

An Integration of Stationary Wavelet Transform and Nonlinear Autoregressive Neural Network with Exogenous Input for Baseline and Future Forecasting of Reservoir Inflow

Siriporn Supratid¹ • Thannob Aribarg^{1,2} • Seree Supharatid³

Received: 20 March 2017 / Accepted: 17 May 2017 / Published online: 1 June 2017 © Springer Science+Business Media Dordrecht 2017

Abstract For effective water resources management and planning, an accurate reservoir inflow forecast is essential not only in training and testing phases but also in particular future periods. The objective of this study is to develop a reservoir inflow integrated forecasting model, relying on nonlinear autoregressive neural network with exogenous input (NARX) and stationary wavelet transform (SWT), namely SWT-NARX. Due to the elimination of down-sampling operation, SWT provides influential reinforcement of efficiently extracting the hidden significant, temporal features contained in the nonstationary inflow time series without information loss. The decomposed SWT sub-time series are determined as input-output for NARX forecaster; where a multi-model ensemble global mean (MMEGM) of downscaled precipitation based on nine global climate models (GCMs) represents as a climate-change exogenous input. Two major reservoirs in Thailand, Bhumibol and Sirikit ones are focused. Pearson's correlation coefficient (r) and root mean square error (RMSE) are employed for performance evaluation. The achieved results indicate that the SWT-NARX explicitly outperforms the comparable forecasting approaches regarding a historical baseline period (1980-1999). Therefore, such SWT-NARX is further employed for future projection of the reservoir inflow over near (2010– 2039) -, mid (2040–2069) - and far (2070–2099) - future periods against the inflow of the baseline one.

Siriporn Supratid siri_sup1@hotmail.com

¹ College of Information Technology and Communication, Rangsit University, Pathumthani 12000, Thailand

² Climate Change and Disaster Center, Rangsit University, Pathumthani 12000, Thailand

³ Provincial Waterworks Authority, Bangkok 10210, Thailand

Keywords Inflow forecast · Nonlinear autoregressive neural network with exogenous input · Stationary wavelet transform · Bhumibol reservoir · Sirikit reservoir

1 Introduction

Short- and long- term accurate and reliable forecasts of reservoir inflow are essential for effective water resources management and planning. Such accurate forecasts would aid maximizing flood protection; while saving water for drought times of the year over long-run periods. Backpropagation neural network (BPNN) (Rumelhart et al. 1986), the most commonly used artificial neural network (ANN) can estimate the nonlinear input-output relationship without the limits of traditional linear time series models. BPNN, thereby has been widely applied as an effective approach for modeling highly nonlinear phenomenon in hydrological forecasting (Tiwari et al. 2013; Latt 2015; Mohanty et al. 2015). Jothiprakash and Magar (2012) employed BPNN, adaptive neuro fuzzy inference system and linear genetic programming for reservoir inflow prediction.

However, BPNN is executed on a basis of static-based learning. It was indicated in El-Shafie et al. (2012) the advantage of monthly rainfall time series forecasting accuracy based on dynamic over static neural network model during both training and testing stages at Klang river basin, Malaysia. The static network is inferior to the dynamic ones because the inputs of the static network depend solely on observed data, whereas those of the dynamic networks incorporate observed data with time delay units through recurrent connections. Nonlinear autoregressive neural network with exogenous input (NARX) (Chen et al. 1990; Narendra and Parthasarathy 1990), a proficiently dynamic neural forecasting tool exploits recurrent neural architecture. As opposed to other recurrent neural networks (RNN) (Elman 1990), NARX has limited feedback architectures that come only from the output neuron instead of from hidden neurons. It has been reported that such learning architecture can yield more effective results in NARX model than in other recurrent architectures with hidden states (Horne and Giles 1995). According to Chang et al. (2014), NARX indicated the best performance on multi-step-ahead water level forecasting for Taiwanese urban flood control, compared to the static BPNN and Elman's recurrent neural network, including the case where the water level is not available as input for the network. Valipour et al. (2013) pointed the competitive performance of dynamic nonlinear autoregressive neural network (NAR) over static BPNN, autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) (Box et al. 1994) for forecasting the inflow of Dez dam reservoir, Iran. Furthermore, Valipour (2016) demonstrated proficient results of NARX over NAR and nonlinear input-output (NIO) with regard to annual precipitation forecasts in Gilan, Iran. This leads to realize a particular characteristic of the NARX model, saying that the evolution of a phenomenon can be explained by its previous behavior as well as the effect of its related exogenous factors. Current estimation and future projection of precipitation patterns with respect to climate changes is a significant exogenous information, having a main impact on reservoir inflow forecasts (Wehner 2013).

Global climate model (GCM) simulations archived by the coupled model intercomparison project (CMIP) have been one of the most important sources for future quantitative climate projection for the twenty-first century. The simulations from phase 3 of CMIP (CMIP3) (Meehl et al. 2007) have been extensively analyzed and incorporated into the fourth assessment report (AR4) of the intergovernmental panel on climate change (IPCC) (Solomon et al. 2007). The recently completed simulations for phase 5 of CMIP (CMIP5) (Taylor et al. 2012)

are expected to be extensively featured in the upcoming IPCC fifth assessment report. Compared to CMIP3, CMIP5 models typically have finer resolution processes, incorporation of additional physics, and better-developed or well-integrated earth system components. Such CMIP3 and CMIP5 data have been used to construct multi-model hydroclimate projection. A drying trend in the twenty-first century was found over southwestern North America (Seager et al. 2007; Karl et al. 2009) and the Mediterranean region (Mariotti et al. 2008) based on CMIP3 simulations. Using the CMIP5 data, Hsu et al. (2012, 2013) found a slight enhancement of the hydrological cycle of global monsoons; while Cook and Seager (2013) showed a shift in the seasonality of North American monsoon to late summer under global warming. Brands et al. (2013), Joetzjer et al. (2013), Supharatid et al. (2016), Supharatid (2016) and Kumar et al. (2014) have performed the intercomparison between the performance of CMIP3 and CMIP5, under IPCC special report on emission scenarios (SRES) (Nakicenovic et al. 2000), e.g., A2 and B1 for CMIP3 and under representative concentration pathways (RCPs) (Moss et al. 2010), e.g., RCP8.5 and RCP4.5 for CMIP5. Such RCP8.5 and RCP4.5 are respectively comparable to SRES scenarios A2 and B1 (Rogelj et al. 2012). CMIP3 and CMIP5 under A2 and RCP8.5 scenarios consecutively depict relatively fast rates of atmospheric greenhouse gas accumulation, resulting high concentration; and vice versa for CMIP3 and CMIP5 under B1 and RCP4.5 successively.

However, there have been suggestions and counter-arguments about the value of using multi-model ensemble global mean (MMEGM) versus a single-best model (Santer et al. 2009; Sanderson and Knutti 2012; Knutti et al. 2010).). The MMEGM tends to be an improvement over any individual model, because the bias in one model is cancelled out by another. Various studies have shown that such an MMEGM yielded better prediction against the observations than any single model when compared over multiple variables (Knutti et al. 2010). Supharatid et al. (2016) and Supharatid (2016) implemented precipitation changes projection respectively for Bangkok and Chao Phraya river basin in Thailand by applying downscaled precipitation based on nine MMEGM pairs from CMIP3 and CMIP5 through distribution mapping (DM) statistical downscaling (SD) (Teutschbein and Seibert 2012), under 4 emission scenarios: B1, A2, RCP4.5 and RCP8.5.

Even if an efficient dynamic nonlinear neural forecasting model is employed with the use of exogenous information of CMIP3- and CMIP5- based MMEGM precipitation as a representative of climate change impact, limitation in dealing with nonstationary time series should be bore in mind. A complex nonstationary time series such as reservoir inflow has a variable variance and mean that does not remain constant to their long-run mean over time. Such nonstationarity possibly severely degrades the inflow forecasting performance. Discrete wavelet transform (DWT) (Mallat 1989) was developed as a powerful tool for effectively coping with such nonstationarity problem. To efficiently extracting hidden significant, temporal features for an individual level of frequency resolution in the inflow time series environment, DWT decomposes the original inflow time series into different sub-time series. The lowfrequency sub-time series or approximate components generally reflect periodicity and trends of the original data whereas the high-frequency sub-time series or detail components uncover sharp fluctuations (Kucuk and Oglu 2006). In the literature, there have been numbers of applications using DWT with artificial neural network (ANN) in water resource variables forecasting, including inflow forecasting (Okkan 2012; Krishna 2014; Sehgal et al. 2014a, 2014b). Kumar et al. (2015) performed reservoir inflow forecasting using ensemble models based on ANN, DWT and bootstrap method. Shafaei and Kisi (2016) applied integrated DWT and ARMA, adaptive neuro fuzzy inference system (ANFIS) (Jang et al. 1997) and support vector regression (SVR) (Vapnik 1995) to forecast monthly lake level fluctuations. Shenify et al. (2016) applied artificial neural network (ANN), genetic programming (GP) (Koza 1992) and SVR with DWT for estimating monthly precipitation. Besides, DWT was employed with NARX for Thailand tourist arrival forecast (Kummong and Supratid 2016).

Nevertheless, a critical disadvantage of DWT is that it is not a shift-invariant transform due to down-sampling in each subband. This causes information loss in the respective sub-time series. The stationary wavelet transform (SWT) (Fowler 2005) was designed to overcome the lack of such shift-invariance, regarding DWT. Since down-sampling procedure is removed, SWT sub-time series coefficients contain many redundant information where fine details can be preserved for thorough analysis. SWT was applied with relevance vector regression (Tipping 2000) according to Bai et al. (2014). This work reported that the proposed approach was capable of following the chaos pattern of daily urban water demand reasonably well. Merely few studies have been conducted on using SWT for improving water variables forecasts.

This paper focuses on inflow forecasts into the two major reservoirs in Thailand, Bhumibol and Sirikit ones using an integrated model of the SWT and NARX, namely SWT-NARX. With respect to the proposed method, SWT is executed to achieve efficient feature extraction for an individual level of frequency resolution in the inflow time series environment. Then, the extracted feature set or sub-time series at a particular frequency resolution level along with related exogenous input factors are fed into NARX for further forecasting. The related exogenous input factors here refer to downscaled monthly precipitation around the reservoirs'near-by rainfall gauge stations, based on nine multi-model ensemble global mean (MMEGM) pairs from CMIP3 under B1, A2 scenarios and CMIP5 under RCP4.5, RCP8.5. Such exogenous input factor, MMEGM represents an impact of climate changes on the reservoir inflow. At last, the NARX forecasting outputs from all the resolution levels are reconstructed. Forecasting performance measures rely on Pearson's correlation coefficient (r) and root mean square error (RMSE). Here, not only the historical baseline period (1980–1999) is examined, but also near (2010–2039) -, mid (2040–2069) – and far (2070–2099) -future projection are investigated.

2 Materials and Methods

2.1 Study Areas

Two study areas, Bhumibol and Sirikit reservoirs are respectively described in section 2.1.1 and 2.1.2.

2.1.1 Bhumibol Reservoir

Bhumibol reservoir, originally called Yanhee reservoir is located at Tak province. The reservoir was constructed across Ping River in 1951 with the purposes of water storage and irrigation for agricultures, flood control, fisheries, saltwater intrusion management as well as hydroelectric power generation. The Bhumibol hydropower plant, situated at the reservoir has a total installed capacity of 743.8 MW from 8 turbines. Fig. 1 demonstrates that the reservoir is located at 17.24 North latitude and 98.97 East longitude. The reservoir having a capacity of 13.5 billion m³. The monthly mean temperature over Bhumibol reservoir ranges from 22.55 °C



Fig. 1 Location map of Bhumibol reservoir and the near-by rainfall gauge stations

to 31.55 °C with average annual rainfall of 83.15 mm, most of which occurs in months from September to October. There are 3 near-by rainfall gauge stations, including Bhumibol Dam, Maejo and Chiang Mai affecting the inflow of Bhumibol reservoir.

2.1.2 Sirikit Reservoir

Sirikit reservoir is located across Nan river, about 60 km upstream from the city of Uttaradit province. The reservoir, constructed in 1968 is Thailand's largest earthfilled with crest elevation of 169 m above mean sea level. Like Bhumibol reservoir, Sirikit reservoir has been built for multi-purpose. The maximum water storage capacity of the impounded reservoir is 9.51 billion m³. The reservoir power station contains 4 turbines for an installed capacity of 500 MW. The monthly mean temperature over the Sirikit reservoir

ranges from 23.37 °C to 30.87 °C with average annual rainfall of 114.98 mm, most of which occurs in months from August to September. The location of the Sirikit reservoir is 17.76 North latitude and 100.56 East longitude as shown in Fig. 2 As seen in the figure, a rainfall over Nan rainfall station flows from upstream to Sirikit reservoir; also a rainfall over Uttaradit rainfall station may flow into the reservoir.

2.2 Datasets Used in the Study

In this study, monthly reservoir inflows were collected from electricity generating authority of Thailand (EGAT) during 1980–1999, referred as historical baseline time series data. From such available baseline dataset, the inflow from January, 1984 to December, 1999 and from January, 1980 to December, 1983 are considered for training and testing data.



Fig. 2 Location map of Sirikit reservoir and the near-by rainfall gauge stations

Besides, the future-period (2010–2099) reservoir inflow, divided into near (2010–2039) -, mid (2040–2069) - and far (2070–2099) -future ones are also to be investigated in this study. The exogenous input factors, used for enhancing the inflow forecasting performance are represented by multi-model ensemble global mean (MMEGM) monthly precipitation downscaled around rainfall gauge stations, near by the reservoirs. Such MMEGM have been extracted and averaged with equal weights from nine climate model pairs based on IPCC CMIP3 and CMIP5 GCMs. Table 1 demonstrates a list of the nine climate model pairs. These models have been selected based on data availability and horizontal grid resolution (Supharatid et al. 2016; Supharatid 2016).

Such CMIP3-CMIP5 MMEGM precipitation of the near-by gauge stations are employed for baseline period forecast as well as for near -, mid - and far-future period forecasts. All monthly MMEGM precipitation in the near (2010–2039) -, mid (2040–2069) - and far (2070– 2099) -future periods have been taken from comparable greenhouse warming scenarios, SRES B1 and A2 from CMIP3, and RCP4.5 and RCP8.5 from CMIP5 models (Rogelj et al. 2012; Stocker et al. 2013). All such nine CMIP3 and CMIP5 models were regridded to a 0.5° with 720 longitude × 278 latitude as shown in Fig. 3; bi-linear interpolation is also applied to the locations of 3 (Bhumibol dam, Chiangmai and Maejo) and 2 (Nan and Uttaradit) near-by rainfall gauge stations regarding Bhumibol and Sirikit reservoirs respectively. The nine CMIP3 and CMIP5 GCMs models are downscaled through distribution mapping (DM) statistical downscaling.

2.3 Nonlinear Autoregressive Neural Network with Exogenous Factors (NARX)

NARX forecasting neural network has been applied for modeling discrete-time nonlinear dynamical systems (Leontaritis and Billings 1985; Norgaard et al. 2000). In this network, the next value of the dependent output signal at time t + 1, namely as l(t+1) is regressed on previous values of the output signal and exogenous input signal at time t, namely as l(t). Thus, the NARX model can be mathematically represented as (1).

$$l(t+1) = f \left| l(t), l(t-1), \dots, l(t-d_r); p_i(t), p_i(t-1), \dots, p_i(t-d_{p_i}) \right|$$
(1)

According to (1), the input of the NARX network is formed by two types of regressors, one is the exogenous input regressor $p_i(t) \in \mathbb{R}$, here represented by MMEGM precipitation

CMIP3	Resolution	CMIP5	Resolution	Center
CNRM-CM3	128 × 64	CNRM-CM5	256 × 128	Centre National de Recherches Meteorologiques, France
CSIRO-Mk3.0	192 × 96	CSIRO-Mk3.6	192 × 96	CSIRO, Australia
GFDL-CM2.0	144 × 90	GFDL-CM3	144 × 90	Geophysical Fluid Dynamics Laboratory, NOAA
GFDL-CM2.1	144 × 90	GFDL-ESM2M	144 × 90	Geophysical Fluid Dynamics Laboratory, NOAA
GISS-ER	72 × 46	GISS-E2-H	144×90	Goddard Institute for Space Studies, USA
IN-CM3.0	72 × 45	INM-CM4	180×120	Institute of Numerical Mathematics, Russia
IPSL-CM4	96 × 72	IPSL-CM5A-LR	96 × 96	Institut Pierre Simon Laplace, France
MIROC3.2	128×64	MIROC5	256 × 128	CCSR/NIRS/FRCGC, Japan
MRI-CGCM2.3.2	192 × 96	MRI-CGCM3	320 × 160	Meteorological Research Institute, Japan

Table 1 List of IPCC CMIP3 and CMIP5 GCMs used in this study



Fig. 3 Climate model grid boxes for Thailand

at a particular station *i*; the precipitation at more than one station may be involved. The other one is the output regressor $l(t) \in \mathbb{R}$, here referred as reservoir inflow at time *t*. $d_{p_i} \in \mathbb{Z}$ and $d_i \in \mathbb{Z}$ consecutively are the *lags* of the exogenous input *i* and output regressors of the system, where $d_i \ge d_{p_i} \ge 1$. The nonlinear mapping $f(\cdot)$ is generally an unknown smooth function; and can be approximated by a standard multilayer perceptron network. The output of the NARX network, represented by the reservoir inflow is basically estimated under one of the following two modes. The first one is a series-parallel (SP) mode, where the output's regressor is formed by actual values of the system's output during both training and testing phases, as shown in (2):

$$\hat{l}(t+1) = \hat{f}\Big[l(t), l(t-1), \dots, l(t-d_r); \hat{p}_i(t), \hat{p}_i(t-1), \dots, \hat{p}_i(t-d_p)\Big]$$
(2)

where $\hat{l}(t+1)$, $\hat{p}_i(t)$ and $\hat{f}(\cdot)$ are an estimators of l(t+1), $p_i(t)$ and $f(\cdot)$ respectively. NARX using SP-mode, thereby is utilized for the historical baseline forecast due to data availability. The other mode refers to a parallel (P) mode, where actual values of the reservoir inflow which is a system's output are available during only the training of the network. During the testing

one, estimated outputs are fed back and included in the output's regressor mainly due to lack of complete data, as shown in (3). By this reason, NARX using P-mode should be applied for future projection.

$$\hat{l}(t+1) = \hat{f} \left[\hat{l}(t), \hat{l}(t-1), \dots, \hat{l}(t-d_r); \hat{p}_i(t), \hat{p}_i(t-1), \dots, \hat{p}_i(t-d_p) \right]$$
(3)

2.4 Stationary Wavelet Transform (SWT)

To enhance NARX forecasting capability, stationary wavelet transform (SWT) (Fowler 2005) is adopted here. Stationary wavelet transform (SWT) was designed to overcome the lack of translation-invariance of the discrete wavelet transform (DWT) by removing the down-sampling in the DWT and up-sampling the filter coefficients by a factor of 2^{i-1} in the jth level of the decomposition algorithm (Shensa 1992). The up-sampling procedure is carried out before performing convolution at each scale. The redundancy of this type of transform facilitates the identification of salient, fine detail features in the reservoir inflow time series data. Similar to DWT, the SWT decomposes the inflow time series l(t)by applying recursively a succession of the low and high -pass filters based on wavelet filtering basis functions to l(t), which allows separating its high frequency components from the low frequency ones. The decomposition can be illustrated as a form of dyadic tree. Fig. 4 delineates a dyadic tree of 3-level discrete stationary wavelet decomposition.

First l(t) is decomposed into the trend or the approximate, low frequency components at the 1-level scale decomposition, $A_1(t)$ and the deviations from the trend or the detail, high frequency components at the same scale decomposition level, $D_1(t)$. $A_1(t)$ is successively decomposed into low and high frequency components at the 2-level, denoted as $A_2(t)$ and $D_2(t)$ consecutively, and so on. Based on this dyadic tree, the reservoir inflow time series l(t) can be reconstructed by combining $A_3(t)$, $D_3(t)$, $D_2(t)$ and $D_1(t)$, that refers to the inverse wavelet transform.

 $A_3(t)$, $D_3(t)$, $D_2(t)$ and $D_1(t)$, appeared in Fig. 4 represent the hidden features extracted by SWT. $A_3(t)$ is the low-frequency approximate components; whereas $D_3(t)$, $D_2(t)$ and $D_1(t)$ are





the high-frequency detail components, ordered from coarsest to finest. Based on SWT algorithm, $A_3(t)$ as well as $D_3(t)$, $D_2(t)$ and $D_I(t)$ are extracted respectively by (4) and (5):

$$A_{3}(t) = \sum_{\forall n} a_{3,n} \phi_{3,n}(t)$$
(4)

$$D_j(t) = \sum_{\forall n} d_{j,n} \psi_{j,n}(t) \tag{5}$$

where j = 1, 2, 3 and $n \ge 0$ refers to time interval which is an integer type. From (4), one can say that $A_3(t)$, an approximate sub-time series is a linear combination of the wavelet low-pass filtering basis function, $\phi(t)$. In a similar manner, from (5) $D_3(t)$, $D_2(t)$ and $D_1(t)$, called detail sub-time series are linear combination of the wavelet high-pass filtering basis function, $\psi(t)$. SWT finds the coefficients $a_{3,n}$ and $d_{j,n}$ by projecting the reservoir inflow time series l(t) onto the wavelet low and high -pass filtering functions consecutively at different time interval and particular frequency resolution levels. The number of coefficients for each level is about half that of the preceding level in the DWT; whereas the number of coefficients is the same for each level in the SWT. Such retention of redundant makes the SWT translation invariant, which provides powerful support to analyze each given reservoir inflow sub-time series at various frequency fluctuation levels.

2.5 The Proposed Method

The proposed forecasting approach, SWT-NARX serves as a combination of SWT and NARX neural network. To enhance the forecasting capability, the reservoir inflow forecast is performed with the use of the climate-change exogenous factors, which refers to MMEGM monthly precipitation downscaled to the near-by rainfall gauge stations i = 1, ..., N, represented as $\hat{\mathbf{p}}(t) = \{\hat{p}_1(t)...\hat{p}_N(t)\}$, as seen in Fig. 5. Such MMEGM monthly precipitation or $\hat{\mathbf{p}}(t)$ relies on B1 and A2 scenarios of CMIP3-based and RCP4.5 and RCP8.5 of CMIP5-based GCMs. An overall framework of the proposed SWT-NARX for a 3-level wavelet decomposition, for instance is presented in Fig. 5. It can be explained in terms of training and testing phases as follows:

- 1) Training phase:
 - 1.1 The original reservoir inflow time-series with respect to the training baseline period (January, 1984 to December 1999), is decomposed by SWT into an approximate, A₃(t) and details, D₃(t), D₂(t) and D₁(t) sub-time series.
 - 1.2 Each sub-time series of both approximate and details is separately fed along with the corresponding exogenous input, MMEGM monthly precipitation, $\hat{\mathbf{p}}(t)$ to an individual NARX using series-parallel (SP) mode for model training purpose.
 - 1.3. At the end of this learning phase, the wavelet as well as NARX trained parameters related to each sub-time series are obtained for further using in the testing phase.
- Testing phase:

This phase relates to two main sets of time series data. The first one refers to the time series with respect to the baseline-period (January, 1980 to December 1983) validation-testing data;



Fig. 5 An overall framework of the proposed SWT-NARX for a 3-level wavelet decomposition

whereas the second one relies on the future-period (January, 2010 to December, 2099) unseentesting data.

- 2.1) For the baseline-period validation-testing data, the original reservoir inflow timeseries is decomposed into $A_3(t)$, $D_3(t)$, $D_2(t)$ and $D_1(t)$ based on SWT parameters obtained from the training phase; whilst for the future-period unseen-testing data, only the starting seed, the inflow in January and February 2010 is decomposed in the same manner.
- 2.2) Each sub-time series of both approximate and details is separately fed along with the corresponding exogenous input, MMEGM monthly precipitation $\hat{\mathbf{p}}(t)$ to an individual NARX for forecasting purpose by using the NARX trained parameters obtained from the training phase.
- 2.2.1) For the baseline-period validation-testing data, NARX with series-parallel (SP) mode is executed, where the sub-time series of the system's output actual values, represented by the reservoir inflow are employed as output regressor.
- 2.2.2) For the future-periods unseen-testing data, NARX with parallel (P) mode is employed, where the actual values of the future-period starting seed is provided as input for the first step of forecasting process; for further forecasting steps, calculated or estimated outputs are fed back to the output regressor due to lack of inflow information.
- 2.3) All the forecasted components are supplied to inverse wavelet transform process for finally reconstructing the forecasted reservoir inflow.

2.6 Evaluation Criteria

To evaluate the performance of the SWT-NARX, NARX, BPNN and Elman's RNN, Pearson's correlation coefficient (*r*) and root mean square error (RMSE) are used as follows:

$$r = \frac{\sum_{i=1}^{n} \left((l_o)_i - \bar{l}_o \right) \left((l_p)_i - \bar{l}_p \right)}{\sqrt{\sum_{i=1}^{n} \left((l_o)_i - \bar{l}_o \right)^2 \sum_{i=1}^{n} \left((l_p)_i - \bar{l}_p \right)^2}}$$
(6)

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{i=1}^{n} \left(\left(l_p \right)_i - \left(l_o \right)_i \right)^2}{n}} \tag{7}$$

where l is the reservoir inflow; and the subscripts o and p denote the observed and predicted values, respectively.

3 Results and Discussion

First of all the experiment considers historical baseline time series, that is divided into 192 training (80%) and 48 testing data (20%), which respectively refer to January 1984 to December 1999 and January 1980 to December 1983. The comparison tests involve the proposed SWT-NARX as well as some other related forecasters, including NARX alone, BPNN and Elman's RNN. Necessary parameters used in SWT-NARX for historical baseline forecasting relying on CMIP3 – and CMIP5 –based MMEGM monthly precipitation are declared in Table 2. The parameters are selected for SWT-NARX so as to achieve the effective testing results. The same numbers of input lag as well as hidden neurons are utilized for other related forecasting methods. Wavelet parameters, chosen for SWT-NARX depends on a competency to evaluate local and global behavior of the time series. Such wavelet parameters selection relies on popular, effective wavelet low and high –pass filtering functions: Daubechies (db), Coiflet (coif), Symlet (sym) as well as Haar using one to five scale decomposition levels. Numbers of input lag that yield the best performance for NARX is applied to all other forecasting methods.

The SWT-NARX model that yields the best performance in the testing phase, based on the historical baseline data is then employed for the near (2010–2039) -, mid (2040–2059) - and far (2060–2099) -future inflow projection under CMIP3- and CMIP5 –based precipitation

Bhumibol		Sirikit		
CMIP3 3-input lag, 3-hidden neuron, coif4 using 4 decomposition level	CMIP5 2-input lag, 10-hidden neuron, sym4 using 4 decomposition level	CMIP3 3-input lag, 5-hidden neuron, db2 using 4 de- composition level	CMIP5 4-hidden neuron, haar using 1 decomposition level	

Table 2 SWT-NARX parameters used for historical baseline forecasting

scenarios. Moreover, gamma probability density function (pdf) (Wilks 1990) is also employed here to estimate probability of different amount of inflow for the historical baseline and the particular future time intervals. Such gamma pdf is frequently used to represent hydrological time series such as precipitation and reservoir inflow (Vu et al. 2016; Campos et al. 2014). This is because it provides a flexible representation of a variety of distribution shapes with significant statistical variables such as a mode value, while utilizing only two parameters, shape and scale.

The experimental results, relating to Bhumibol and Sirikit reservoir inflows are separately described in the following two sub-sections.

3.1 Bhumibol Reservoir Inflow Forecast

Relying on historical baseline period (January 1980–December 1999) under CMIP3 - and CMIP5 –based precipitation scenarios, experimental results for testing cases with respect to the comparative forecasting methods using series-parallel mode are shown in Table 3. SWT-NARX generates the best performance in terms of Pearson's correlation coefficient (r) as well as root mean square error (RMSE) for all cases as seen in such Table 3.

From Table 3, relying on CMIP3 - and CMIP5 –based precipitation scenarios, 17.19% and 16.79% of r improvements based on SWT-NARX over the traditional NARX are respectively yielded. In addition for CMIP3 and CMIP5, tremendous RMSE improvements of 56.37% and 64.22% based on SWT-NARX over NARX are respectively produced. Such improvements point remarkable benefits of exploiting SWT as feature extraction analysis with NARX forecaster, in comparison to using NARX alone. On the other hand, NARX indicates better forecasting performance over BPNN and Elman's RNN for all the testing cases.

As a result, the SWT-NARX model using the best-performance parameters defined in Table 3 is further employed for future projection. Fig. 6 shows the historical baseline observations against future projection of averaged monthly inflow as well as MMEGM precipitation at the near-by staions under CMIP3-based B1 and A2 scenarios and CMIP5–based RCP4.5 and RCP8.5 scenarios. Such near-by stations include Bhumibol dam (BB), Chiangmai (CM) and Maejo (MJ) stations.

According to Fig. 6, in general, trends of the inflow along with the MMEGM precipitation at each near-by station in the future periods still follow the trend of the observations in the historical baseline one, with regard to all CMIP3 – and CMIP5 –based senarios. Although there exist double peaks MMEGM precipitation in May and September–October at all the near-by stations for both CMIP3 and CMIP5, there exist only one peak reservoir inflow in September. This implies that only the peak precipitation over 6 mm/day, occurred in September–October possibly has an important impact on the reservoir inflow. In September which refers to the peak-inflow month, the inflow under CMIP3-based A2

Forecasting methods	r		RMSE	
	CMIP3	CMIP5	CMIP3	CMIP5
SWT-NARX	0.9662	0.9803	4.3184	3.3274
NARX	0.8245	0.8394	9.8971	9.2998
BPNN	0.7233	0.8217	12.4272	9.6822
RNN	0.3813	0.3751	19.3640	19.7230

Table 3 Performance measures in terms of r and RMSE

Bold entries refers to the best values of r and RMSE



Fig. 6 Historical baseline observations against future projection of averaged monthly inflow as well as MMEGM precipitation at the near-by stations regarding Bhumibol reservoir under (a) CMIP3-based B1 and A2 scenarios and (xb) CMIP5-based RCP4.5 and RCP8.5 scenarios

and CMIP5-based RCP8.5 scenarios is respectively higher than those under CMIP3-based B1 and CMIP5-based RCP4.5 scenarios due to the faster rates of atmospheric greenhouse gas accumulation; in addition, successive increase of inflow amount follows sequential increase of precipitation at all the near-by stations from the near -, mid – and far -future periods for the whole CMIP3 – and CMIP5 –based scenarios. In such peak-inflow month, it indicates extreme 11.70% inflow increase in average during the far-future period regarding RCP8.5 over the historical baseline period. On the other side, the severely lowest amount of inflow 13 mcm in average comes up in February during the far-future under A2 based on CMIP3, represented as 59.38% decrease in average compared to the lowest 32 mcm in March during the historical baseline period.

Furthermore, gamma probability distribution of reservoir inflow with regard to wet (May– October) and dry (November – March) seasons in the historical baseline and future periods for CMIP3 - and CMIP5 –based scenarios are illustrated in Fig. 7. In the wet season, the future periods belong to more shape-dominated regime than the historical baseline one. This implies that fewer extreme events exist in the future periods with higher probability than in the historical baseline one; while more variations of inflow amounts occur in the baseline period. In the dry season, a few variation from the mode value is illustrated with respect to the historical baseline and all scenarios of the future periods. The mode value about 571 mcm and 148 mcm are denoted in the wet and dry season during the historical baseline period. However, A2 scenario based on CMIP3 shows extreme cases of around 709 and 115 mcm mode values of inflow for the wet and dry seasons in the far-future period. One can say, for the extreme cases, the modes of inflow amount approximately increases by 24.17% and decreases by 22.30% with a few probabilities lesser than 0.003 and 0.007 respectively in the wet and dry seasons.



Fig. 7 Gamma probability distribution of Bhumibol reservoir inflow with regard to wet (May–October) and dry (November – March) seasons for CMIP3 - and CMIP5 –based scenarios. **a** wet season for CMIP3, **b** wet season for CMIP5, **c** dry season for CMIP3, **d** dry season for CMIP5

3.2 Sirikit Reservoir Inflow Forecast

Table 4 illustrates experimental results for testing cases with respect to the comparative forecasting methods using series-parallel mode. Such results relies on historical baseline period (January

Forecasting methods	r		RMSE	
	CMIP3	CMIP5	CMIP3	CMIP5
SWT-NARX	0.9304	0.8882	8.8653	9.9322
NARX	0.8179	0.6584	13.8065	16.9268
BPNN	0.7686	0.6568	14.5509	17.7040
RNN	0.4839	0.4574	22.2620	21.1311

Table 4 Performance measures in terms of r and RMSE

Bold entries refers to the best values of r and RMSE

1980–December 1999) under CMIP3 - and CMIP5 –based precipitation scenarios. Once again, SWT-NARX yields the best performance in terms of *r* as well as RMSE for all cases.

From Table 4, similar to Bhumibol reservoir inflow forecast, relying on CMIP3 and CMIP5, 13.75% and 34.90% of r improvements based on SWT-NARX over NARX alone are respectively yielded. In addition for CMIP3 and CMIP5, tremendous RMSE improvements of 39.07% and 43.90% based on SWT-NARX over NARX respectively are produced. Based on such performance improvements, remarkable benefits of exploiting SWT as feature extraction analysis with NARX forecaster are denoted once again, in comparison to using NARX alone. Like Bhumibol reservoir inflow forecasting results, NARX yields better forecasting performance than BPNN and Elman's RNN for all testing cases.

As a consequence, future projection is further performed relying on the SWT-NARX model using the best-performance parameters defined in Table 3. Fig. 8 illustrates the historical baseline observations against future projection results of averaged monthly inflow as well as MMEGM precipitation at the near-by stations under CMIP3-based B1 and A2 scenarios and CMIP5–based RCP4.5 and RCP8.5 scenarios. Such near-by stations include Nan and Uttaradit (UT) stations.

According to Fig. 8, in general, trends of the inflow along with the MMEGM precipitation in the future periods still follow the trend of the observations in the historical baseline one, with regard to all CMIP3 – and CMIP5 –based scenarios. Although there exist double peaks MMEGM precipitation in May as well as August-September at such couple near-by stations for both CMIP3 and CMIP5, there exist only one peak reservoir inflow in August-September. This implies that only the peak precipitation over 8 mm/day, occurred in August-September possibly has an important impact on the reservoir inflow. In September, which is the same peak-inflow month as Bhumibol reservoir case, the inflow amount under CMIP3-based A2 and CMIP5-based RCP8.5 scenarios is respectively greater than those under CMIP3-based B1 and CMIP5-based RCP4.5 scenarios for most circumstances due to the faster rates of atmospheric greenhouse gas accumulation; in addition, successive increase of inflow amount follows sequential increase of precipitation at Nan as well as UT stations from the near -, mid - and far -future periods for all CMIP3 - and CMIP5 -based scenarios. In such peak-inflow month, it indicates the extreme 11.82% inflow increase in average during the far-future period under CMIP5-based RCP8.5 scenario over the baseline period. On the other side, the severely lowest amount of inflow 43 mcm in average appears in March during the farfuture under RCP8.5 based on CMIP5, represented as 42.67% decrease in average compared to the lowest 75 mcm in March during the historical baseline period.

Like Bhumibol reservoir case, gamma probability distribution of Sirikit reservoir inflow with regard to wet and dry seasons in historical and future periods for CMIP3 - and CMIP5 – based scenarios are illustrated in Fig. 9. For the dry season, a sharp shape-dominated characteristic belonging to the historical baseline period is explicit; in contrast, solid scale-



Fig. 8 Historical baseline observations against future projection of averaged monthly inflow as well as MMEGM precipitation at the near-by stations regarding Sirikit reservoir under (a) CMIP3-based B1 and A2 scenarios and (b) CMIP5-based RCP4.5 and RCP8.5 scenarios

dominated regimes are mostly manifested in the far-future period. This implies very few extreme events exist in the historical baseline period; whereas rather solid variations of inflow amount mostly occur in the far-future. In the wet season, fewer extreme events exist in the future periods than in the historical baseline, especially shown in CMIP5-based scenarios. In the wet and dry seasons during historical baseline period, the inflow amounts of about 658 and 115 mcm mode values of inflow consecutively appears. With respect to the far-future period under CMIP3-based B1 scenario, an extreme case in wet season shows approximate 746 mcm mode value of inflow amount; while a severe drought demonstrate around 12 mcm mode value of inflow amount. One can say, for the most serious cases, the modes of inflow approximately increases by 13.37% and decreases by 89.57% respectively in wet and dry seasons. However, the severe latter case may come up with only a few probability, lesser than 0.004.

4 Conclusions

Within the scope of the historical baseline period (1980–1999), for both Bhumibol and Sirikit reservoirs, SWT-NARX yields the best performances in terms of Pearson's correlation coefficient



Fig. 9 Gamma probability distribution of reservoir inflow with regard to annual, wet (May–October) and dry (November – March) seasons for CMIP3 - and CMIP5 –based scenarios. **a** wet season for CMIP3, **b** wet season for CMIP3, **c** dry season for CMIP5, **c** dry season for CMIP5, **d** dry season for CMIP5

(*r*) as well as root mean square error (RMSE) for all testing cases under CMIP3 - and CMIP5 - based models, compared to other related traditional models. Therefore, such SWT-NARX is further employed for the near (2010–2039) -, mid (2040–2069) - and far (2070–2099) -future inflow projections, relying on MMEGM monthly precipitation under CMIP3-based B1, A2 scenarios and CMIP5-based RCP4.5, RCP8.5 ones. The projection results indicate that trends of monthly inflow as well as the MMEGM monthly precipitation at the stations, near by the

mentioned reservoirs during the near -, mid – and far –future periods still follow those in the historical baseline period for all CMIP3 – and CMIP5 –based scenarios. The most extreme cases of peak and lowest inflow amount are usually found respectively in September and February/ March during the far-future periods for both Bhumibol and Sirikit reservoirs. The severe inflow decrease in the dry season may happen with a few probability during such far-future period at Sirikit reservoir. Nevertheless, one may also realize some amount of inflow decrease in average, possibly happened in the near-future at both reservoirs.

The findings with regard to this work will be useful to water resource policy makers in pondering whether the current drainage system is appropriate to meet the inflow changes in short – and long –term periods. However, this study still has some limitations that the results, presented here are average values with respect to short – and long -term projections. Such projection values represent large-time-scale changes; thus, face difficulties in translation to a local-time-scale. In addition, significant changes in land use or urbanization around the reservoirs as well as the near-by rainfall gauge stations affect the hydrological process; and in turn, affect the inflow behaviors.

Acknowledgements We gratefully thank to the National Research Council of Thailand (NRCT) for funding this work. We also thank to the electricity generating authority of Thailand (EGAT) and Thai Meteorological Department for providing the observed reservoir inflow and rainfall data respectively.

References

- Bai Y, Wang P, Li C, Xie J, Wang Y (2014) A multi-scale relevance vector regression approach for daily urban water demand forecasting. J Hydrol 517:236–245
- Box GEP, Jenkins GM, Reinsel GC (1994) Time Series Analysis: Forecasting and Control, 3rd edn. Prentice hall, Englewood cliffs pp 197–199
- Brands S, Herrera S, Ferna'ndez J, Gutie'rrez JM (2013) How well do CMIP5 earth system models simulate present climate conditions in Europe and Africa? Clim Dyn 41(3):803–817
- Campos JNB, Souza Filho FA, Lima HVC (2014) Risks and uncertainties in reservoir yield in highly variable intermittent rivers: case of the Castanhão reservoir in semi-arid Brazil. Hydrol Sci J 59(6):1184–1195
- Chang FJ, Chen PA, Lu YR, Huang E, Chang KY (2014) Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. J Hydrol 517:836–846
- Chen S, Billings SA, Grant PM (1990) Non-linear system identification using neural networks. Int J Control 51(6):1191–1214
- Cook BI, Seager R (2013) The response of the North American monsoon to increased greenhouse gas forcing. J Geophys Res Atmos 118(4):1690–1699
- Elman JL (1990) Finding structure time. Cogn Sci 14(2):179-221
- El-Shafie A, Noureldin A, Taha M, Hussain A, Mukhlisin M (2012) Dynamic versus static neural network model for rainfall forecasting at Klang River basin, Malaysia. Hydrol Earth Syst Sci 16(4):1151–1169
- Fowler JE (2005) The redundant discrete wavelet transform and additive noise. IEEE Signal Process Lett 12(9):629-632
- Horne BG, Giles CL (1995) An experimental comparison of recurrent neural networks. MIT Press, In Advanc in Neural Info Process Syst, pp 697–704
- Hsu P-C et al (2012) Increase of global monsoon area and precipitation under global warming: a robust signal? Geophys Res Lett 39(6):L06701
- Hsu P-C, Li T, Murakami H, Kitoh A (2013) Future change of the global monsoon revealed from 19 CMIP5 models. J Geophys Res Atmos 118(3):1247–1260
- Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice-Hall, Englewood Cliffs pp 665–685
- Joetzjer E, Douville H, Delire C, Ciais P (2013) Present-day and future Amazonian precipitation in global climate models: CMIP5 versus CMIP3. Clim Dyn 41(11):2921–2936
- Jothiprakash V, Magar RB (2012) Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. J Hydrol 450–451:293–307

- Karl TR, Melillo JM, Peterson TC (eds) (2009) Global climate change impacts in the United States. Cambridge University Press, New York
- Knutti R, Abramowitz G, Collins M, Eyring V, Gleckler PJ, Hewitson BM, Mearns L (2010) Good practice guidance paper on Assessing and combining multi model climate projections. Meeting report of the intergovernmental panel on climate change expert meeting on Assessing and combining multi model climate projections. University of Bern, Bern, IPCC Working Group I Technical Support Unit
- Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge
- Krishna B (2014) Comparison of wavelet based ANN and regression models for reservoir inflow forecasting. J Hydrol Eng 19(7):1385–1400
- Kucuk M, Oglu NA (2006) Wavelet regression technique for stream flow prediction. J Appl Stat 33(9):943-960
- Kumar D, Kodra E, Ganguly AR (2014) Regional and seasonal intercomparison of CMIP3 and CMIP climate model ensembles for temperature and precipitation. Clim Dyn 43(9):2491–2518
- Kumar S, Tiwari MK, Chatterjee C, Mishra A (2015) Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method. Water Resour Manag 29(13):4863–4883
- Kummong R, Supratid S (2016) Thailand tourism forecasting based on a hybrid of discrete wavelet decomposition and NARX neural network. Ind Manag Data Syst 116(6):1242–1258
- Latt ZZ (2015) Application of feedforward artificial neural network in Muskingum flood routing: a black-box forecasting approach for a Natural River system. Water Resour Manag 29(14):4995–5014
- Leontaritis IJ, Billings SA (1985) Input-output parametric models for non-linear systems part I: deterministic non-linear systems. Int J Control 41(2):303–328
- Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693
- Mariotti A et al (2008) Mediterranean water cycle changes: transition to drier twenty-first century conditions in observations and CMIP3 simulations. Environ Res Lett 3(4):044001
- Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE (2007) The WCRP CMIP3 multimodel dataset: a new era in climate change research. Bull Am Meteorol Soc 88(9): 1383–1394
- Vu MT, Aribarg T, Supratid S, Raghavan SV, Liong S-Y (2016) Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? Theor Appl Climatol 126(3):453–467
- Mohanty S, Jha MK, Raul SK, Panda RK, Sudheer KP (2015) Using artificial neural network approach for simultaneous forecasting of weekly groundwater levels at multiple sites. Water Resour Manag 29(15):5521– 5532
- Moss RH, Edmonds JA, Hibbard KA et al (2010) The next generation of scenarios for climate change research and assessment. Nature 463(7282):747–756
- Nakicenovic N et al (2000) Special Report on Emissions Scenarios (SRES): A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press. http://www.grida. no/climate/ipcc/emission/index.htm. Accesses 2 Oct 2016
- Narendra KS, Parthasarathy K (1990) Identification and control of dynamical systems using neural networks. IEEE Trans Neural Netw 1(1):4–27
- Norgaard M, Ravn O, Poulsen NK, Hansen LK (2000) Neural networks for Modelling and control of dynamic systems. Springer, Berlin
- Okkan U (2012) Wavelet neural network model for reservoir inflow prediction. Sci Iran 19(6):1445-1455
- Rogelj J, Meinshausen M, Knutti R (2012) Global warming under old and new scenarios using IPCC climate sensitivity range estimates. Nat Clim Chang 2(4):248–253
- Rumelhart DE, JL MC, Corporate PDP Research Group (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol 1. MIT Press, Cambridge
- Sanderson BM, Knutti R (2012) On the interpretation of constrained climate model ensembles. Geophys Res Lett 39(16):1–6 (L16708)
- Santer BD, Taylor KE, Glecker PJ et al (2009) Incorporating model quality information in climate change detection and attribution studies. Proc Natl Acad Sci 106(35):14778–14783
- Seager R et al (2007) Model projections of an imminent transition to a more arid climate in southwestern North America. Science 316(5828):1181–1184
- Sehgal V, Sahay RR, Chatterjee C (2014a) Effect of utilization of discrete wavelet components on flood forecasting performance of wavelet based ANFIS models. Water Resour Manag 28(6):1733–1749
- Sehgal V, Tiwari MK, Chatterjee C (2014b) Wavelet bootstrap multiple linear regression based hybrid modeling for daily river discharge forecasting. Water Resour Manag 28(10):2793–2811
- Shafaei M, Kisi O (2016) Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. Water Resour Manag 30(1):79–97

- Shenify M, Danesh AS, Gocić M, Taher RS, Wahab AWA, Gani A, Shamshirband S, Dalibor P (2016) Precipitation estimation using support vector machine with discrete wavelet transform. Water Resour Manag 30(2):641–652
- Shensa MJ (1992) The discrete wavelet transform: wedding the a Trous and Mallat algorithms. IEEE Trans Signal Process 40(10):2464–2482
- Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) (2007) Climate change 2007: the physical Science basis. Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change
- Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J (2013) Climate change 2013. The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change
- Supharatid S, Aribarg T, Supratid S (2016) Assessing potential flood vulnerability to climate change by CMIP3 and CMIP5 models: case study of the 2011 Thailand great flood. J Water Clim Change 7(1):52–67
- Supharatid S (2016) Skill of precipitation projection in the Chao Phraya river basin by multi-model ensembleCMIP3-CMIP5. Weather Clim Extremes 12:1–14
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experimental design. Bull Am Meteorol Soc 93(4):485–498
- Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climatechange impact studies: review and evaluation of different methods. Hydrol 456-457:12–29
- Tipping ME (2000) The relevance vector machine. Adv Neural Inf Proces Syst 12:652-658
- Tiwari MK, Song KY, Chatterjee C, Gupta MM (2013) Improving reliability of river flow forecasting using neural networks, wavelets and self-organizing maps. J Hydroinf 15(2):486–502
- Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J Hydrol 476(7): 433–441
- Valipour M (2016) Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. Meteorol Appl 23(1):91–100
- Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
- Wehner M (2013) Methods of projecting future changes in Extremes. Extremes in a Changing Clim 65:223–237
- Wilks D (1990) Maximum likelihood estimation for the gamma distribution using data containing zeros. J Clim 3(12):1495–1501