,1) and Naïve forecasting methods give more reliable and generalizable results than the chaotic prediction method, then why will one prefer to use more complex chaotic prediction methods? (2) Is there any necessity to use a chaotic prediction method in forecasting streamflows? Finally, (3) while the linear stochastic models (ARMA, ARIMA) was reported to give better results than a nonlinear chaotic method (Yu et al. 2004), will rarely used nonlinear SETAR model outperform a chaotic prediction method in forecasting streamflows?

In this study, to answer these questions and to gain more insights into forecasting performances of a nonlinear stochastic time series model (SETAR) and a chaotic prediction method (a local *k*-nn model), nine streamflow time series with different mean flows are modelled at daily time scales. To the best of our knowledge, this is the first study to apply a SETAR and a local *k*-nn model jointly in daily and multi-step forecasting of streamflows with different hydrological characteristics. Nine streamflow time series in the western United States are selected from the study of Sivakumar (2003) in which they are categorized as low, medium and high flows according to the mean flow value.

In this study, by comparing the performances of stochastic (SETAR) and chaotic models (*k*-nn), we also try to add a contribution to the following comment of Sivakumar (2003) 'Consequently, a general assessment regarding whether one approach [i.e., chaotic or stochastic approach] is better than the other or vice-versa is difficult to provide. It must be emphasized, however, that studies that have employed both approaches have revealed that the chaotic approach was better than the stochastic approach for the streamflow series analysed…The author is not aware of any studies that have reported the opposite situation'. However, it must be noted that the findings of this study are limited to the *k*-nn and SETAR models. We do not intend to reach such a conclusion that the stochastic or chaotic approach gives a better forecasting performance, but rather show the capabilities of both models in forecasting daily streamflows for one and multi-step ahead forecasting by only using past observed discharges as inputs.

2 Streamflow Data

The data used in the analyses were obtained from the study of Sivakumar (2003) where the author categorized the streamflows of 79 gauging stations in the western United States into three groups as low, medium and high flows according to the value of the mean flow as follows: (1) low-flow stations with mean streamflow values of less than 2.8 m^3s^{-1}, (2) high-flow stations with mean streamflow values of more than 28.3 m^3s^{-1} and (3) medium-flow stations with mean streamflow values between the above two values. Nine stations from these three groups were selected ensuring sufficient variability among drainage basin areas, elevations and mean flows (Table 1).

Because of different climatic regions and drainage basin characteristics, the magnitude of streamflows varies greatly among the stations. Most of the drainage basins are medium to small sized and are located in areas with middle to high elevations (mostly higher than 500 m) (Sivakumar 2003). The climate is tender from south (desert climate) to north (more rainy regions) and to represent climate variability, low-medium-high flow stations were chosen from the south-middle-north regions, respectively. The selected regions are: CA, CO/WY, ID, MT, NM, NV, OR, UT and WA. The nine stations were randomly selected as one station to be in

Table 1 Statistics of streamflow data obtained from Sivakumar (2003)

Stream flow	Serial no.	State	Area (km^2)	Elevation of Station (m)	Discharge Mean (m^3s^{-1})	Std. Deviation (m^3s^{-1})	Coefficient of variation	Min. (m^3s^{-1})	Max. (m^3s^{-1})	Station Number
Low-Flow	01	CA	23	1067	0.15	0.59	3.93	0.00	29.73	11058500
	02	UT	648	2298	1.93	3.26	1.69	0.03	40.21	10131000
	03	NM	490	3021	2.83	4.15	1.48	0.17	56.07	8378500
Medium-Flow	04	CO/WY	3706	2713	11.33	16.68	1.47	0.43	178.11	6620000
	05	NV	13087	1966	11.62	19.78	1.70	0.09	221.44	10322500
	06	ID	1476	1484	19.46	31.45	1.62	0.65	252.87	12306500
High-Flow	07	WA	831	1564	40.35	45.82	1.14	2.46	509.70	12451000
	08	OR	13183	1341	55.38	74.56	1.35	0.18	1030.73	14046500
	09	MT	30580	939	196.62	231.91	1.18	12.74	2268.18	6214500

The states are: *CA* California, *UT* Utah, *NM* New Mexico, *CO-WY* Colorado-Wyoming, *NV* Nevada, *ID* Idaho, *WA* Washington, *OR* Oregon and *MT* Montana

each of these regions. All flow time series span from 01.01.1932 to 31.12.2010. The data were divided into three subsets for calibration, which includes training and testing data sets, and validation. The first 70 % of the record is designated for calibration, 80 % of the calibration data set was assigned as training set and 20 % was assigned as testing set, and the last 30 % of the data set was assigned for the validation.

3 Methodology

3.1 BDS Test

The BDS test (Brock et al. 1987) is a nonparametric method to test for the existence of nonlinearity in a time series. It is particularly useful for chaotic and nonlinear stochastic systems of which effectiveness is shown in various studies (e.g., Lisi and Villi (2001), Amendola et al. (2006)). A scalar time series $\{X_t\}$ of length N should be reconstructed into a m-dimensional space to obtain a new series $\{Y_t\}$, with $Y_t=(X_t, X_{t-1}, \ldots, X_{t-(m-1)\tau})$, where τ is a delay time. For each couple of parameters m and r, the correlation integral is used to examine the dependence of the series that can be computed as follows (Grassberger and Procaccia 1983):

$$C(m,r,N) = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=1}^{M} H\left(r - \| Y_i - Y_j \|\right) \tag{1}$$

where $M=N-(m-1)\tau$ is the number of embedded points in a m-dimensional space, H is the Heaviside step function, with $H(u)=1$ for $u>0$, and $H(u)=0$ for $u\leq 0$ where $u=r-\|Y_i-Y_j\|$, r is the radius of a sphere centered on Y_i or Y_j, and N is the number of data points. The sup-norm $\|\bullet\|$, can be any of the three usual norms, the maximum norm (maximum absolute difference between Y_i and Y_j), the diamond norm (sum of all the absolute differences) or the standard Euclidean norm. It is then possible to estimate the BDS statistic

$$BDS = \sqrt{\mathrm{M}}(C(m,r,N) - C(1,r,N)^m)/\sqrt{\mathrm{V}} \tag{2}$$

where V is the variance of $C(m,r,N)$ under the null hypothesis. The analytical expression of V can be found in Cromwell et al. (1994). Under the null hypothesis of an independent and identical distribution, the BDS statistic is asymptotically normal with zero mean and unit variance (Brock et al. 1987; Lisi and Villi 2001).

3.2 Self-Exciting Threshold Autoregressive Models

First proposed by Tong (1978), the threshold autoregressive (TAR) model describes a given stochastic process by a class of linear auto-regressions, where the output of linear auto-regression at any instant depends upon the value taken by a threshold variable (Clements and Krolzig 1998). TAR is considered as SETAR when the threshold variable is taken as a lagged value of the considered time series. Let y_t be a discharge time series that follows a l -regimes threshold autoregressive process with threshold variable y_{t-d}:

$$y_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} y_{t-p} + \varepsilon_t^{(j)} \text{if} \quad r_{j-1} \leq y_{t-d} < r_j \tag{3}$$

where, p is the degree of autoregressive, l and d are positive integers, $j = 1, 2, 3, .., l$, r_i are real numbers that satisfy the following condition:

$$-\infty = r_0 < r_1 < \ldots < r_{l-1} < r_l = \infty \tag{4}$$

the superscript (j) is used to identify the regime, $\{\varepsilon_t^{(j)}\}$ are independent identical distributed sequences with zero mean and variance σ_j^2 and are mutually independent for different j. The parameter d is the delay parameter, r_j are thresholds and ϕ_i are autoregressive coefficients. Equation (3) indicates that a SETAR model is a piecewise linear AR model in the threshold space. Details of determining the structural parameters $\{\phi, r, d\}$ can be found in Pinson et al. (2008).

3.3 *K*-nearest Neighbour Model

The k-nearest neighbour model (k-nn) is a local approximation method which uses nearby states to make nonlinear predictions (Sharifazari and Araghinejad 2015). Let $\overline{y}_t^m = \left(y_t, \ldots, y_{t-m+1}\right)$ be a feature vector of past discharge records that summarises the whole history in a way that it is supposed to contain most of the information related to the forecast. It is assumed that the probability distribution of the observed variable conditioned on the entire past is the same for the observed variable conditioned on only the past observations. In other words, the probability distribution of $(y_{t+1}|y_t, y_{t-1}, \ldots)$ is the same as the probability distribution of $\left(y_{t+1}\middle|\overline{y}_t^m\right)$. Even if the $\overline{y}_t^m$ does not satisfy the above mentioned condition, as Karlsson and Yakowitz (1987) proved, the k-nn forecaster will be asymptotically optimal among all the forecasters defined on $\overline{y}_t^m$ (Toth et al. 2000).

The expectation of the forecasted value can be written as conditioned on the feature vector as:

$$\hat{y}_{t+1} = E\left[y_{t+1}\middle|\overline{y}_t^{\,m}\right] \tag{5}$$

To estimate $\hat{y}_{t+1}$, the k-nn model employs a metric ($\|\cdot\|$) defined on the feature vector to find the set of k observed nearest neighbours of $\overline{y}_t^m$, i.e., the k-dimensional vectors of past observations as $\overline{y}_{t_j}^m$ which minimizes:

$$\left\|\overline{y}_t^{\,m} - \overline{y}_{t_j}^{\,m}\right\| \tag{6}$$

where $j = 1, 2, \ldots, k$. One of the most employed metrics to identify the number of nearest neighbours is the Euclidean norm. If R^m represents a vector of coordinates $\xi_1, \xi_2, \ldots, \xi_m$, the differences between the current feature vector and past data, the Euclidean distance, is defined as:

$$\|R\| = \left(\sum_{i=1}^{m} \xi_i^2\right)^{1/2} \tag{7}$$

For one-step ahead forecasting, the estimation of $\hat{y}_{t+1}$ can then be taken as the average of the temporal evolution (i.e., the successive values) of the most resembling k-past values as:

$$\hat{y}_{t+1} = \frac{1}{k}\sum_{j=1}^{k} y_{t_j+1} \tag{8}$$

Intuitively speaking, the forecast $\hat{y}_{t+1}$ by the *k*-nn model is the sample average of the succeeding values of the *k* nearest neighbours in the data set. For more detailed information the reader can refer to Toth et al. (2000).

To select appropriate models from their classes and to evaluate forecasting performances following performance indices were used: the Akaike Information Criterion $AIC(p_1,p_2) = \ln\sigma^2 + 2(p_1+p_2)/N$, the Nash-Sutcliffe efficiency coefficient $NSE = 1-\sum_{i=1}^{N}(Q_i-Q_i)^2/\sum_{i=1}^{N}\left(Q_i-\overline{Q}\right)^2$, a modified form of this $NSE_m = 1-\sum_{i=1}^{N}|Q_i-Q_i|/\sum_{i=1}^{N}\left|Q_i-\overline{Q}\right|$, the percent bias $PBIAS\ (\%) = 100\times\sum_{i=1}^{N}Q_i-Q_i/\sum_{i=1}^{N}Q_i$, the index of agreement $d = 1-\sum_{i=1}^{N}\left(Q_i-\hat{Q}_i\right)^2/\sum_{i=1}^{N}\left(\left|\hat{Q}_i-\overline{Q}\right|+\left|Q_i-\overline{Q}\right|\right)^2$ and the root mean square error $RMSE = \left\{N^{-1}\sum_{i=1}^{N}\left(Q_i-\hat{Q}_i\right)^2\right\}^{0.5}$ where N is the number of data points, p_1 and p_2 are the appropriate model orders for two different regimes, $\hat{\sigma}^2$ is the variance of residuals, Q_i, $\hat{Q}_i$ and $\overline{Q}$ are the observed, forecasted and the mean of the observed discharge values, respectively. Detailed information about these indices can be found in Legates and McCabe (1999), Krause et al. (2005), and Ajmal et al. (2015).

4 Results and Discussion

4.1 Autocorrelations, Mutual Information Functions and BDS Test

Data characteristics were firstly analysed through the autocorrelogram and the standardized mutual information function as depicted in Fig. 1. The faster decay of the autocorrelation function compared to the mutual information function is an indication of the nonlinear information is more persistent than the linear information in the daily time series.

To confirm that nonlinear mechanisms are more prevalent than linear mechanisms, we applied the BDS test to the streamflow time series. Before applying the test, since the daily streamflow data exhibit strong interdependency (Fig. 1), linear dependencies were reduced within the data by pre-processing as suggested by Wang et al. (2006). All logarithmized streamflow time series were standardized and then pre-whitened with autoregressive (AR) models of which degrees were determined according to minimum AIC values (Table 2). Then, the BDS test was applied to the obtained residuals from these models (Table 2). Generally, the BDS statistic is calculated for a set of parameters (m,r). In this study, these parameters were chosen as $m=2,3,4$ and $r=0.1\sigma$, 0.3σ, 0.5σ, 0.75σ, 1.25σ, 1.5σ similarly to Lisi and Villi (2001). The results of the BDS test showed that the asymptotically normal statistic varied between 48.9 and 64.7 which is larger than standard normal statistics (i.e., $z_{critical}=1.96$) and therefore the hypothesis of independence (all $p<0.05$) are strongly rejected, indicating all daily streamflow time series are nonlinear.

4.2 Self-Exciting Threshold Autoregressive Model

Various studies have shown that SETAR models having one threshold variable and thus two regimes (i.e., low and high dynamics that are below and above a threshold) are quite efficient in modelling nonlinear hydrological time series (e.g., Amendola et al. 2006; Chen et al. 2008; Komorník et al. 2006). To reduce complexity originating from transition between the regimes, the model parameterization and computing time, the number of regimes was fixed to two. The maximum allowed model degrees in each regime were determined with the partial autocorrelation function (PACF).

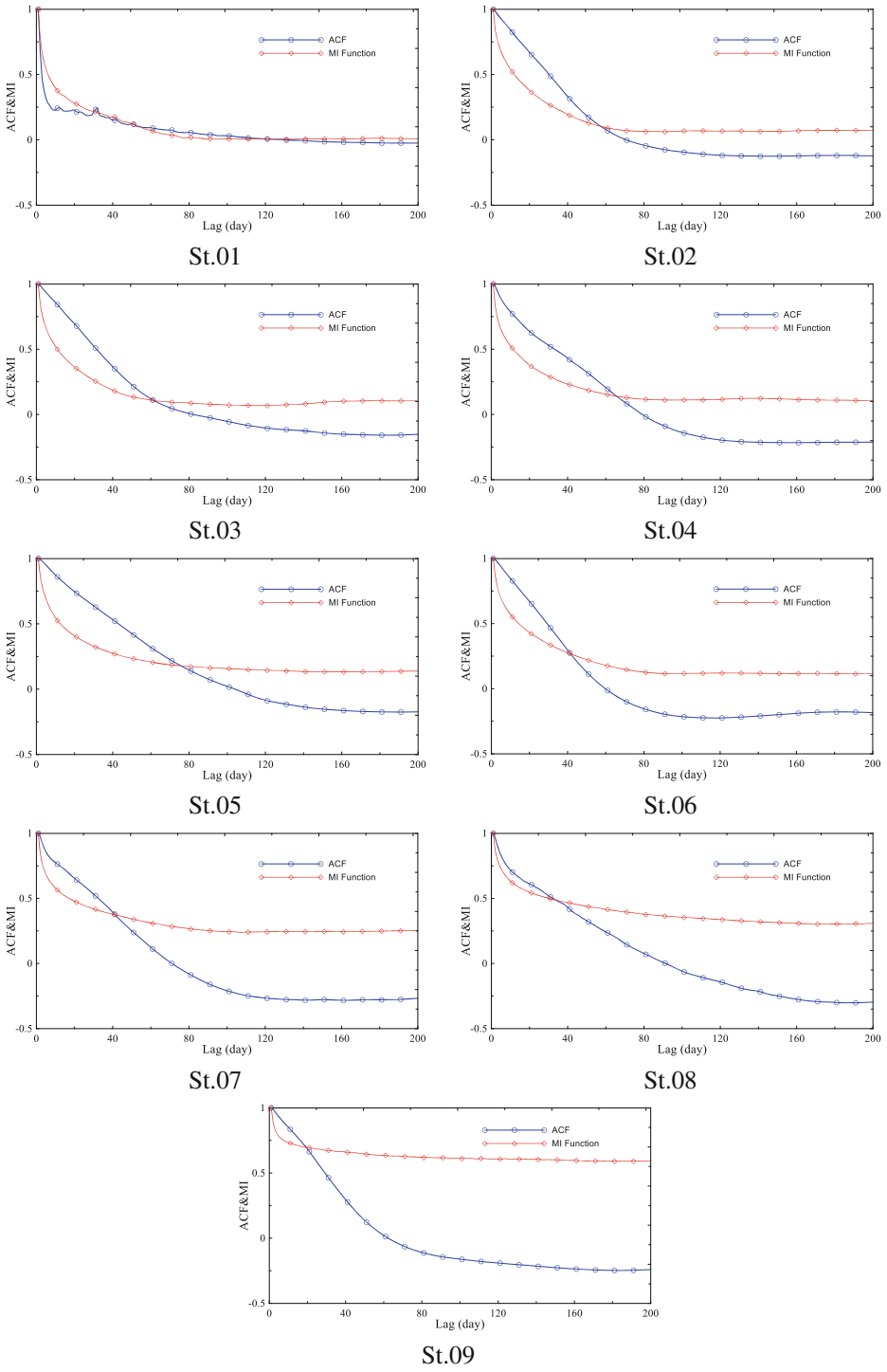

Fig. 1 Autocorrelation and standardized mutual information functions of nine daily streamflows

Table 2 Degree of AR models constructed for streamflow time series, BDS test results and SETAR models with threshold values

Station	AR degree	$m = 2$		$m = 3$		$m = 4$		SETAR models	Threshold (m^3s^{-1})
		Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value		
St.01	43	53.6	0	53.3	0	51.2	0	SETAR (2,3,2)	0.23
St.02	42	53.7	0	60.9	0	64.0	0	SETAR (4,3,2)	5.15
St.03	37	54.6	0	57.3	0	57.7	0	SETAR (2,3,2)	6.32
St.04	42	48.9	0	56.8	0	60.3	0	SETAR (4,5,2)	31.15
St.05	42	54.8	0	59.3	0	61.2	0	SETAR (5,5,2)	31.15
St.06	43	60.0	0	62.7	0	62.2	0	SETAR (4,4,2)	68.24
St.07	43	64.4	0	64.7	0	62.3	0	SETAR (2,5,2)	104.20
St.08	43	56.3	0	59.9	0	60.2	0	SETAR (2,3,2)	141.90
St.09	42	55.3	0	61.6	0	64.2	0	SETAR (3,4,2)	461.60

The empirical procedure described by Tong (1983) led to models that minimize the AIC (Table 2) for the delay parameter, $d = 1$. Additionally, likelihood ratio test for threshold nonlinearity (see Chan 1991; Chan and Tong 1986; Moeanaddin and Tong 1988) was applied to evaluate the appropriateness of the SETAR model for the streamflows, with the null hypothesis of being a normal autoregressive process and the alternative hypothesis of being a threshold autoregressive process. The obtained likelihood ratio statistic (the results are not shown here) varies from 49.08 to 415.47 and all corresponding *p*-values are statistically significant (i.e., lower than 0.05) which indicates the necessity of SETAR for streamflows time series.

For daily time series (Fig. 2), all SETAR models showed good forecasting performances except for St.01, caused by the model parameters of the upper regime. As can be seen in Fig. 2., the forecasted values above the threshold (i.e., 0.23 m^3s^{-1}) are constantly below the observed values for St.01. When examining the high and low flows, all SETAR models successfully forecasted peak flows except St.01, and successfully captured the low flow dynamics.

The adequacy of the SETAR model in forecasting streamflow dynamics can be confirmed in Table 3. *NSE*, *d* and *RMSE* indices are more sensitive to peak flows due to their squared forms. All these indices showed that SETAR models are efficient in high flow forecasting. There is not any effect of stations being in a low, medium or high flow region on performance indices except *RMSE*.

The *RMSE* index is highly dependent on the magnitude of the flow and it is not appropriate to compare this index between stations, but only for the same station. *PBIAS* values indicate the over or underestimation of the forecasts. Since the *PBIAS* values of daily time series are lower than ±25 % (Makungo et al. 2010), it can be concluded that all SETAR models are acceptable in daily forecasting. The NSE_m index is an overall sensitivity measure for the quality of the model results. As a more global measure, the NSE_m index stands in the middle between the squared forms and the relative forms of performance indices (Krause et al. 2005). The NSE_m index is between around 0.70 and 0.95 for all stations. It should be noted that it is rather difficult to obtain higher values for the NSE_m than the original form. Therefore, overall it can be said that the SETAR models are quite efficient in daily streamflow forecasting in terms of considered performance indices.

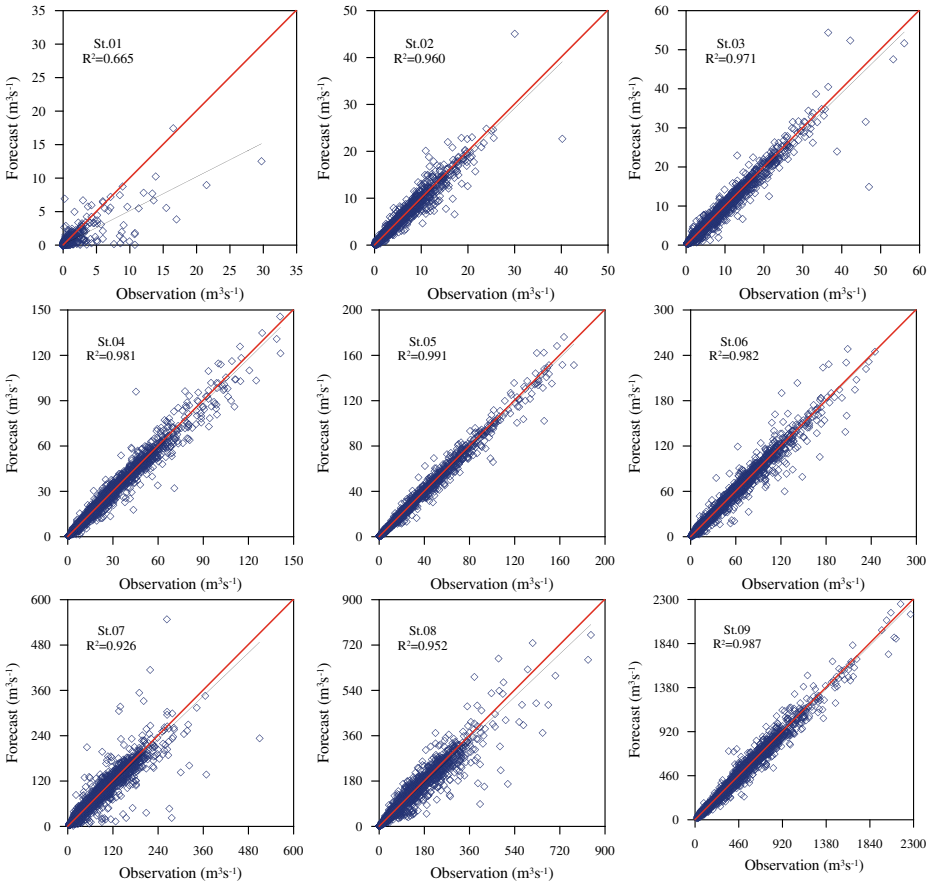

Fig. 2 Scatter diagrams of daily observed and forecasted streamflow values for the self-exciting threshold autoregressive (SETAR) model

4.3 *K*-nearest Neighbour Model

The necessary parameters for building up a phase-space forecasting or *k*-nearest neighbour model are usually found by chaotic approaches such as the correlation dimension method and the false nearest neighbour method. However, since we aim to compare the forecasting performances of SETAR and *k*-nn rather than investigating chaotic dynamics in streamflow, the necessary parameters *m* and *k* were determined by a trial-error process employed for the training set that minimizes the *RMSE* calculated in the testing set.

The embedding dimensions were varied from $m=2$ to $m=10$ and *k* was varied from 1 to 100. A lower bound of two was selected since at least one delay coordinate is needed in the state vector. Also, it is not meaningful to search for a higher embedding dimension than 10, because only low to moderate dimensional chaos is amenable to short-term prediction (Phoon et al. 2002). All the minimal *RMSE* values were obtained for $m=2$.

We would expect to obtain a higher number of nearest neighbours (Table 3) that have higher dynamics (i.e., high BDS statistic) and vice versa, but the number of nearest neighbours was found to be independent from having higher or lower dynamics. In one aspect, it is

Table 3 Forecasting performances of daily SETAR and *k*-nn models with optimal number of nearest neighbours (*k*)

	Perf. Indices	St.01	St.02	St.03	St.04	St.05	St.06	St.07	St.08	St.09
SETAR	NSE (−)	0.665	0.960	0.971	0.981	0.991	0.982	0.926	0.952	0.987
	d (−)	0.845	0.990	0.993	0.995	0.998	0.995	0.981	0.987	0.997
	RMSE (m^3s^{-1})	0.477	0.551	0.732	2.019	1.676	3.846	11.916	15.718	25.867
	PBIAS (%)	−9.672	0.165	−0.164	0.198	−0.091	0.362	0.340	0.072	0.061
	NSE_m (−)	0.691	0.883	0.901	0.916	0.950	0.937	0.875	0.894	0.928
	Perf. Indices	St.01 k = 11	St.02 k = 8	St.03 k = 10	St.04 k = 7	St.05 k = 3	St.06 k = 24	St.07 k = 66	St.08 k = 29	St.09 k = 82
k-nn	NSE (−)	0.448	0.941	0.949	0.936	0.973	0.960	0.886	0.904	0.956
	d (−)	0.738	0.985	0.987	0.983	0.993	0.990	0.969	0.974	0.988
	RMSE (m^3s^{-1})	0.584	0.672	0.974	3.724	2.934	5.632	14.742	22.110	47.113
	PBIAS (%)	−7.589	0.428	−0.158	0.054	0.123	0.311	0.259	0.002	−0.312
	NSE_m (−)	0.604	0.839	0.855	0.842	0.899	0.888	0.813	0.832	0.869

reasonable that the performance of the employed *k*-nn model is independent from the system dynamics. It only depends on the *k*-familiar observations of the considered value of which succeeding values need to be forecasted.

The *k*-nn model performance largely depends on the selection of the number of nearest neighbours and the similarity between calibration and validation series. For the daily forecasts (Fig. 3), the *k*-nn model failed to capture peak values since these peak values were not observed in the calibration set and/or the forecasted peak values are lower than the observed ones because of the "averaging effect" which can be explained as follows. For instance, if theoretically, in St.06 there are four observed equal values to the forecasted value and the number of nearest neighbours is 24, then the remaining 20 observed values will decrease the forecasted value.

A good overall match between the daily observed and forecasted values can be seen. Especially, the low streamflows were well forecasted due to the frequent occurrences of low streamflows that allowing the trained model to have a better generalization of these values.

For the validation period, the *NSE* index indicates a "very good" forecasting performance except for St.01 (Table 3) for which we obtained a poor forecasting performance. The *d* and

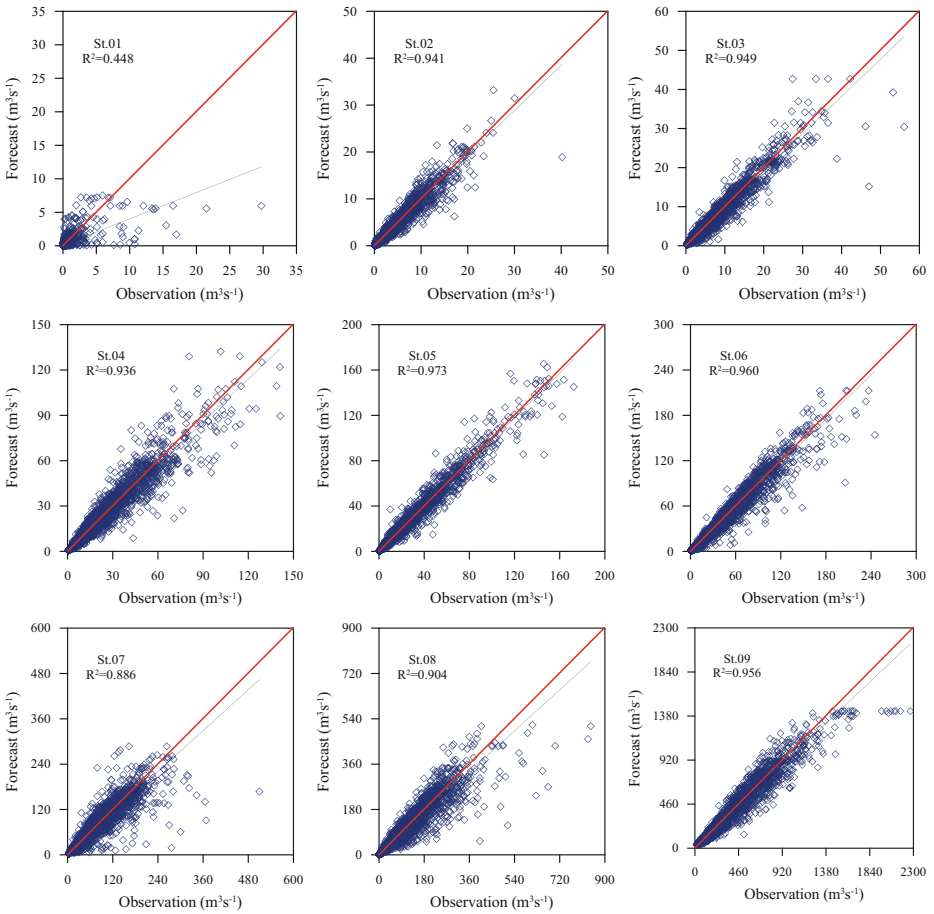

Fig. 3 Scatter diagrams of daily observed and forecasted streamflow values for the validation set for the *k*-nearest neighbour (*k*-nn) model

NSE indices are sensitive to peak flows and therefore relatively lower values were obtained for St.01. The *PBIAS* index indicates "very good" forecasting performances except for St.01. For this station, *PBIAS* indicates a "good" forecasting performance. Although it is rather difficult to obtain higher values for NSE_m, we obtained values above 0.80 for all stations except for St.01.

To understand whether climate change led such a poor forecasting performance in St.01, we applied Mann-Kendall trend analysis to the monthly discharges of stations (not shown here) and any statistically meaningful increasing or decreasing trend, except St.02, St.04 and St.05 where models were well-performed, was not found. Thus, the poor forecasting performance for St.01 could be partially explained by its distinct hydrological and geological characteristics as follows. The climate of San Bernardino region, where St.01 is located, is broadly characterized as Mediterranean but form of precipitation changes significantly with elevation (Hawley and Bledsoe 2011). Distinct from the other stations, winter frontal storms which can be intense, lead to flashy regime with short-lived instantaneous peak flows in St.01 coupled with steep topography and narrow mountain valleys (Modrick and Georgakakos 2014). Additionally, in the mountainous regions of the western United States, streamflow derived from snow is very sensitive to changes in temperature in compare to elsewhere in the United States (Safeeq et al. 2013). The characteristics of mean annual hydrograph of the stations can be informative about the basin characteristics and annual variations in temperature, precipitation and snowpack (Kunkel and Pierce 2010). Only annual hydrograph of St.01 (not shown here) shows complex-hydrograph features with fluctuations over short periods in both the climbing and the falling limbs (Kunkel and Pierce 2010) among the other stations. Additionally, the highest coefficient of variation (CV) value (defined as the standard deviation divided by the mean) that shows variability of a time series was obtained for this station (Table 1). It is obvious that; being small catchment that is more non-linear than larger ones with the highest CV, showing complex-hydrograph features, and flash flood occurrences coupled with sensitiveness to snow-melting led poor forecasting performances in St.01. Thus, *k*-nn and SETAR models could be less useful for modelling such a high-complex dynamics.

4.4 Multi-Step Forecasting

We applied *Recursive Forecasting Strategy* which has been widely used in multi-step forecasting of hydrological time series (e.g., Aqil et al. (2007), Chen et al. (2013)). In this method, observed sample points of the time series are gradually replaced by previously predicted time series at each forecasting step.

The SETAR models yielded better results than the *k*-nn models for up to five-day ahead streamflow forecasting (Table 4). Since the forecasted outputs were fed back into the models to forecast further values, the forecast error variances also increased as the forecast horizon increased, as expected. The obtained multi-step forecasts (i.e., 2-day, 3-day, 4-day and 5-day) of streamflows for one low, one medium and one high flow station where both models show acceptable performances (i.e., St.03, St.05 and St.09), can be seen in Fig. 4.

The SETAR models better responded to most of the fluctuations within the data in addition to the peak flows. In multi-step forecasting, the SETAR models are less sensitive to the propagated error variances than the *k*-nn models and the SETAR model is still superior over the *k*-nn model. This can be best observed for the peak flow forecasts. The propagated errors led to a decrement in the peaks as the forecast horizon increases, especially for St.09. In addition, the forecasts of the SETAR models are less scattered around the 45° fit, which

Table 4 Performance indices of multi-step ahead daily forecasts obtained by SETAR and *k*-nn models

Models	Performance Indices	Setar				*k*-nn			
		2-day	3-day	4-day	5-day	2-day	3-day	4-day	5-day
St.01	NSE_m (−)	0.606	0.554	0.485	0.436	0.505	0.414	0.364	0.322
	d (−)	0.729	0.655	0.601	0.483	0.659	0.479	0.384	0.336
	RMSE (m^3s^{-1})	0.581	0.623	0.644	0.696	0.646	0.716	0.743	0.755
St.02	NSE_m (−)	0.849	0.824	0.797	0.777	0.789	0.744	0.711	0.683
	d (−)	0.984	0.980	0.974	0.966	0.976	0.966	0.957	0.949
	RMSE (m^3s^{-1})	0.691	0.757	0.858	0.990	0.843	0.991	1.107	1.198
St.03	NSE_m (−)	0.864	0.833	0.804	0.775	0.802	0.746	0.704	0.665
	d (−)	0.985	0.982	0.971	0.961	0.975	0.961	0.948	0.936
	RMSE (m^3s^{-1})	1.051	1.131	1.429	1.651	1.318	1.631	1.874	2.066
St.04	NSE_m (−)	0.873	0.830	0.799	0.770	0.783	0.717	0.675	0.642
	d (−)	0.988	0.979	0.971	0.964	0.969	0.946	0.931	0.917
	RMSE (m^3s^{-1})	3.108	4.182	4.832	5.285	5.071	6.558	7.333	7.952
St.05	NSE_m (−)	0.926	0.902	0.881	0.862	0.863	0.815	0.774	0.738
	d (−)	0.995	0.992	0.988	0.985	0.988	0.977	0.967	0.956
	RMSE (m^3s^{-1})	2.461	3.177	3.862	4.252	3.881	5.272	6.335	7.243
St.06	NSE_m (−)	0.904	0.878	0.849	0.820	0.843	0.791	0.751	0.714
	d (−)	0.990	0.986	0.978	0.973	0.981	0.968	0.957	0.946
	RMSE (m^3s^{-1})	5.590	6.681	8.270	9.212	7.626	9.794	11.310	12.629
St.07	NSE_m (−)	0.822	0.787	0.748	0.705	0.748	0.683	0.642	0.608
	d (−)	0.966	0.961	0.948	0.929	0.949	0.926	0.910	0.896
	RMSE (m^3s^{-1})	15.524	16.697	18.858	22.138	18.844	22.234	24.236	25.798
St.08	NSE_m (−)	0.843	0.805	0.765	0.733	0.774	0.710	0.665	0.626
	d (−)	0.974	0.964	0.953	0.944	0.955	0.931	0.912	0.895
	RMSE (m^3s^{-1})	22.169	25.701	29.039	31.716	28.621	35.063	39.126	42.244
St.09	NSE_m (−)	0.889	0.860	0.827	0.797	0.820	0.767	0.731	0.698
	d (−)	0.992	0.988	0.980	0.973	0.980	0.967	0.957	0.946
	RMSE (m^3s^{-1})	39.941	48.561	62.050	71.203	61.599	77.941	88.438	98.075

describes the degree collinearity between observed and forecasted data, in contrast to the *k*-nn models.

5 Conclusions

The purpose of this study is to compare two nonlinear models, SETAR and *k*-nn, that belong to two different modelling approaches (stochastic and chaotic) for nine rivers in the western United States with low, medium and high flows. To the authors' knowledge, SETAR and *k*-nn models have so far not been tested and compared for forecasting streamflows, jointly. The autocorrelation, mutual information functions and nonlinear BDS test indicated that nonlinear mechanisms are prevailing over linear ones in streamflows. Since the number of nearest neighbours of the *k*-nn model was found to be independent from river flow dynamics, the SETAR model has an advantage in analysing basin dynamics resulting from the model

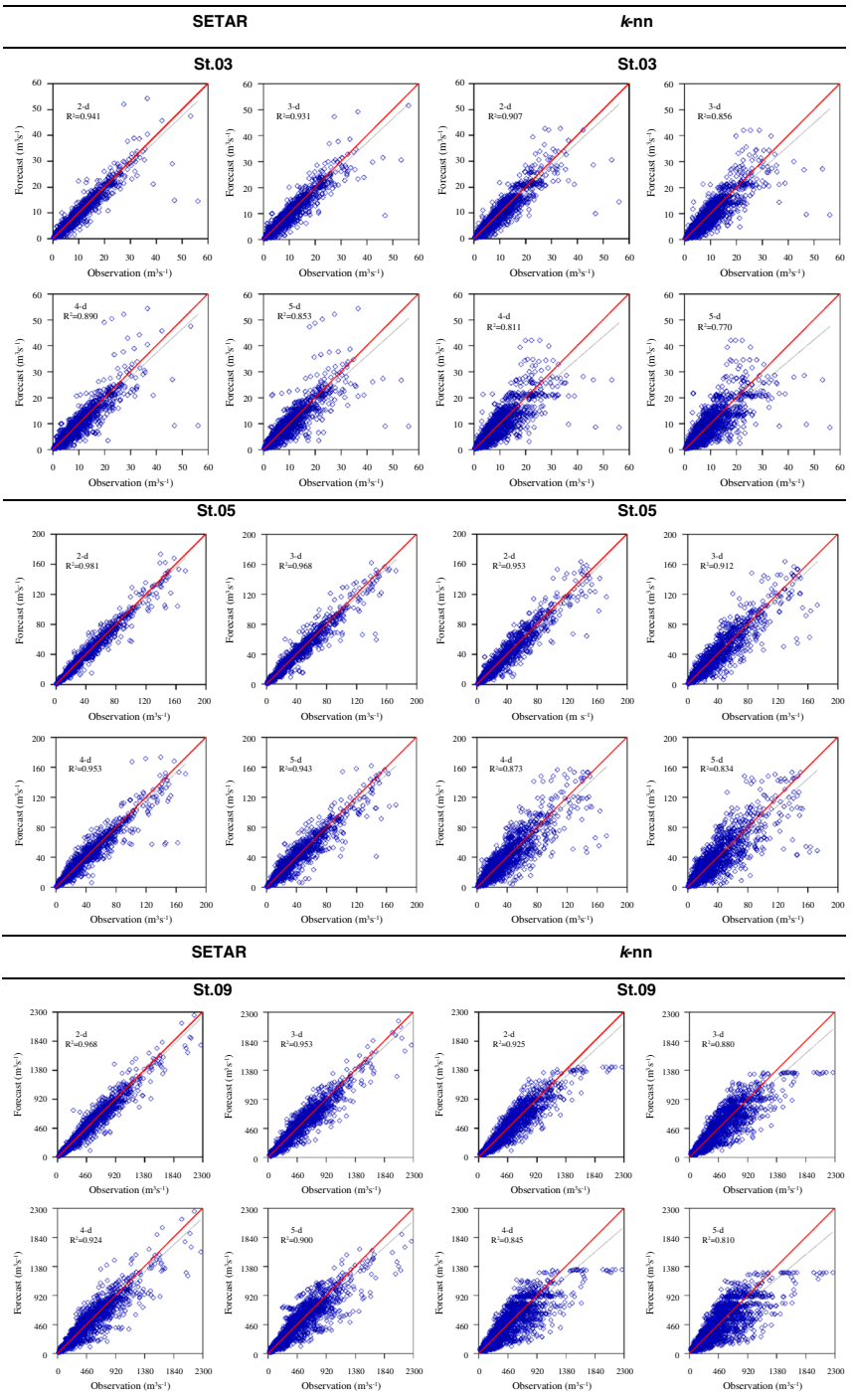

Fig. 4 Scatter plots of observed and multi-time-step forecasted discharges of St.03, St.05 and St.09 by the SETAR and *k*-nn models

structure. The one-day and the recursive multi-step ahead forecasting results showed a clear superiority of the SETAR model over the *k*-nn model. The multi-step ahead forecasting results also showed that the SETAR model is less sensitive to the propagated error variances compared to the *k*-nn model which can be best observed for peak flow forecasts for higher forecast horizons (e.g., 5-day ahead forecasts). The *k*-nn model should be carefully used in multi-step ahead forecasting when peak flow forecasting is important.

In conclusion, the analyses showed that both SETAR and *k*-nn models have the potential to forecast streamflows for one-day and multi-days ahead, but that the SETAR model is superior to the *k*-nn model for various forecast horizons up to 5 days. However, in a pluralistic modelling culture, each modelling strategy reveals some aspects of the data and hydrological processes that other models may overlook and an important aim of modelling is to gain insight into the underlying contributory factors. 'In the pluralistic culture, models are not regarded as expressions of truth but constructs to explain the data' (Khatibi et al. 2014). Therefore, it is important to determine whether the differences in forecasting performances of the models in different hydrological time series are statistically significant in a further study.

Acknowledgments The authors would like to thank the U.S. Geological Survey for providing data used in this study.

References

Amendola A, Niglio M, Vitale C (2006) Multi-step SETARMA predictors in the analysis of hydrological time series. Phys Chem Earth A/B/C 31:1118–1126

Ajmal M, Waseem M, Ahn J-H, Kim T-W (2015) Improved runoff estimation using event-based rainfall-runoff models. Water Resour Manag 29:1995–2010. doi:10.1007/s11269-015-0924-z

Aqil M, Kita I, Yano A, Nishiyama S (2007) Neural networks for real time catchment flow modeling and prediction. Water Resour Manag 21:1781–1796. doi:10.1007/s11269-006-9127-y

Brock W, Dechert WD, Scheinkman J (1987) A test for independence based on the correlation dimension. Working paper, University of Wisconsin

Chan KS (1991) Percentage points of likelihood ratio tests for threshold autoregression. J R Stat Soc Ser B Methodol 53:691–696. doi:10.2307/2345598

Chan KS, Tong H (1986) On estimating thresholds in autoregressive models. J Time Ser Anal 7:179–190

Chen Y, Chang L, Huang C, Chu H (2013) Applying genetic algorithm and neural network to the conjunctive use of surface and subsurface. Water Res Manag 27:4731–4757. doi:10.1007/s11269-013-0418-9

Chen C-S, Liu C-H, Su H-C (2008) A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. Hydrol Process 22:3697–3711. doi:10.1002/hyp.6973

Clements MP, Krolzig H-M (1998) A comparison of the forecast performance of markov-switching and threshold autoregressive models of US GNP. Econ J 1:47–75. doi:10.1111/1368-423X.11004

Cromwell JB, Labys WC, Terraza M (1994) Univariate tests for time series models. Sage Publications, Thousand Oaks

Dutta D, Welsh W, Vaze J, Kim SH, Nicholls D (2012) A comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in ewater source. Water Resour Manag 26: 4397–4415. doi:10.1007/s11269-012-0151-9

Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208

Hawley RJ, Bledsoe BP (2011) How do flow peaks and durations change in suburbanizing semi-arid watersheds? a southern California case study. J Hydrol 405:69–82

Jayawardena AW, Gurung AB (2000) Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach. J Hydrol 228:242–264. doi:10.1016/S0022-1694(00)00142-6

Karlsson M, Yakowitz S (1987) Rainfall-runoff forecasting of methods, old and new. Stoch Hydrol Hydraul 1: 303–318

Khatibi R, Ghorbani MA, Naghipour L, Jothiprakash V, Fathima TA, Fazelifard MH (2014) Inter-comparison of time series models of lake levels predicted by several modeling strategies. J Hydrol 511:530–545. doi:10.1016/j.jhydrol.2014.01.009

Komorník J, Komornikova M, Mesiar R, Szökeova D, Szolgay J (2006) Comparison of forecasting performance of nonlinear models of hydrological time series. Phys Chem Earth 31:1127–1145

Krause P, Boyle DP, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97

Kunkel ML, Pierce JL (2010) Reconstructing snowmelt in Idaho's watershed using historic streamflow records. Clim Chang 98:155–176

Latt Z, Wittenberg H (2014) Improving flood forecasting in a developing country: a comparative study of stepwise multiple linear regression and artificial neural network. Water Resour Manag 28:2109–2128. doi:10.1007/s11269-014-0600-8

Legates DR, McCabe GJ (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35:233–241. doi:10.1029/1998wr900018

Lisi F, Villi V (2001) Chaotic forecasting of discharge time series: a case study. J Am Water Resour Assoc 37:271–279

Liu Q, Islam S, Rodriguez-Iturbe I, Le Y (1998) Phase-space analysis of daily streamflow: characterization and prediction. Adv Water Resour 21:463–475

Makungo R, Odiyo JO, Ndiritu JG, Mwaka B (2010) Rainfall–runoff modelling approach for ungauged catchments: a case study of nzhelele river sub-quaternary catchment. Phys Chem Earth A/B/C 35:596–607. doi:10.1016/j.pce.2010.08.001

Modrick TM, Georgakakos KP (2014) Regional bankfull geometry relationships for southern California mountain streams and hydrologic applications. Geomorphology 221:242–260

Moeanaddin R, Tong H (1988) A comparison of likelihood ratio test and cusum test for threshold autoregression. J R Stat Soc Ser D 37:213–225

Ng WW, Panu US, Lennox WC (2007) Chaos based analytical techniques for daily extreme hydrological observations. J Hydrol 342:17–41

Patel S, Ramachandran P (2015) A comparison of machine learning techniques for modeling river flow time series: the case of upper Cauvery river basin. Water Resour Manag 29:589–602. doi:10.1007/s11269-014-0705-0

Phoon K, Islam M, Liaw C, Liong S (2002) Practical inverse approach for forecasting nonlinear hydrological time series. J Hydrol Eng 7:116–128

Pinson P, Christensen LEA, Madsen H, Sorensen PE, Donovan MH, Jensen LE (2008) Regime-switching modelling of the fluctuations of offshore wind generation. J Wind Eng Ind Aerodyn 96:2327–2347

Porporato A, Ridolfi L (1997) Nonlinear analysis of river flow time sequences. Water Resour Res 33:1353–1367

Safeeq M, Grant GE, Lewis SL, Tague C (2013) Coupling snowpack and groundwater dynamics to interpret historical streamflow trends in the western United States. Hydrol Process 27:655–668

Sharifazari S, Araghinejad S (2015) Development of a nonparametric model for multivariate hydrological monthly series simulation considering climate change impacts. Water Resour Manag 29:5309–5322. doi:10.1007/s11269-015-1119-3

Sivakumar B (2003) Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach. Environ Model Softw 18:721–728

Tong H (1978) On a threshold model. In: Chen CH (ed) Pattern recognition and signal processing. Sijhoff and Noordoff, Amsterdam, pp 101–141

Tong H (1983) Threshold models in non-linear time series analysis. Springer, New York

Tongal H, Demirel MC, Booij MJ (2013) Seasonality of low flows and dominant processes in the Rhine river. Stoch Env Res Risk A 27:489–503. doi:10.1007/s00477-012-0594-9

Toth E, Brath A, Montanari A (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. J Hydrol 239:132–147. doi:10.1016/S0022-1694(00)00344-9

Wang W, Vrijling JK, Van Gelder PHAJM, Ma J (2006) Testing for nonlinearity of streamflow processes at different timescales. J Hydrol 322:247–268

Yu XY, Liong SY, Babovic V (2004) EC-SVM approach for real-time hydrologic forecasting. J Hydroinf 6:209–233