Monthly rainfall prediction using wavelet regression and neural network: an analysis of 1901–2002 data, Assam, India

Article *in* Theoretical and Applied Climatology · October 2014

DOI: 10.1007/s00704-013-1029-3

Manish Kumar Goyal
Indian Institute of Technology Indore

Theor Appl Climatol (2014) 118:25–34
DOI 10.1007/s00704-013-1029-3

ORIGINAL PAPER

Monthly rainfall prediction using wavelet regression and neural network: an analysis of 1901–2002 data, Assam, India

Manish Kumar Goyal

Received: 12 July 2013 / Accepted: 1 October 2013 / Published online: 21 November 2013

Abstract Rainfall is a principal element of the hydrological cycle and its variability is important from both the scientific as well as practical point of view. Wavelet regression (WR) technique is proposed and developed to analyze and predict the rainfall forecast in this study. The WR model is improved combining two methods, discrete wavelet transform and linear regression model. This study uses rainfall data from 21 stations in Assam, India over 102 years from 1901 to 2002. The calibration and validation performance of the models is evaluated with appropriate statistical methods. The root mean square errors (RMSE), N-S index, and correlation coefficient (R) statistics were used for evaluating the accuracy of the WR models. The accuracy of the WR models was then compared with those of the artificial neural networks (ANN) models. The results of monthly rainfall series modeling indicate that the performances of wavelet regression models are found to be more accurate than the ANN models.

1 Introduction

Rainfall being the most important variables is an end product of number of complex processes in atmosphere, which vary both in space and time (Baratta et al. 2003). Associated with global warming, changing rainfall pattern, and its impact on surface water resources, is an important climatic problem facing society today. For water resources planning purposes, a long-term rainfall series is required in hydrological and simulation models (Tantanee et al. 2005). Artificial neural networks (ANN) are being increasingly used to model hydrological processes such as rainfall–runoff modeling, rainfall prediction, flood forecasting, water quality modeling, ground water modeling, development of water management policy, and reservoir operation studies due to their capability to represent any arbitrary nonlinear function given sufficient complexity of the trained networks (Raman et al. 1996; Jain et al. 1999).

Since past decade, wavelet transform technique has become a useful approach for investigating variations, periodicities, and trends in data time series. Labat et al. (2000) used wavelet methods to rainfall rates and runoffs measured at different sampling rates. For the decomposition of the unit hydrograph, Chou and Wang (2002) used on-line estimation of unit hydrograph using the discrete wavelet transform (DWT) components. The rainfall spectrum and its impact on North China during the rainy season with summer monsoon decaying in interdecadal time scales was investigated using wavelet transform (Xingang et al. 2003). Coulibaly and Burn (2004) used wavelet analysis to identify and describe variability in annual Canadian streamflows. Labat et al. (2005) demonstrated that the application of combined continuous and multiresolution analysis, wavelet entropy, wavelet coherence, wavelet cross-correlation leads to several improvements in the analysis of global hydrological signal fluctuations. A possible trend in annual precipitation data series has been investigated using wavelet analysis in Turkey (Partal and Kucuk 2006). Zhou et al. (2008) developed a wavelet predictor–corrector model for the simulation and prediction of monthly discharge time series. Kisi (2009) used wavelet regression (WR) model as an alternative to neural networks for monthly streamflow forecasting in Turkey and WR is found to be better than the ANN models. Kisi (2011) analyzed the performance of the WR technique in daily river stage forecasting and determined that the WR model was improved combining two methods, discrete wavelet transform and a linear regression model. Chou (2011) proposes a new framework for determining the number of resolution levels required in the application of a

M. K. Goyal (✉)
Department of Civil Engineering, Indian Institute of Technology, Guwahati, India
e-mail: vipmkgoyal@gmail.com

M. K. Goyal
e-mail: mkgoyal@iitg.ernet.in

wavelet transform to rainfall time series analysis. An application of wavelet analysis is done with long time series of the total monthly rainfall amount at the capital cities of northeastern Brazil (Santos and Freire 2012). Chou (2013) applied Wavelet transform for simultaneous decomposition of rainfall and runoff time series data in two case studies of the Feng-Hua Bridge and Wu-Tu watershed and found proposed wavelet-based method slightly outperforms the conventional method of using data only at the original scale. Sang (2013), developed a method for discrete wavelet decomposition and an improved wavelet modeling framework, WMF for short, was proposed for hydrologic time series forecasting. These studies indicate that wavelet transform is an effective tool for precisely locating irregularly distributed multi-scale features of climate elements in space and time.

One of the most vulnerable regions in India is its northeastern part comprising of the Assam state. Assam is blessed with fertile soil and a climate conducive to agriculture. Agriculture and related sectors such as food security, and energy security of India are crucially dependent on the timely availability of adequate amount of water (Jain and Kumar 2012). The economy of the state continues to be predominantly agrarian. About 75 % of the state's population is directly or indirectly dependent on agriculture. Prediction of rainfall in Assam did not receive enough concern globally and locally and no comprehensive studies have been carried out in the state as per author's best knowledge. The purpose of this study was to investigate the performance of wavelet regression models for monthly rainfall prediction and to compare them with the performance of artificial neural network models. This study will be useful for water resources planner, managers and agricultural scientists to work out better water management options for local level in various part of state.

2 Materials and methods

2.1 Study area

Assam is the largest state (accounts for ~2.4 % of India's total geographical area) in northeast India (Fig. 1). This state is situated at the foothills of the eastern Himalayas and lies in the middle reach of the river Brahmaputra and Barak. The humidity that is brought into Assam by the southwest monsoons shower an average annual rainfall of 305 cm or more on the Brahmaputra valley and the surrounding region. The Brahmaputra river flows through Assam from east to west over a length of approximately 650 km. This state is home to about 51 forest and sub-forest types, and the confluence of diverse patterns of vegetation due to its topography and the warm and humid climate Mahanta et al. (2003).

Being located in eastern Himalayan range and also seismic active region, Assam is highly vulnerable to floods, river bank erosion, and landslides; and storms are also aggravated. Agriculture sector accounts for about a third of the state domestic product. The total production of food grains and other crops have recorded an increase in the last fifty years, but the per capita food grain has showed a decline.

2.2 Data

The data of monthly precipitation of 21 stations were obtained from Indian Meteorological Department (IMD) site India water portal (http://www.indiawaterportal.org/metdata) covering a period of 102 years (1901–2002). As all series were complete, no gap filling method was needed. Four stations, namely, Tinsukia, Kokrajhar, Nagoan, and Hailakandi, were selected to show the amount and seasonality of precipitation which is located in east, west, central, and south of Assam and the bar plot of monthly precipitation for above station is shown is Fig. 2. To forecast the rainfall at 21 rain gauge station in Assam (Fig. 1), monthly rainfall of 102 years from January 1901 to December 2002 was used.

Table 1 shows the statistics of the total data for monthly rainfall. The recorded monthly maximum rainfall is 1, 594.27 mm at Kamrup station, while the mean monthly rainfall was about 292 mm. The standard deviation varied between 155.35 and 301.43 mm for various stations. The skewness, which is a measure of asymmetry in a probability distribution around the mean, varied between 0.67 and 0.92, positive skewness indicating that annual precipitation during the period is asymmetric and it lies to the right of the mean over all the stations. The observed monthly rainfall shows low positive skewness and indicates that the data has less scattered distribution. The first 60 years (60 % data) from January 1901 to December 1960 was used for calibration of the model, and the remaining 42 years (about 40 % data) from January 1961 to December 2002 data were used for validation.

2.3 Methodology: wavelet analysis

The wavelet analysis is a recent advance in signal processing that has attracted much attention in communications, astronomy, acoustics, data compression, nuclear engineering, signal processing, neurophysiology, image processing, and optical engineering applications since its theoretical development in 1984 by Grossman and Morlet (1984). It is a method based on expressing signals as sums of little waves (Quiroz et al. 2011). It allows analyzing different scales of temporal variability and it does not need a stationary series compared to classical spectral analysis. Fourier transforms concentrate mainly on their frequency and usually assume infinite length signals. In the wavelets tool, wavelets can provide the exact locality of

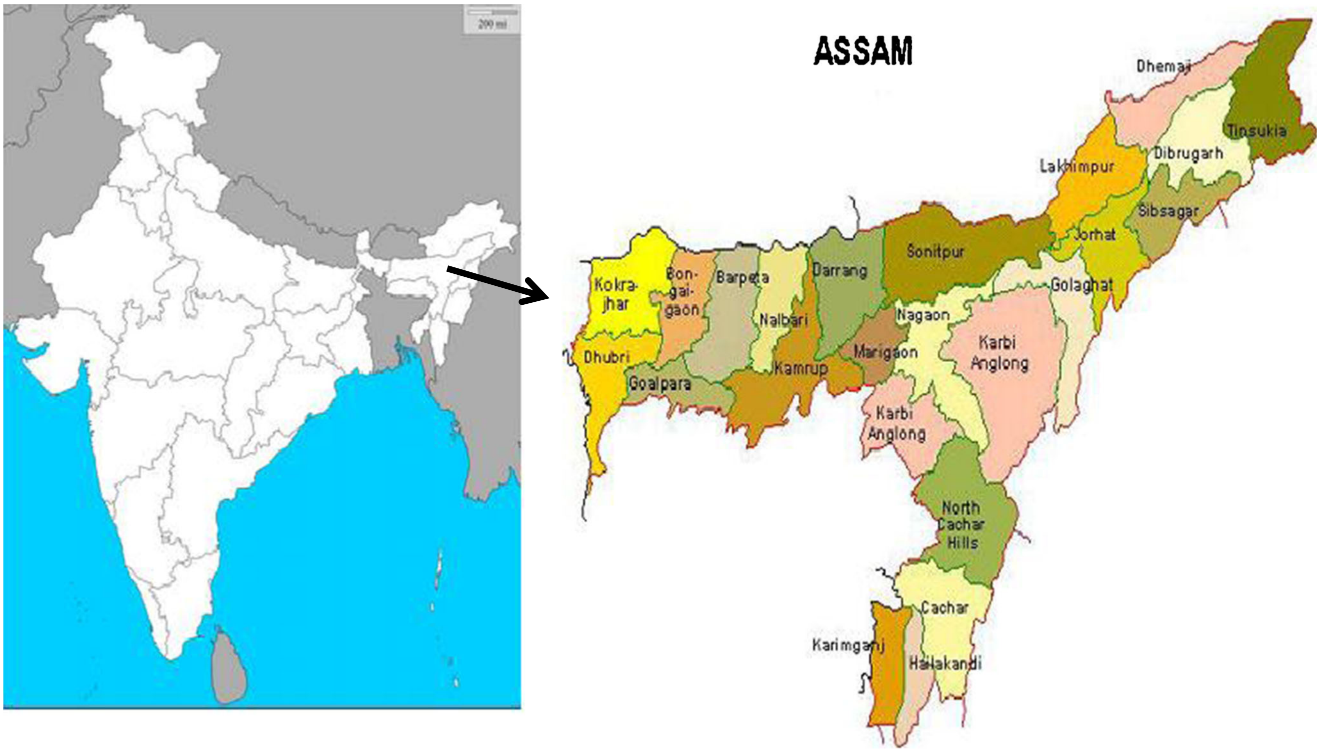

Fig. 1 Study area

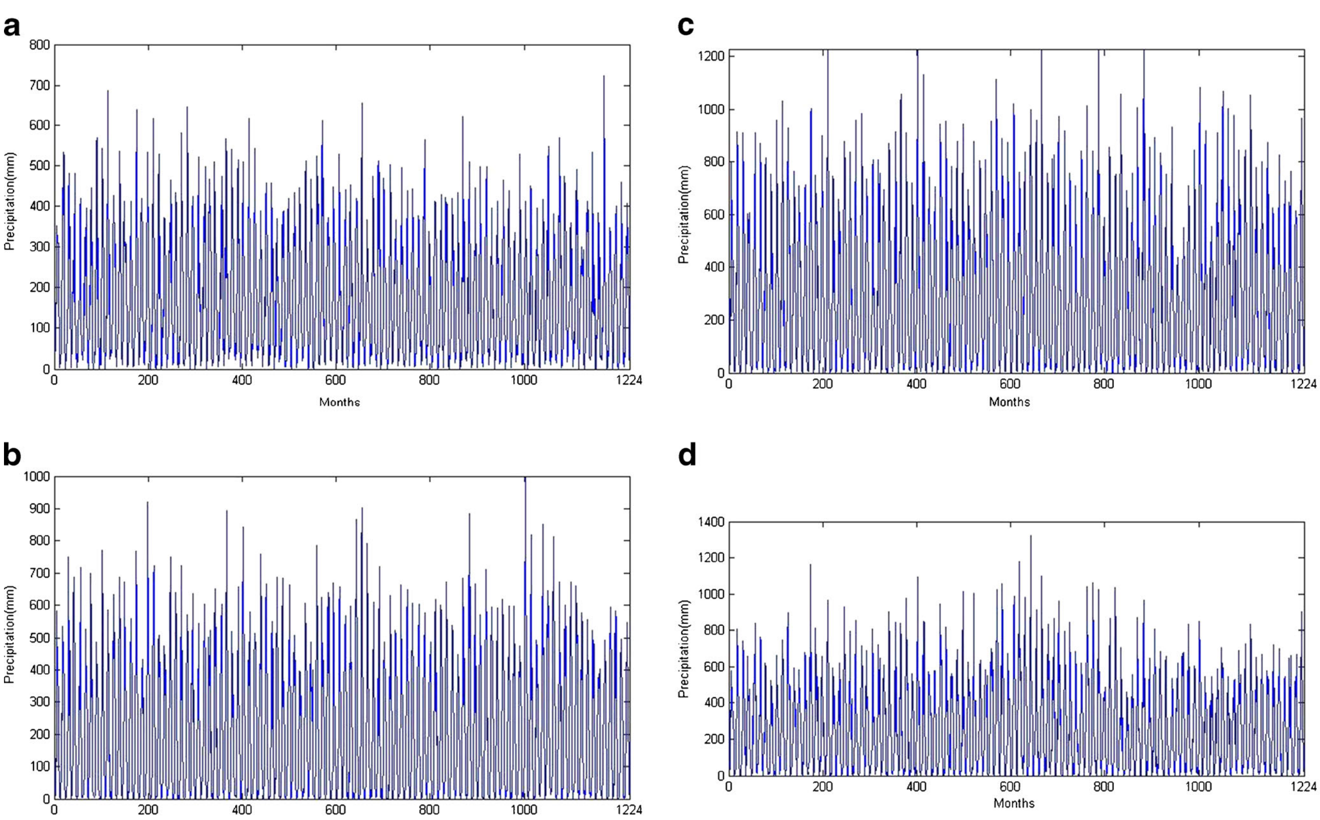

Fig. 2 Plot of monthly precipitation during 1901–2002 at **a** Tinsukia, **b** Kokrajhar, **c** Nagoan, and **d** Hailakandi station

Table 1 Statistical properties of rainfall for 21 stations

Sl No.	Station	Minimum	Maximum	Mean	Standard deviation	Skewness
1	Jorhat	0.19	677.06	183.69	164.32	0.73
2	Barpeta	0.00	1,033.69	230.73	239.38	0.83
3	Cachar	0.00	1,293.43	275.71	265.96	0.87
4	Darrang	0.00	1,069.20	234.79	237.25	0.83
5	Dhemaji	0.20	685.51	180.44	160.98	0.69
6	Dhubri	0.00	1,020.58	185.85	198.97	0.88
7	Dibrugarh	0.20	703.83	183.69	164.32	0.73
8	Goalapara	0.00	1,108.47	222.13	231.72	0.83
9	Golaghat	0.06	705.11	186.69	170.58	0.69
10	Hailakandi	0.00	1,323.74	261.82	256.93	0.92
11	Kamrup	0.00	1,594.27	292.04	301.43	0.89
12	karbi Anglong	0.03	1,234.26	270.23	265.14	0.87
13	Karimganj	0.00	1,197.53	233.98	230.07	0.91
14	Kokrajhar	0.00	997.85	208.74	221.32	0.83
15	Lakhimpur	0.13	674.85	178.10	162.08	0.67
16	Nagaon	0.00	1,525.96	293.95	294.85	0.89
17	Nalbhari	0.00	1,055.19	231.85	239.49	0.85
18	North Chahar Hills	0.00	1,233.48	294.86	280.44	0.80
19	Sibsagar	0.19	694.85	179.75	161.60	0.71
20	Sonitpur	0.00	784.22	192.80	185.47	0.71
21	Tinsukia	0.19	722.23	172.87	155.35	0.78

any changes in the dynamical patterns of the sequence and can be applied to any kind and any size of time series, even when these sequences are not homogeneously sampled in time (Antonios and Constantine 2003). Therefore, this tool is emerged as an alternative to the Fourier transform in preserving local, non-periodic and multiscaled phenomena. Thus, it is appropriated to analyze irregular distributed events and time series that contain nonstationary power at many different frequencies. In general, wavelet transforms can be used to explore, denoise, and smoothen time series which aid in forecasting and other empirical analysis (Lindsay et al. 1996). Wavelet analysis is becoming a common tool for analyzing localized variations of power within a time series. By decomposing a time series into time–frequency space, one is able to determine both the dominant modes of variability and how those modes vary in time (Torrence and Compo 1998; Kisi 2009).

2.4 Discrete wavelet transform (DWT)

Wavelet analysis is sued to determine the frequency content of a signal and then it assesses and determines the temporal variation of this frequency content. The wavelet transform is the tool of choice when signals are characterized by localized high frequency events or when signals are characterized by a large numbers of scale variable processes. Because of its localization properties in both time and scale, the wavelet transform allows for tracking the time evolution of processes at different scales in the signal (Kisi 2009).

Wavelet function $\psi(t)$, can be written as $\int_{-\infty}^{+\infty}\Psi(t)dt=0$.$\Psi_{a,b}(t)$ can be obtained through compressing and expanding

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}}\Psi\left(\frac{t-b}{a}\right) \quad b\in R, a\in R, a\in \neq 0 \tag{1}$$

Where $\psi_{a,b}(t)$ is the successive wavelet, a=scalar or frequency factor, b=a time factor; R=domain of the real numbers.

If $\psi_{a,b}(t)$ satisfies Eq. (1), for the time series f (t)∈ L^2(R) or finite energy signal, the successive wavelet transform of $f(t)$ is defined as (Rosso et al. 2004; Zhou et al. 2008; Kisi 2009):

$$W_\Psi f(a,b) = |a|^{-\frac{1}{2}}\int_R f(t)\Psi\left(\frac{t-b}{a}\right)dt \tag{2}$$

Where ψ (t) = complex conjugate function of $f(t)$. It is clear from Eq. (2) that the wavelet transform is the decomposition of $f(t)$ at different resolution level.

The successive wavelet is often discrete in real application. Let $a=a_0^{-j/2}\int_R f(t)\Psi(a_0^{-j}t-kb_0)dt$ are integer number. The discrete wavelet transform of can be written as

$$W_\Psi f(j,k) = a_0^{-j/2}\int_R f(t)\ \Psi\left(a_0^{-j}t-kb_0\right)dt \tag{3}$$

The most common preference for the parameter and is 2 and 1 time steps, respectively. This power of two logarithmic scaling of time and scale is known as a dyadic grid arrangement and is the simplest and most efficient case for practical purposes (Mallat 1989). Equation (3) becomes a binary wavelet transform when a_0=2, b_0=1 (Kisi 2009)

$$W_{\Psi}f(j,k) = 2^{-j/2}\int_R f(t)\Psi\left(2^{-j}t-k\right)dt \tag{4}$$

The characteristics of the original time series in frequency (a or j) and time domain (b or k) are reflected at the same time by $W\psi\, f(a,b)$ or $W\psi\, f(j,k)$. When the frequency resolution of the wavelet transform is low, but the time domain resolution is high, a or j becomes small. When the frequency resolution of the wavelet transform is high, but time domain resolution is low a or j become large (Wang and Ding 2003; Kisi 2009).

For a discrete time series $f(t)$, where occurs at different time t (i.e., here integer time steps are used), the DWT can be defined as:

$$W_{\Psi}f(j,k) = 2^{-j/2}\sum_{t=0}^{N-1} f(t)\Psi\left(2^{-j}t-k\right) \tag{5}$$

Where $W\psi\, f(j,k)$ is the wavelet coefficient for the discrete wavelet at scale $a=2^j, b=2^jk$.

DWT operates two sets of functions viewed as high-pass and low-pass filters. The original time series are passed through high-pass and low-pass filters and separated at different scales. The time series is decomposed into two: one comprising its trend (the approximation) and the other comprising high frequencies and the fast events (Kisi 2009).

2.5 ANN—an overview

Artificial neural networks (ANNs) have their capability in representing any nonlinear processes by given sufficient complexity of the trained networks and are discussed in detail in a number of hydrologic papers because of (Minns and Hall 1996; Sudheer et al. 2002; Senthil Kumar et al. 2005; Vos and Rientjes 2008; Srinivasulu and Jain 2009). An artificial neural network consists of simple synchronous element, called neurons, which are analogous to the biological neurons in the human brain. The architecture of a feed forward ANN can have many layers where a layer represents a set of parallel neurons. These neurons are arranged in layers in a network. The neurons in one layer are connected to those in the adjacent layers and strength of connection between the two neurons in adjacent layers is called "weight." There are weights on each of the inter-connections and it is these weights that are altered during the training process to ensure that the inputs produce an output that is close to the desired value with an appropriate training rule being used to adjust the weights in accordance with the data that are presented to the network (Ajmera and Goyal 2012). An ANN normally consists of three layers, an input layer, a hidden layer, and an output layer. In a feed-forward network, the weighted connections feed activations only in the forward direction from an input layer to the output layer. Each node in a layer receives and processes weighted input from a previous layer and transmits its output to nodes in the following layer through links. The incoming data are processed by nonlinear functions at hidden and output layers to get the output. The commonly used nonlinear function is logsigmoid function (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a, b).

Table 2 Model input variables

Model	Approach	Input Variables
WRM1	WR	$R(t)=f(R[t-1])$
ANNM1	ANN	
WRM2	WR	$R(t)=f(R[t-1],R[t-2])$
ANNM2	ANN	

The feed forward ANN is generally adapted in all studies because of its applicability to a variety of different problems (Hsu et al. 1995; Rumelhart et al. 1986). Maier and Dandy (2000) report that not more than one hidden layer is required in feed forward networks because a three-layer network can generate arbitrarily complex decision regions. Also, the appropriate input vector to the ANN model can be identified according to the procedure of Sudheer et al. (2002). Back propagation is the most popular algorithm used for the training of the feed forward ANNs (Thirumalaiah and Deo 1998; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a; Jain and Srinivasulu 2004; Fernando and Shamseldin 2009). The Levenberg-Marquardt is a trust region based method with hyper-spherical trust region (Burney et al. 2005). The Levenberg-Marquardt algorithm uses the following approximation to obtain the revised weights:

$$X_{k+1} = X_k - \left[J_k^T J + \mu I\right]^{-1} J^T e \tag{6}$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases; e is a vector of network errors and I is an identity matrix.

The Levenberg–Marquardt algorithm is based on approaching second-order training speeds without having the computation of Hessian matrix. When the scalar μ is very large, the Levenberg–Marquardt algorithm approximates as the steepest descent method. However, when μ is small, it is the same as the Gauss–Newton method. Since the Gauss–Newton method converges faster and more accurately towards an error minimum, the goal is to shift towards the Gauss–Newton method as quickly as possible (Hagan and Meghraj 1994). Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a

Table 3 The RMSE, N-S index, and R statistics of WR and ANN models

Location	Model	Calibration			Validation		
		Correlation coefficient	RMSE	N-S index	Correlation coefficient	RMSE	N-S index
Kamrup	WRM1	0.75	201.12	0.55	0.74	208.42	0.53
	WRM2	0.83	174.44	0.66	0.84	175.02	0.67
	ANNM1	0.72	206.47	0.52	0.71	216.44	0.50
	ANNM2	0.69	215.87	0.48	0.62	240.07	0.38
Jorhat	WRM1	0.74	114.99	0.53	0.75	105.44	0.54
	WRM2	0.84	96.58	0.67	0.83	90.58	0.66
	ANNM1	0.71	118.84	0.50	0.70	110.56	0.49
	ANNM2	0.68	123.32	0.46	0.67	116.24	0.44
Barpeta	WRM1	0.74	162.72	0.53	0.74	164.36	0.53
	WRM2	0.83	141.80	0.65	0.83	140.82	0.66
	ANNM1	0.61	189.87	0.37	0.63	189.66	0.38
	ANNM2	0.74	159.62	0.55	0.74	161.70	0.55
Cachar	WRM1	0.74	190.06	0.53	0.75	166.76	0.55
	WRM2	0.80	174.55	0.60	0.84	144.45	0.67
	ANNM1	0.64	211.73	0.41	0.66	188.77	0.43
	ANNM2	0.69	200.55	0.47	0.71	177.28	0.50
Darrang	WRM1	0.75	157.89	0.55	0.74	164.05	0.53
	WRM2	0.82	141.02	0.64	0.84	136.68	0.68
	ANNM1	0.72	162.95	0.52	0.71	169.17	0.50
	ANNM2	0.75	155.66	0.56	0.72	166.55	0.52
Dhemaji	WRM1	0.74	112.67	0.53	0.74	106.16	0.53
	WRM2	0.84	95.78	0.66	0.83	91.37	0.65
	ANNM1	0.72	114.29	0.52	0.71	110.15	0.50
	ANNM2	0.47	145.31	0.22	0.48	136.34	0.23
Dhubri	WRM1	0.76	132.57	0.56	0.74	135.16	0.53
	WRM2	0.83	119.14	0.64	0.83	115.55	0.66
	ANNM1	0.72	138.74	0.52	0.71	138.40	0.51
	ANNM2	0.79	123.21	0.62	0.77	125.28	0.60
Dibrugarh	WRM1	0.73	117.17	0.52	0.73	108.50	0.52
	WRM2	0.82	101.42	0.64	0.82	95.29	0.63
	ANNM1	0.70	120.86	0.49	0.70	112.74	0.48
	ANNM2	0.72	117.03	0.52	0.70	112.30	0.49
Golaghat	WRM1	0.75	118.45	0.54	0.75	109.22	0.55
	WRM2	0.84	100.41	0.67	0.85	92.21	0.68
	ANNM1	0.71	124.14	0.50	0.72	113.90	0.51
	ANNM2	0.73	119.71	0.53	0.73	111.79	0.53
Golapara	WRM1	0.75	155.49	0.55	0.74	158.60	0.53
	WRM2	0.84	132.70	0.67	0.83	138.29	0.65
	ANNM1	0.70	164.52	0.49	0.69	168.65	0.47
	ANNM2	0.78	144.28	0.61	0.77	149.29	0.59
Hailakandi	WRM1	0.74	182.98	0.53	0.75	163.31	0.55
	WRM2	0.80	169.78	0.59	0.83	143.54	0.65
	ANNM1	0.66	199.67	0.44	0.68	179.07	0.45
	ANNM2	0.70	191.24	0.48	0.70	174.58	0.48
Karbi Anglong	WRM1	0.75	181.45	0.55	0.75	170.38	0.56
	WRM2	0.81	167.20	0.62	0.86	139.47	0.70
	ANNM1	0.69	197.05	0.47	0.70	183.59	0.49

Table 3 (continued)

Location	Model	Calibration			Validation		
		Correlation coefficient	RMSE	N-S index	Correlation coefficient	RMSE	N-S index
	ANNM2	0.52	231.15	0.27	0.51	220.62	0.26
Karimganj	WRM1	0.71	172.77	0.48	0.75	145.44	0.54
	WRM2	0.81	147.04	0.62	0.83	127.74	0.64
	ANNM1	**0.61**	**190.74**	**0.37**	**0.64**	**166.11**	**0.40**
	ANNM2	0.67	177.71	0.45	0.71	151.77	0.50
Kokrajhar	WRM1	0.76	146.27	0.57	0.76	144.01	0.57
	WRM2	0.82	134.80	0.63	0.86	118.31	0.71
	ANNM1	0.71	156.40	0.51	0.72	152.86	0.51
	ANNM2	0.79	137.33	0.62	0.79	134.56	0.62
Lakhimpur	WRM1	0.75	111.46	0.54	0.75	107.16	0.54
	WRM2	0.85	93.22	0.68	0.84	90.85	0.67
	ANNM1	0.72	113.65	0.52	0.72	109.56	0.52
	ANNM2	0.77	106.13	0.59	0.74	106.64	0.54
Nagaon	WRM1	0.77	106.13	0.59	0.74	106.64	0.54
	WRM2	0.82	182.24	0.62	0.85	158.53	0.70
	ANNM1	0.66	224.52	0.43	0.62	228.54	0.38
	ANNM2	0.68	219.17	0.46	0.64	224.44	0.41
Naibhari	WRM1	0.74	163.14	0.53	0.74	166.76	0.53
	WRM2	0.82	143.43	0.64	0.83	142.24	0.65
	ANNM1	0.71	166.91	0.51	0.71	170.52	0.50
	ANNM2	0.49	207.04	0.24	0.50	209.83	0.25
North Chahar Hills	WRM1	0.75	196.44	0.54	0.76	175.04	0.56
	WRM2	0.81	179.98	0.62	0.86	146.49	0.70
	ANNM1	0.68	213.56	0.46	0.70	190.34	0.49
	ANNM2	0.71	204.12	0.50	0.74	181.19	0.53
Sibsagar	WRM1	0.74	113.93	0.53	0.74	105.22	0.53
	WRM2	0.83	98.39	0.65	0.83	91.36	0.65
	ANNM1	0.70	118.57	0.49	0.70	110.23	0.49
	ANNM2	0.75	110.71	0.56	0.72	107.28	0.51
Sonitpur	WRM1	0.75	124.38	0.55	0.74	125.77	0.54
	WRM2	0.85	105.15	0.68	0.84	104.79	0.68
	ANNM1	0.73	127.73	0.53	0.72	127.23	0.53
	ANNM2	0.77	118.56	0.59	0.74	124.07	0.55
Tisukia	WRM1	0.74	109.94	0.53	0.74	101.30	0.53
	WRM2	0.81	100.07	0.61	0.82	87.88	0.65
	ANNM1	0.69	115.29	0.48	0.69	107.44	0.47
	ANNM2	0.72	111.18	0.52	0.70	106.33	0.48

tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm (Senthil Kumar et al. 2012).

3 Results and discussion

To forecast the rainfall at several stations in Assam, the monthly rainfall data of 102 years was used. The first 60 years data were used for calibration of the model, and the remaining 42 years data were used for validation. Kamrup station is used as representative station. In this study, wavelet regression (WR) models were developed with help of discrete wavelet transform (DWT) and linear regression (LR). The WR model is an LR model which uses sub-time series components obtained using DWT on original rainfall data. Mallat DWT algorithm was used to decompose original rainfall time series into a certain number of sub-time series components (Ds;

Mallat 1989). WR was constructed in which the Ds. of the original input time series are the input of the LR and the original output time series are the output of the LR. The model inputs were decomposed by wavelets as shown in Table 2 and decomposed sub-series were taken as input to LR model.

Root mean square error (RMSE), N-S index, and correlation coefficient (R) statistics were used to evaluate the accuracy of the WR model. These measures are defined below.

A. Coefficient of Correlation: The Coefficient of Correlation (CC) can be defined as:

$$CC = \frac{\sum_{i=1}^{N}\left(Y_o - \overline{Y}_o\right)\left(Y_c - \overline{Y}_c\right)}{\left[\sum_{i=1}^{N}\left(Y_o - \overline{Y}_o\right)^2 \cdot \sum_{i=1}^{N}\left(Y_c - \overline{Y}_c\right)^2\right]^{1/2}} \tag{7}$$

B. Root mean square error: the root mean square error (RMSE) between observed and computed outputs can be defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(Y_c - Y_o\right)^2}{N}} \tag{8}$$

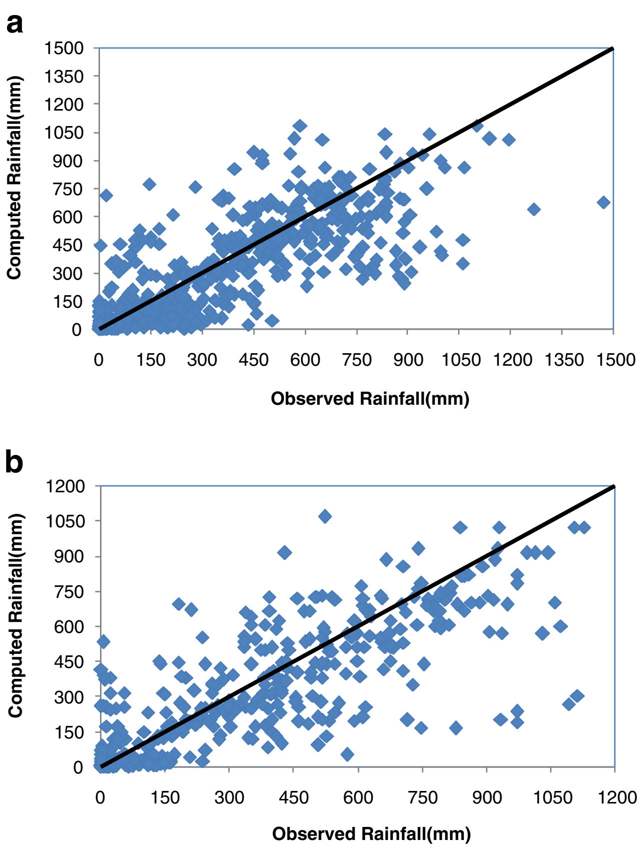

Fig. 3 Scatter plot for the result of WRM2 model during **a** calibration and **b** validation for Kamrup station

C. Nash–Sutcliffe efficiency index: The Nash–Sutcliffe efficiency index (η) can be defined as:

$$\eta = 1 - \frac{\sum_{i=1}^{N}\left(Y_c - Y_o\right)^2}{\sum_{i=1}^{N}\left(Y_o - \overline{Y}_o\right)^2} \tag{9}$$

where N represents the number of instances, Y_o and Y_c denote the observed and the computed values respectively, $\overline{Y}_o$ is the mean of the observed variables.

ANN was trained using back propagation with LM algorithm. The optimal number of hidden neurons for ANN was determined as several by trial and error procedure. The final structure of the ANN model is the number of inputs used in ANN, 7 hidden neurons and 1 output neuron. The values of the performance criteria from various models for both calibration and validation are presented in various tables (Table 3). The calibration and validation results are compared with the performance indices of the various best models. Kamrup station is used as representative station for analyzing purposes.

3.1 Calibration /training results

It can be observed from Table 3 that the model WRM2 using WR algorithm outperformed all other models investigated in

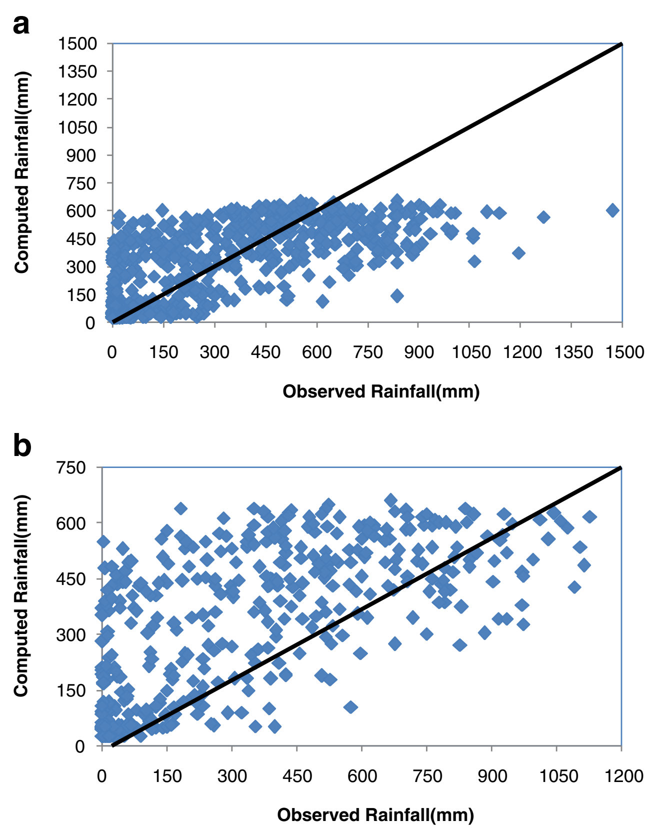

Fig. 4 Scatter plot for the result of ANNM2 model during **a** calibration and **b** validation for Kamrup station

this study. In terms of RMSE, model ANNM2 using ANN performed worst (215.87) and WRM2 model performed the best (174.44). The coefficient of correlation for the WRM2 model was 0.83, whereas the value of the coefficient of correlation for the ANN with LM back propagation model ANNM2 was 0.69. In terms of N-S index, WRM2 model using performed best (65.81 %) while model ANNM2 performed the worst (47.64 %).

3.2 Validation/testing results

Model WRM2 using WR algorithm outperformed all other models investigated in this study. RMSE was 175.02 for WRM2 while the model ANNM2 performed worst (240.07). The best value of correlation coefficient was 0.84 from WRM2 model while the model ANNM2 performed worst (0.62). In term of N-S index, WRM2 performed best (67.20 %) while ANNM2 performed worst (38.29 %).

From Table 3, it is found that low RMSE and high correlation coefficient values for WR models when compared to ANN models. Figures 3 and 4 show the observed and modeled graphs during calibration and validation for WR and ANN models respectively for representative Kamrup station. It is found that values modeled from WR model properly matched with the observed values, whereas ANN model underestimated the observed values. From this analysis, it is evident that the performance of WR was much better than ANN models in forecasting the rainfall.

Models using 2 lag rainfalls (R[$t-1$],R[$t-2$]) as input data were performed well in comparison to models using single lag rainfall data (R[$t-1$]). It can also be inferred from Table 3 that stations located in upper Assam performed much better than compared to the other parts of state. RMSE varied from 90.85 for Lakhimpur station to 220.62 for Karbi Anglong for validation dataset. Some stations such as for Karbi Anglong, North Cachar Hills etc. performed poor. This may correspond to hilly area locations and decrease in forest cover due to shifting cultivation (FSI 2009).

4 Conclusion

This paper reports a wavelet regression model for time series modeling of monthly rainfall at 21 stations in Assam, India. The proposed model is a combination of wavelet analysis and regression approach. Wavelet decomposes the time series into multilevel of details and it can adopt multi resolution analysis and effectively diagnose the main frequency component of the signal and abstract local information of the time series. The time series data of rainfall was decomposed into sub series by DWT. Appropriate sub-series of the variable used as inputs to the LR model and original time series of the variable as output. Model parameters are calibrated using 60 years of data and rest of the data is used for model validation. From this analysis, it was found that efficiency index is more than 67 % for WR models whereas it is 50 % for ANN models respectively for representative Kamrup station. It is hoped that this analysis will provide input data for a management system and to enable the development of optimal water allocation policies and management strategies for water and agriculture manager to bridge the gap between water needs and obtainable water supply.

References

Ajmera TK, Goyal MK (2012) Development of stage discharge rating curve using model tree and neural networks: an application to peachtree creek in Atlanta. Expert Systems With Applications, Elsevier Ltd. 39(5):5702–5710

Antonios A, Constantine EV (2003) Wavelet exploratory analysis of the FTSE ALL SHARE index. In proceedings of the 2nd WSEAS international conference on non-linear analysis. Non-linear systems and Chaos, Athens.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) Artificial neural networks in hydrology-I: Preliminary concepts. J Hydrol Engrg 5(2):115–123

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b) Artificial neural networks in hydrology-II: Hydrologic applications. J Hydrol Engrg 5(2):124–137

Baratta D, Cicioni G, Masulli F, Studer L (2003) Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting. Neural Networks 16:375–387

Burney SMA, Jilani TA, Ardil C (2005) Levenberg-Marquardt Algorithm for Karachi Stock Exchange Share Rates Forecasting. World Acad Sci Eng Technol 3:171–176

Chien-Ming C (2011) Wavelet-based multi-scale entropy analysis of complex rainfall time series. Entropy 13:241–253. doi:10.3390/e13010241

Chien-ming C (2013) Enhanced accuracy of rainfall–runoff modeling with wavelet transform. J Hydroinf 15(2):392–404

Chou CM, Wang RY (2002) On-line estimation of unit hydrographs using the wavelet-based LMS algorithm. Sci J Hydrol 47(5):721–738

Coulibaly P, Burn HD (2004) Wavelet analysis of variability in annual Canadian streamflows. Water Resour Res 40, W03105

Fernando DAK, Shamseldin AY (2009) Investigation of the internal functioning of the radial basis function neural network river flow forecasting models. J Hydrol Eng 14(3):286–292. doi:10.1061/(ASCE)1084-0699(2009)14:3(286)

FSI, 2009. State of Forest Report, 2009. Forest Survey of India (FSI), Ministry of Environment and Forests, Government of India

Grossman A, Morlet J (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. SIAM J Math Anal 15:723–736

Hagan MT, Menhaj MB (1994) Training feed forward networks with the Marquardt algorithm. IEEE Trans Neural Netw 6:861–867

Hsu K-L, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. Water Resour Res 31(10):2517–2530

Jain A, Srinivasulu S (2004) Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms, and artificial neural network techniques. Water Resour Res 40(4):W04302 doi:10.1029/2003WR002355

Jain SK, Kumar V (2012) Trend analysis of rainfall and temperature data for India. Curr Sci 102(1):37–49
Jain SK, Das A, Srivastava DK (1999) Application of ANN for reservoir inflow prediction and operation. J Water Resour Plan Manag ASCE 125(5):263–271
Kisi O (2009) Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. Hydrol Process 23: 3583–3597
Kisi O (2011) Wavelet regression model as an alternative to neural networks for river stage forecasting. Water Resour Manag 25(2):579–600
Labat D, Ababou R, Mangin A (2000) Rainfall–runoff relations for karstic springs. Part II. Continuous wavelet and discrete orthogonal multiresolution analyses. J Hydrol 238(3–4):149–178
Labat D, Ronchail J, Guyot JL (2005) Recent advances in wavelet analyses. Part 2—Amazon, Parana, Orinoco and Congo discharges time scale variability. J Hydrol 314(1–4):289–311
Lindsay RW, Percival DB, Rothrock DA (1996) The discrete wavelet transform and the scale analysis of the surface properties of sea ice. IEEE Trans Geosci Remote Sens 34:771–787
Mahanta et al. (2003) Assam Human Development Report 2003, available at http://hdr.undp.org/en/reports/nationalreports/asiathepacific/india/name,3268,en.html. Accessed 20 July 2013
Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw 15:101–124
Mallat SG (1989) A theory for multi resolution signal decomposition: the wavelet representation. IEEE T Pattern Anal Mach Intell 11(7): 674–693
Minns AW, Hall MJ (1996) Artificial neural networks as rainfall runoff models. Hydrol Sci J 41(3):399–418
Partal T, Kucuk M (2006) Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region. Phys Chem Earth 31:1189–1200, Turkey
Quiroz R, Yarlequé C, Posadas A, Mares V, Immerzee Walter W (2011) Improving daily rainfall estimation from NDVI using a wavelet transform. Environ Model Softw 26:201–209
Raman H, Chandramauli V (1996) Deriving a general operating policy for reservoirs using neural network. J Water Resour Plan Manag ASCE 122(5):342–347
Rosso OA, Figliola A, Blanco S, Jacovkis PM (2004) Signal separation with almost periodic components: a wavelets based method. Rev Mex Fis 50(1):179–186
Rumelhart DE, Hinton E, Williams J (1986) Learning internal representation by error propagation, Parallel Distributed Processing, Vol. 1. MIT Press, Cambridge, pp 318–362
Sang Y (2013) Improved wavelet modeling framework for hydrologic time series forecasting. Water Resour Manag 27(8): 2807–2821
Santos Celso AG, Freire Paula KMM (2012) Analysis of Precipitation Time Series of Urban Centers of Northeastern Brazil using Wavelet Transform. World Acad Sci Eng Technol 67:845–850
Senthil Kumar AR, Ojha CSP, Goyal MK, Singh RD, Swamee PK, (2012) Modelling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic and decision tree algorithms. ASCE's J Hydrol Eng 17(3):394–404
Senthil Kumar AR, Sudheer KP, Jain SK, Agarwal PK (2005) Rainfall-runoff modelling using artificial neural networks: comparison of network types. Hydrol Process 19:1277–1291
Srinivasulu S, Jain A (2009) River flow prediction using an integrated approach. J Hydrol Engrg, ASCE 14(1):75–83
Sudheer KP, Gosain AK, Ramasastri KS (2002) A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrol Process 16:1325–1330
Tantanee S, Patamatammakul S, Oki T, Sriboonlue V, Prempree T, (2005) Coupled wavelet-autoregressive model for annual rainfall prediction. J Environ Hydrol 13
Thirumalaiah K, Deo MC (1998) River stage forecasting using artificial neural networks. J Hydrol Engrg 3(1):26–32
Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteorol Soc 79(1):61–78. doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2
Vos NJ, Rientjes THM (2008) Multiobjective training of artificial neural networks for rainfall-runoff modelling. Water Resour Res 44(W08434):1–15
Wang W, Ding J (2003) Wavelet network model and its application to the prediction of the hydrology. Nature and Science 1(1):67–71
Xingang D, Ping W, Jifan C (2003) Multiscale characteristics of the rainy season rainfall and interdecadal decaying of summer monsoon in North China. Chin Sci Bull 48:2730–2734
Zhou HC, Peng Y, Liang G-H (2008) The research of monthly discharge predictor–corrector model based on wavelet decomposition. Water Resour Manag 22(2):217–227