Journal of Water Resource and Protection, **2013, 5, 395-404**
http://dx.doi.org/10.4236/jwarp.2013.54039 Published Online April 2013 (http://www.scirp.org/journal/jwarp)

Coupling Singular Spectrum Analysis with Artificial Neural Network to Improve Accuracy of Sediment Load Prediction

Sokchhay Heng, Tadashi Suetsugi
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu, Japan
Email: heng_sokchhay@yahoo.com

Received January 8, 2013; revised February 19, 2013; accepted February 28, 2013

ABSTRACT

Sediment load estimation is generally required for study and development of water resources system. In this regard, artificial neural network (ANN) is the most widely used modeling tool especially in data-constraint regions. This research attempts to combine SSA (singular spectrum analysis) with ANN, hereafter called SSA-ANN model, with expectation to improve the accuracy of sediment load predicted by the existing ANN approach. Two different catchments located in the Lower Mekong Basin (LMB) were selected for the study and the model performance was measured by several statistical indices. In comparing with ANN, the proposed SSA-ANN model shows its better performance repeatedly in both catchments. In validation stage, SSA-ANN is superior for larger Nash-Sutcliffe Efficiency about 24% in Ban Nong Kiang catchment and 7% in Nam Mae Pun Luang catchment. Other statistical measures of SSA-ANN are better than those of ANN as well. This improvement reveals the importance of SSA which filters noise containing in the raw time series and transforms the original input data to be near normal distribution which is favorable to model simulation. This coupled model is also recommended for the prediction of other water resources variables because extra input data are not required. Only additional computation, time series decomposition, is needed. The proposed technique could be potentially used to minimize the costly operation of sediment measurement in the LMB which is relatively rich in hydrometeorological records.

Keywords: Artificial Neural Network; Singular Spectrum Analysis; Coupled Model; Sediment Load; Mekong Basin

1. Introduction

Quantification of sediment load is necessary for study and development of water resources system such as reservoir storage, dam, irrigation/navigation channel, soil and water conservation measure, environmental impact assessment, etc. [1-5]. Sediments are the end products of land surface erosion governed mainly by hydrometeorology, topography, geology and land use/cover [1,2]. Sediment data are lacking for rivers in many areas of the world, especially in developing and remote regions [6]. However, it can be estimated with the aid of modeling approaches. The hydrologic and terrain conditions of a river basin change spatio-temporally and this causes difficulties in determining their effects on sediment erosion and transport. This drawback has encouraged the application of black box models, e.g. artificial neural network (ANN). ANN forecasts outputs using experiences learned from historical data. Its application can be found in many sectors including finance, medicine, water resources, and so forth. There are many types of ANN and the recognized ones are feedforward, kohonen and hopfield networks [7]. In predicting and forecasting water resources variables, feedforward networks are almost exclusively applied [8]. The term "ANN" used in this paper is referred to feedforward artificial neural network.

The ANN model is commonly used in river basins with data scarcity because it does not require detailed physical information of the system. By just providing hydrometeorogical information as inputs, ANN can predict sediment load at the watershed outlets with high accuracy. Kisi and Shiri [5] applied ANNs to predict suspended sediment concentration (*SSC*) in Eel River (USA) with rainfall and discharge as inputs and obtained very satisfactory results with Nash-Sutcliffe Efficiency (*NSE*) between 0.80 and 0.84 in validation stage. Sediment

yield of various sub-watersheds in Kapgari River Basin (India) is modeled well by ANN (input: rainfall and temperature) with *NSE* ranging from 0.76 to 0.83 in validation stage [9]. In Pari River (Malaysia), ANNs (input: discharge) perform very well in simulating suspended sediment load (*SSL*) with *NSE* equal to 0.99 and 1.00 in validation stage [10]. ANN can be employed also to analyze the hysteretic phenomenon of sediment transport [11]. It is a very practical and promising modeling tool in the context of sediment load prediction [12] and its outputs can be potentially used for design and management purposes in water-related development projects [7].

Although ANN has been proved to perform well in modeling sediment load and other hydrological variables, many researches have been carried out further in order to improve its accuracy by coupling with other methods. Sediment load is generally predicted by using hydrometeorological variables and the most common of which are rainfall and discharge. Naturally, the time series of such variables are very noisy due to the effects of climate variation and other human activities. Thus, one common way to improve the prediction accuracy of ANN is to perform some pre-processing of the inputs and this requires another method. This kind of technique is known as a coupled approach which has been getting more interest recently. Kisi [13] developed a range-dependent neural network (RDNN) for predicting sediment load at two stations operated by the US Geological Survey. RDNN splits the original data series into three ranges which are afterward used as ANN inputs. In term of model efficiency measured by determination coefficient (R^2), RDNN is slightly better than ANN for larger R^2 about 0.5% at Santa Clara Station, and both models perform comparably at Calleguas Station. If considering root mean square error (*RMSE*) and mean absolute error (*MAE*), RDNN is much better than ANN at both stations.

Selection of a method for input pre-processing should match ideally the specific learning problems. In this study, singular spectrum analysis (SSA) was proposed because it is generally seen as an adaptive noise-reduction algorithm [14]. SSA is a tool decomposing a time series into a number of components with simple structures, which can be often identified as trends, seasonality and other oscillatory series, or noise components, and it does not require any statistical assumptions while performing the analysis [15,16]. The application of SSA in analyzing hydrometeorological time series (e.g. rainfall, discharge, temperature) can be found in Hanson *et al.* [17] and Marques *et al.* [18]. This method can be used particularly to extract the main components of rainfall and discharge series and to provide good forecast for them [18]. Sivapragasam *et al.* [14] combined SSA with the support vector machine method (the latter called SSA-SVM approach) to predict rainfall at Station 23 (Singapore) and runoff from Tryggevælde catchment (Denmark), and the results were compared with those of the non-linear prediction (NLP) method. For rainfall prediction, SSA-SVM performs much better than NLP for less *RMSE* 36% in calibration stage and 28% in validation stage. For runoff prediction, SSA-SVM is also superior to NLP for less *RMSE* 64% in calibration phase and 59% in validation phase. To our knowledge, there are no any studies associating SSA with ANN for predicting sediment load yet.

The present study attempts to combine SSA with ANN, hereafter called SSA-ANN model, for sediment load prediction with expectation to obtain more accurate results than using ANN alone. The specific objectives are to examine the application of the SSA-ANN model in predicting monthly average $SSL\left(SSL_m\right)$ and compare its performance with that of the existing ANN approach. The case study was firstly tried in Ban Nong Kiang (BNK) catchment. In order to show consistency, another case study was conducted in Nam Mae Pun Luang (NMPL) catchment. Both catchments are located in the Lower Mekong Basin (LMB).

2. Materials and Methods

2.1. Study Catchments

The LMB is a trans-boundary river basin which partially covers four Southeast Asian countries: Lao PDR, Thailand, Cambodia and Vietnam. This basin is relatively rich in hydrometeorological records except sediment [19]. As illustrated in **Figure 1**, BNK catchment is located in the western part of the basin and drains approximately 1405 km^2. The elevation in this catchment decreases from north to south and varies from just over 1300 m to about 200 m. The average catchment slope is around 22.5%. Rainfall in this area is influenced by the southwest monsoon (May-October) blowing from Bay of Bengal bringing humid and hot weather. From November to April, this period is known as dry season. Natural topography and mountain ranges make this catchment oriented in a leeward direction creating a rain shadow and therefore little rainfall amount, about 1080 mm/year. The annual discharge is 17.10 m^3/s. Sediment yield in BNK catchment is around 44 t/year/km^2. The dominant land use is tree cover or forest and the dominant soil type is Orthic Acrisols.

Situated in the northwest of the LMB (**Figure 1**), NMPL catchment has a drainage area of about 260 km^2. The feature of catchment topography is west-east gradient with elevation varying between 510 and 1670 m. The average catchment slope is approximately 32.5%. This catchment receives rainfall around 1950 mm annually and produces an average discharge of 2.22 m^3/s. Rainfall pattern in this area is influenced by the southwest mon-

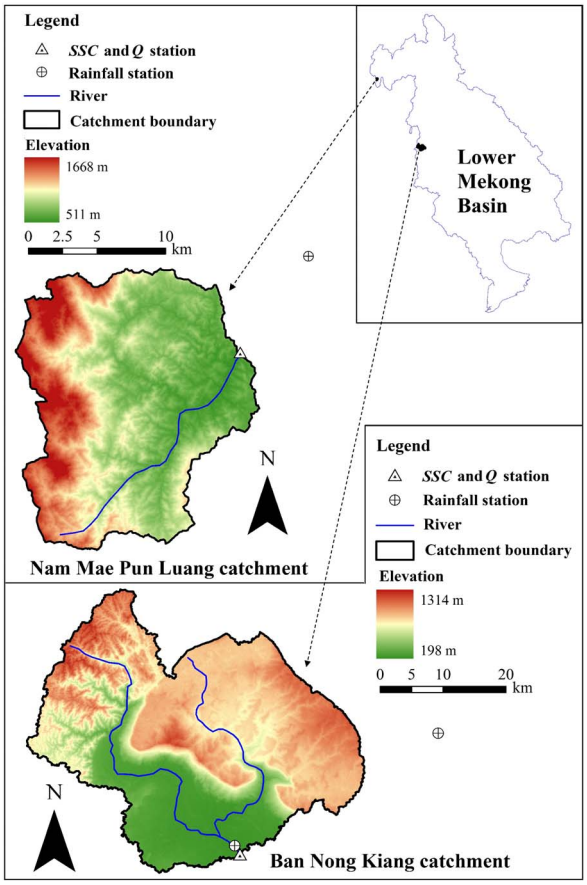

Figure 1. Map of the study catchments.

soon as well but the amount is much higher than that in BNK catchment because NMPL catchment is oriented in windward direction. Sediment yield of this catchment is about 58 t/year/km^2. The larger sediment yield can be explained by topographic feature of each catchment. Mosaics and shrub cover dominates the land use in the catchment and the dominant soil type is Orthic Acrisols.

Poor gauging stations in term of data availability and completeness are commonly found in developing and remote regions as located the LMB. These two catchments were selected based on data availability: 20 years (1982-2003, no data in 1986 and 1987) in BNK catchment and 22 years (1980-2001) in NMPL catchment.

2.2. Data

The main data used in this study are rainfall (R), discharge (Q) and suspended sediment load (SSL). SSL is the product of Q and SSC. The daily time series of R, Q and SSC were obtained from Mekong River Commission. R and Q series are continuous but SSC series are discontinuous with few samples per month. The average sampling frequency in BNK and NMPL catchment is about 2 and 4 samples per month, respectively. This provokes the study in monthly basis. The monthly average $SSL(SSL_m)$ is the product of monthly average $Q(Q_m)$ and monthly average $SSC(SSC_m)\cdot Q_m$ and R_m (monthly average rainfall) were employed as inputs for model calculation and SSL_m was used for comparison with the model outputs. Due to data limitation, the model inputs consist of only R_m and Q_m. Rainfall and discharge are the main erosion and transport agents [1,2] and both variables are generally used in many existing researches. Some case studies (e.g. Mustafa *et al.* [10], Memarian and Balasundram [20]) employ only discharge or rainfall as the input. There are no other reasons besides data unavailability. However, the model accuracy must pass the minimum satisfactory level. In this study, the entire dataset in each catchment was divided into two parts, the first 75% for calibration and the remaining 25% for validation. This combination (75×25) is very common in the study of sediment modeling [21].

The effect of land use changes and other human activities might cause great variation of sediment load over the simulation period (about 20 years) and this could lead to low accuracy of the model results. Based on the Mann-Kendall and the Pettit test (0.01 significance level) on the SSL annual series, there are no significant trends and change points detected at any of the two catchments. Therefore, it can be concluded that the SSL_m data series used in this study have no significant influence from the said effects.

At the catchment outlets, it is very likely that there is lag-time between R and SSL as well as Q and SSL due to clockwise hysteretic effect [22-24]. Hence, consideration of R and Q from previous time step could improve the model accuracy. The present study was conducted in monthly time scale. Therefore, the consideration of antecedent R and Q would have no much effect on the model results because the lag-time is just few days. Melesse *et al.* [3] applied ANN model to simulate daily and weekly SSL in Mississippi and Missouri River (USA) by considering two different input combinations (I1 and I2). I1 includes one-day antecedent Q and I2 does not. As a result in daily basis, the model prediction using I1 is just slightly better than the one using I2. NSE (I1) is larger than NSE (I2) about 6% in Mississippi River and 3% in Missouri River. In weekly basis, the model efficiency decreases dramatically in comparing with the daily simulation and NSE (I1) becomes less than NSE (I2) in Missouri River. Similar situation is also observed in the case study of four rivers in Turkey conducted by Kisi *et al.* [25]. In consequence, the model performance will not be much different for monthly time scale simulation and the reason that this research does not take into account the antecedent R and Q.

2.3. ANN and SSA-ANN Model

ANN is a flexible and potential tool in determining non-

linear processes such as sediment transport. The main differences of ANN structures are network architectures, training algorithms and transfer/activation functions. In this study, the multi-layer perceptron with the back-propagation algorithm and sigmoid transfer function was selected. This kind of structure is commonly used in water resources modeling and provides better results than others [7,8,26]. As presented in **Figure 2**, the designed model structure composes of 1 input, 1 hidden and 1 output layer. The input layer has two nodes, one for R_m and another for Q_m. The number of nodes in the hidden layer was determined by trial and error method because so far, there is no guideline for this purpose. The single node in the output layer is SSL_m.

Firstly, each input node receives a set of input data (x) and in this case R_m and Q_m. The connections between the input and hidden layer contain weights (w) which are determined through the system training. Then, in the hidden layer, the weighted average of input (z) is computed by using summation functions [21]:

$$z = \sum_{i=1}^{n} w_i x_i + \beta \tag{1}$$

where w_i is the weight vector, x_i is the input vector $(i = 1, 2, \cdots, n)$ and β is the bias term. Afterward, z is transferred to y (output) and in this case SSL_m, through the sigmoid transfer function [21]:

$$y = \frac{1}{1 + \mathrm{e}^{-z}} \tag{2}$$

In the output layer, y (the predicted SSL_m) is compared with the target value (the observed SSL_m) in order to detect the error or difference between the predicted and observed SSL_m. Subsequently, the error is corrected by adjusting w. After assigning the new w, the same calculation steps are performed. This procedure is repeated until obtaining a desirable y or acceptable level of error. To sum up, the ANN model training is a process of weight adjustment attempting to produce a desirable outcome with minimum residuals.

For the SSA-ANN model, the methodologies are similar to those of ANN but a new form of R_m and Q_m was accounted as inputs instead of their original one. SSA was applied to decompose the original dataset of R_m and Q_m into a number of components which are then input to the ANN model for predicting SSL_m. The SSA algorithm for one-dimensional time series analysis consists of 1) transformation of the original time series $F = (f_1, \cdots, f_N)$ to multi-dimensional series $X_i = (f_i, \cdots, f_{i+L-1})^{\mathrm{T}}$ where $i = 1, \cdots, K = N - L + 1$, by means of one-parameter (window length L) delay procedure; 2) singular value decomposition of the trajectory matrix $X = [X_1 : \cdots : X_N]$ into a sum of rank-one biorthogonal elementary matrices $X = X_i + \cdots + X_L$; 3) split of the elementary matrices into m groups and within each group, determination of the summed matrices $X = Y_1 + \cdots + Y_m$; and 4) transfer of each summed matrix into a new dimensional series of the same length N. The first two steps make up the decomposition stage and the remaining two do the reconstruction stage. In short, the initial time series F is decomposed into the sum of m time series: $F = F_1 + \cdots + F_m$. The basic concept and detailed methodology of SSA can be found in Golyandina *et al.* [15].

In this paper, the original time series of R_m and Q_m were decomposed into two components. Since this is the first trial study, a number of components other than two were not examined because many components would provoke difficulty (time consuming) in training the model. Optimizing the number of components is subjected to future study. In addition, interpreting physical meaning of each extracted component is beyond the scope of this research. The main purpose here is to examine the potential of SSA in combination with ANN for SSL_m prediction. The model structure of SSA-ANN designed for this particular study is illustrated in **Figure 3**.

2.4. Model Evaluation and Comparison

The efficiency of each model was measured by *NSE* which is the most widely used goodness-of-fit indicator in predictive hydrological models. Basically, *NSE* compares the residual variance with the observed data variance and at the same time, it also reflects the prediction accuracy of the modeling approach in comparing with the observed mean value [27]. Negative *NSE* indicates that the observed mean value is a better predictor than the model being used. With *NSE* greater than 0.50, the model performance is judged as satisfactory [28]. *NSE*,

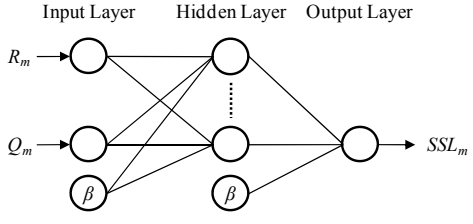

Figure 2. ANN model structure.

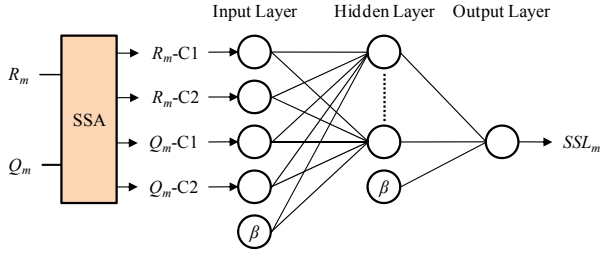

Figure 3. SSA-ANN model structure.

RMSE and *MAE* were applied to optimize the model architecture (number of hidden nodes) of ANN and SSA-ANN. The optimum architecture should contain the highest value of *NSE* and the lowest value of *RMSE* and *MAE*. *RMSE* punishes severely the high values while *MAE* is more sensitive to the moderate values. These three indicators were also employed for model comparison. Since total *SSL*(*SSL*$_t$) is important for dam-reservoir management [13,26], the model performance for this purpose was also investigated and absolute percentage bias (*APBIAS*) was used as an indicator. *SSL*$_t$ is the integral of *SSL*$_m$ series within a particular period (calibration or validation period). The model result of *SSL*$_t$ prediction is considered as acceptable if *APBIAS* is less than 55% [28]. *NSE*, *RMSE*, *MAE* and *APBIAS* were calculated respectively using Equations (3)-(6) [21,28]:

$$NSE = 1 - \frac{\sum (O-P)^2}{\sum (O-O_{avg})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum (O-P)^2} \quad (4)$$

$$MAE = \frac{1}{n}\sum |O-P| \quad (5)$$

$$APBIAS = 100 \times \frac{|O_t - P_t|}{O_t} \quad (6)$$

where O is the observed SSL_m with the mean value O_{avg}, P is the predicted SSL_m, n is the sample size, O_t is the observed SSL_t, and P_t is the predicted SSL_t.

3. Results and Discussion

3.1. Statistical Analysis of Datasets

Figure 4 shows the results of SSA in decomposing R_m and Q_m in BNK catchment. For the case of R_m (**Figure 4(a)**), the first component (C1) behaves lower frequency then the second one (C2) and it is also apparent that SSA removes the discontinuity characterized by many zeros (dry periods) existing in the original time series. For the case of Q_m (**Figure 4(b)**), time series of C1 contains lower frequency than that of C2 as well. From **Figure 4**, it is clearly seen that C1 is the main component. This situation is also found in NMPL catchment. The statisti-

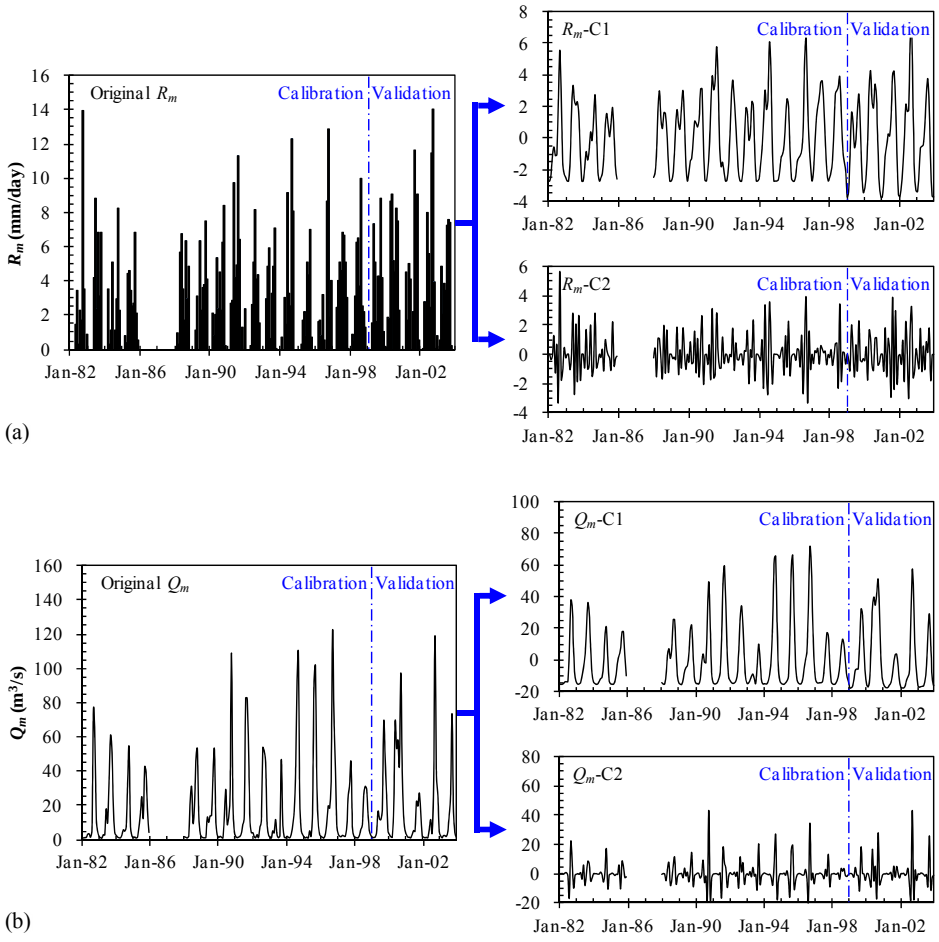

Figure 4. Results of SSA in BNK catchment (no data in 1986 and 1987), (a) R_m and (b) Q_m.

cal difference between the original time series and its extracted components is presented as below.

The results of statistical analysis for both calibration and validation datasets are summarized in **Table 1** (ANN datasets) and **Table 2** (SSA-ANN datasets), and the statistical parameters are the maximum (*Max*), minimum (*Min*), average (*Mean*), correlation coefficient (*CC*) between the inputs and the observed outputs, standard deviation (*SD*) and skewness coefficient (*SKEW*). *SD* is a measure of how widely the data are dispersed from the average value (*Mean*) while *SKEW* indicates the degree of asymmetry of a data distribution [29]. A data normal distribution is corresponding to *SKEW* value about zero. In BNK catchment, the extent of validation datasets (both ANN and SSA-ANN datasets) overall is within the range of calibration datasets. Although there are some over-extrapolations, e.g. the upper bound of R_m dataset of ANN (14.07 mm/day in validation stage and 13.93 mm/day in calibration stage), it is not significant. Discharge generally exhibits higher *CC* than rainfall and this suggests that SSL_m is more dependent on discharge. Since C1 is the main component (both rainfall and discharge), it

Table 1. Statistical characteristics of ANN datasets.

Catchment	Parameter	Calibration			Validation		
		R_m	Q_m	SSL_m	R_m	Q_m	SSL_m
		(mm/day)	(m^3/s)	(t/day)	(mm/day)	(m^3/s)	(t/day)
Ban Nong Kiang (BNK)	*Max*	13.93	122.20	2449.54	14.07	119.34	1718.50
	Min	0.00	1.00	0.18	0.00	1.14	0.11
	Mean	2.70	16.57	173.50	3.66	18.69	154.53
	CC	0.59	0.85	-	0.55	0.91	-
	SD	2.99	24.18	371.91	3.56	25.48	350.67
	SKEW	1.26	2.23	3.54	0.86	2.08	3.37
Nam Mae Pun Luang (NMPL)	*Max*	24.82	9.18	273.60	17.82	8.16	435.42
	Min	0.00	0.35	0.40	0.00	0.36	0.30
	Mean	5.53	2.23	35.48	4.74	2.22	55.74
	CC	0.58	0.89	-	0.49	0.81	-
	SD	5.32	1.77	50.99	4.60	1.70	80.40
	SKEW	0.91	1.65	2.45	0.91	1.64	2.34

Max: Maximum; *Min*: Minimum; *Mean*: Average; *CC*: Correlation coefficient; *SD*: Standard deviation; *SKEW*: Skewness coefficient.

Table 2. Statistical characteristics of SSA-ANN datasets.

Catchment	Parameter	Calibration					Validation				
		R_m (mm/day)		Q_m (m^3/s)		SSL_m	R_m (mm/day)		Q_m (m^3/s)		SSL_m
		C1	C2	C1	C2	(t/day)	C1	C2	C1	C2	(t/day)
Ban Nong Kiang (BNK)	*Max*	6.33	5.66	71.08	43.31	2449.54	7.15	3.90	57.07	43.29	1718.50
	Min	−2.72	−3.31	−15.45	−18.90	0.18	−3.72	−3.01	−17.67	−17.56	0.11
	Mean	−0.01	0.00	−0.08	0.00	173.50	−0.06	0.00	−0.28	−0.01	154.53
	CC	0.59	0.32	0.78	0.65	-	0.52	0.36	0.79	0.69	-
	SD	2.25	1.37	19.79	7.89	371.91	2.76	1.47	20.65	9.75	350.67
	SKEW	0.56	0.76	1.70	1.66	3.54	0.40	0.38	1.21	2.09	3.37
Nam Mae Pun Luang (NMPL)	*Max*	11.75	7.51	5.63	2.77	273.60	9.30	4.39	3.94	2.32	435.42
	Min	−5.56	−4.72	−1.81	−1.59	0.40	−4.80	−4.61	−1.58	−1.37	0.30
	Mean	−0.02	0.00	0.00	0.00	35.48	−0.06	−0.01	0.00	0.00	55.74
	CC	0.59	0.24	0.85	0.52	-	0.48	0.23	0.74	0.58	-
	SD	4.43	1.98	1.51	0.55	50.99	3.85	1.62	1.46	0.54	80.40
	SKEW	0.47	0.74	1.14	1.38	2.45	0.63	0.28	1.14	1.39	2.34

Max: Maximum; *Min*: Minimum; *Mean*: Average; *CC*: Correlation coefficient; *SD*: Standard deviation; *SKEW*: Skewness coefficient; C1: Component 1 (main component); C2: Component 2.

therefore has higher *CC* value than C2. The value of *SD* and *SKEW* is generally low. It should be noted that high value of *SD* and *SKEW* will cause negative effect on the model performance [3,30]. The *SD* and *SKEW* value of the calibration datasets are rather comparable with the corresponding ones of the validation datasets. This is appropriate for modeling because the great difference will lead to poor model performance in validation stage [30]. Remarkably, the SSA-ANN inputs are characterized by lower *SD* and *SKEW* value than the ANN inputs and this condition is favorable to the model simulation. This reveals the potential of SSA in statistical point of view.

In NMPL catchment, the inputs of both ANN and SSA-ANN in validation period do not extend beyond the range of the corresponding ones in calibration period. It is contradictory for SSL_m in which over-extrapolation is significant for the upper bound (435.42 t/day in validation period and 273.60 t/day in calibration period). If excluding this particular event (435.42 t/day), both data ranges become similar. Therefore, this sole unfavorable data point would have no much effect on the model results. This event occurred in August (2001) which is the rainy season. Moreover, NMPL catchment is characterized by steep slope terrain. In consequence, this particular event might associate with local extreme phenomenon (e.g. slope failure, debris flow) occurring episodically and bringing huge amount of sediment in a short time. For the case of lower bound, the difference is not significant. Similar situation is observed for *CC*. Both calibration and validation datasets also contain low *SD* and *SKEW* value and behave similar characteristics. The effect of SSA is the same as observed in BNK catchment.

3.2. Model Performance in BNK Catchment

The performance of each model is summarized in **Table 3**. It can be seen that not only ANN but also SSA-ANN model yields satisfactory results for both SSL_m and SSL_t prediction because *NSE* and *APBIAS* values are respectively greater than 0.50 and less than 55%. *NSE* and *APBIAS* of ANN are correspondingly equal to 0.81 and 5.06% in calibration stage, and 0.52 and 48.04% in validation stage. SSA-ANN contains respectively *NSE* and *APBIAS* value about 0.84 and 0.09% in calibration period, and 0.64 and 38.25% in validation period. The predicted SSL_m resulted from each model is graphically compared with the observed data as depicted in **Figure 5(a)**. Visually, the predicted time series of both models show similar trend with the observed one. **Figure 5(b)** (ANN) and **Figure 5(c)** (SSA-ANN) depict the scatter plots of the predicted versus observed SSL_m which were used to distinguish the model performance in estimating low, medium and high value. In order to clearly investigate the whole extent, from low to high value, both figures were plotted in log-log scale. These two scatter plots obviously demonstrate that both models overestimate the low values. In case of medium and high values, the scattering points are distributed uniformly around the ideal fit line. SSA-ANN predicts better not only the low but also the medium and high SSL_m through reduction of the overestimates at low value and the underestimates at medium and high value. The better prediction of SSA-ANN at medium and high value can be confirmed respectively by the less *MAE* and *RMSE* value (**Table 3**).

For SSL_m prediction, SSA-ANN is superior to ANN for more *NSE* 4%, less *RMSE* 9% and less *MAE* 22% in calibration stage. In validation stage, SSA-ANN is better for more *NSE* 24%, less *RMSE* 14% and less *MAE* 18%. In case of SSL_t prediction, SSA-ANN is more powerful for less *APBIAS* 98% in calibration phase and 20% in validation phase.

3.3. Model Performance in NMPL Catchment

From **Table 3** and **Figure 6**, similar situation is observed. Both models also perform well in this catchment and the advantage of SSA-ANN over ANN also exists. For SSL_m prediction, SSA-ANN is superior to ANN for more *NSE* 1%, less *RMSE* 4% and less *MAE* 3% in calibration stage. In validation stage, SSA-ANN is better for more *NSE* 7%, less *RMSE* 4% and less *MAE* 2%. In case of SSL_t prediction, SSA-ANN is more powerful for less *APBIAS* 65% in calibration phase and 6% in validation phase. The advantage of SSA-ANN in this catchment is rather less in

Table 3. Model performance indicated by *NSE*, *RMSE*, *MAE* and *APBIAS*.

Catchment	Model	Architecture	Calibration				Validation			
			NSE	*RMSE* (t/day)	*MAE* (t/day)	*APBIAS* (%)	*NSE*	*RMSE* (t/day)	*MAE* (t/day)	*APBIAS* (%)
Ban Nong Kiang	ANN	2-2-1	0.81	160.44	92.43	5.06	0.52	242.11	128.04	48.04
	SSA-ANN	4-2-1	0.84	146.34	72.04	0.09	0.64	209.39	105.50	38.25
Nam Mae Pun Luang	ANN	2-3-1	0.88	17.71	10.55	1.88	0.50	56.48	32.10	36.33
	SSA-ANN	4-3-1	0.89	17.03	10.26	0.66	0.54	54.41	31.44	33.99

NSE, *RMSE* and *MAE* for evaluating SSL_m prediction; *APBIAS* for evaluating SSL_t prediction; Architecture (optimum): Number of nodes in the input-hidden-output layer.

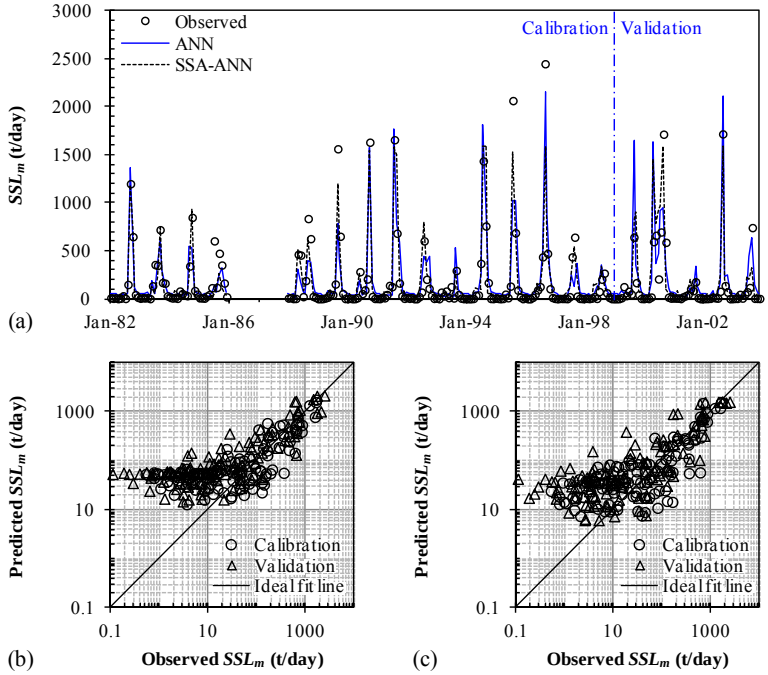

Figure 5. Comparison of the predicted versus observed SSL_m in BNK catchment (no data 1986 and 1987), (a) Time series comparison; (b) Scatter plot of ANN results and (c) Scatter plot of SSA-ANN results.

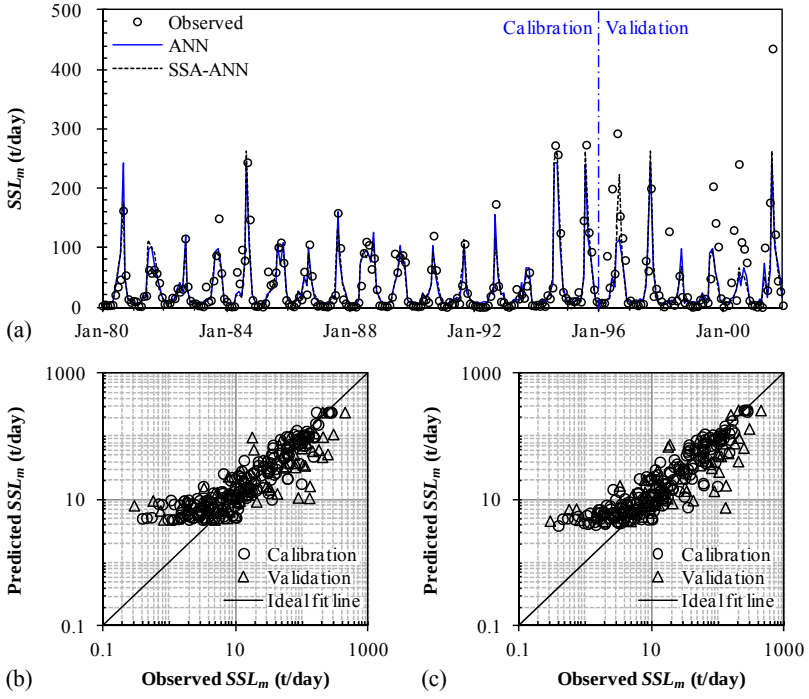

Figure 6. Comparison of the predicted versus observed SSL_m in NMPL catchment, (a) Time series comparison; (b) Scatter plot of ANN results; and (c) Scatter plot of SSA-ANN results.

comparing with that in BNK catchment. This is because the ANN inputs (original datasets) in NMPL catchment are characterized by lower *SD* and *SKEW* value. Therefore, when transformed to become SSA-ANN inputs using SSA, they (*SD* and *SKEW*) are not decreased as much as in BNK catchment, especially C1 which is the main component. For instance, in calibration stage, the decreasing rate of *SKEW* from Q_m to Q_m-C1 is 56% in BNK catchment and it is just 49% in NMPL catchment. Similarly in validation stage, it is 54% and 30% in BNK and

NMPL catchment, respectively.

In calibration period, the efficiency (NSE) of both models in NMPL catchment is slightly better than that in BNK catchment. The difference in model performance between these two catchments may be attributed to different variation of sediment load spatially. This can be explained by the difference in *SD* and *SKEW* value. The SSL_m dataset in NMPL catchment is characterized by lower value of *SD* (50.99) and *SKEW* (2.45) and therefore easier to be calibrated. Looking into validation period, *NSE* value of both methods becomes less in comparing with that in BNK catchment. This could be due to different temporal variation of the SSL_m data which can be explained statistically by the difference between the calibration and validation dataset in each individual catchment. The more similar these two datasets is corresponding to the better model performance in validation period. The difference in *SKEW* value is likely comparable in both catchments but the difference in *SD* value is more significant in NMPL catchment.

4. Conclusions

This research proposed a coupled model (SSA-ANN) to predict sediment load in two catchments, located in the LMB, having different hydrological and terrain characteristics. The performance of this model was compared with that of the existing ANN approach. Satisfactory results were obtained from both methods but SSA-ANN exhibits its better performance repeatedly in both catchments. This improvement reflects the importance of SSA. SSA filters the noise containing in the raw time series. It reduces the value of *SD* and *SKEW*, and transforms the original input data to be near normal distribution which is favorable to modeling. Instead of ANN, the proposed SSA-ANN model is also recommended for the prediction of other water resources variables because extra input data are not required. Only additional computation, time series decomposition, is needed. This new technique could be potentially used to minimize the costly operation of sediment sampling in the LMB which is relatively rich in hydrometeorological records.

In this study, the model simulation was conducted in monthly basis. Therefore, other time scales should be tested. The present research employed SSA to decompose the raw inputs into two components only. Larger amount of components should be examined in order to extensively investigate the potential of SSA-ANN. The present authors expect that the model accuracy will be more improved with more number of components.

5. Acknowledgements

High gratitude is expressed to Japanese Government (Monbukagakusyo: MEXT) and Global Center of Excellent program of University of Yamanashi, Japan, for supporting this research. Sincerest thanks are extended to Mekong River Commission for providing datasets.

REFERENCES

[1] G. L. Morris and J. Fan, "Reservoir Sedimentation Handbook: Design and Management of Dams, Reservoirs, and Watershed for Sustainable Use," McGraw-Hill, New York, 1998.

[2] USBR (United States Bureau of Reclamation), "Erosion and Sedimentation Manual," USBR, Colorado, 2006.

[3] A. M. Melesse, S. Ahmad, M. E. McClain, X. Wang and Y. H. Lim, "Suspended Sediment Load Prediction of River Systems: An Artificial Neural Network Approach," *Agricultural Water Management*, Vol. 98, No. 5, 2011, pp. 855-866. doi:10.1016/j.agwat.2010.12.012

[4] A. Singh, M. Imtiyaz, R. K. Isaac and D. M. Denis, "Comparison of Soil and Water Assessment Tool (SWAT) and Multilayer Perceptron (MLP) Artificial Neural Network for Predicting Sediment Yield in the Nagwa Agricultural Watershed in Jharkhand, India," *Agricultural Water Management*, Vol. 104, 2011, pp. 113-120. doi:10.1016/j.agwat.2011.12.005

[5] O. Kisi and J. Shiri, "River Suspended Sediment Estimation by Climatic Variables Implication: Comparative Study among Soft Computing Techniques," *Computers & Geosciences*, Vol. 43, 2012, pp. 73-82. doi:10.1016/j.cageo.2012.02.007

[6] D. E. Walling and D. Fang, "Recent Trends in the Suspended Sediment Loads of the World's Rivers," *Global and Planetary Change*, Vol. 39, No. 1-2, 2003, pp. 111-126. doi:10.1016/S0921-8181(03)00020-1

[7] O. M. Rezapour, L. T. Shui and D. B. Ahmad, "Review of Artificial Neural Network Model for Suspended Sediment Estimation," *Australian Journal of Basic and Applied Sciences*, Vol. 4, No. 8, 2010, pp. 3347-3353.

[8] H. R. Maier and G. C. Dandy, "Neural Networks for the Prediction and Forecasting of Water Resources Variables: A Review of Modelling Issues and Applications," *Environmental Modelling & Software*, Vol. 15, No. 1, 1999, pp. 101-124. doi:10.1016/S1364-8152(99)00007-9

[9] G. Singh and R. K. Panda, "Daily Sediment Yield Modeling with Artificial Neural Network Using 10-Fold Cross Validation Method: A Small Agricultural Watershed, Kapgari, India," *International Journal of Earth Sciences and Engineering*, Vol. 4, No. 6, 2011, pp. 443-450.

[10] M. R. Mustafa, M. H. Isa and R. B. Rezaur, "A Comparison of Artificial Neural Networks for Prediction of Suspended Sediment Discharge in River—A Case Study in Malaysia," *World Academy of Science, Engineering and Technology*, Vol. 81, 2011, pp. 372-376.

[11] H. K. Cigizoglu, "Suspended Sediment Estimation for Rivers Using Artificial Neural Networks and Sediment Rating Curves," *Turkish Journal of Engineering and Environmental Sciences*, Vol. 26, No. 1, 2002, pp. 27-36.

[12] G. Tayfur, "Artificial Neural Networks for Sheet Sediment Transport," *Hydrological Sciences Journal*, Vol. 47,

No. 6, 2002, pp. 879-892. doi:10.1080/02626660209492997

[13] O. Kisi, "Development of Streamflow-Suspended Sediments Rating Curve Using a Range Dependent Neural Network," *International Journal of Science and Technology*, Vol. 2, No. 1, 2007, pp. 49-61.

[14] C. Sivapragasam, S.-Y. Liong and M. F. K. Pasha, "Rainfall and Runoff Forecasting with SSA-SVM Approach," *Journal of Hydroinformatics*, Vol. 3, No. 3, 2001, pp. 141-152.

[15] N. Golyandina, V. Nekrutkin and A. A. Zhigljavsky, "Analysis of Time Series Structure: SSA and Related Techniques," Chapman and Hall/CRC, Boca Raton, 2001. doi:10.1201/9781420035841

[16] GistaT Group, "Time Series Analysis and Forecasting," 2010. http://www.gistatgroup.com/cat/

[17] R. T. Hanson, M. W. Newhouse and M. D. Dettinger, "A Methodology to Assess Relations between Climatic Variability and Variations in Hydrologic Time Series in the Southwestern United States," *Journal of Hydrology*, Vol. 287, No. 1-4, 2004, pp. 252-269. doi:10.1016/j.jhydrol.2003.10.006

[18] C. A. F. Marques, J. A. Ferreira, A. Rocha, J. M. Castanheira, P. Melo-Goncalves, N. Vaz and J. M. Dias, "Singular Spectrum Analysis and Forecasting of Hydrological Time Series," *Physics and Chemistry of the Earth*, Vol. 31, No. 18, 2006, pp. 1172-1179. doi:10.1016/j.pce.2006.02.061

[19] H. J. Fuchs, "Data Availability for Studies on Effects of Land-Cover Changes on Water Yield, Sediment and Nutrient Load at Catchments of the Lower Mekong Basin," MRC-GTZ Cooperation Programme, Göttingen, 2004.

[20] H. Memarian and S. K. Balasundram, "Comparison between Multi-Layer Perceptron and Radial Basis Function Networks for Sediment Load Estimation in a Tropical Watershed," *Journal of Water Resource and Protection*, Vol. 4, No. 10, 2012, pp. 870-876. doi:10.4236/jwarp.2012.410102

[21] S. Heng and T. Suetsugi, "Using Artificial Neural Network to Estimate Sediment Load in Ungauged Catchments of the Tonle Sap River Basin, Cambodia," *Journal of Water Resource and Protection*, Vol. 5, No. 2, 2013, pp. 111-123. doi:10.4236/jwarp.2013.52013

[22] Y. Tramblay, A. St-Hilaire and T. B. M. J. Ouarda, "Frequency Analysis of Maximum Annual Suspended Sediment Concentrations in North America," *Hydrological Sciences Journal*, Vol. 53, No. 1, 2008, pp. 236-252. doi:10.1623/hysj.53.1.236

[23] P. Gao and M. Josefson, "Event-Based Suspended Sediment Dynamics in a Central New York Watershed," *Geomorphology*, Vol. 139-140, 2011, pp. 425-437. doi:10.1016/j.geomorph.2011.11.007

[24] S. Heng and T. Suetsugi, "Estimating Quantiles of Annual Maximum Suspended Sediment Load in the Tributaries of the Lower Mekong River," *Journal of Water and Climate Change*, Vol. 4, No. 1, 2013, pp. 63-76. doi:10.2166/wcc.2013.023

[25] O. Kisi, I. Yuksel and E. Dogan, "Modelling Daily Suspended Sediment of Rivers in Turkey Using Several Data-Driven Techniques," *Hydrological Sciences Journal*, Vol. 53, No. 6, 2008, pp. 1270-1285. doi:10.1623/hysj.53.6.1270

[26] O. Kisi, "Multi-Layer Perceptrons with Levenberg-Marquardt Training Algorithm for Suspended Sediment Concentration Prediction and Estimation," *Hydrological Sciences Journal*, Vol. 49, No. 6, 2004, pp. 1025-1040. doi:10.1623/hysj.49.6.1025.55720

[27] J. E. Nash and J. V. Sutcliffe, "River Flow Forecasting through Conceptual Models Part I-A Discussion of Principles," *Journal of Hydrology*, Vol. 10, No. 3, 1970, pp. 282-290. doi:10.1016/0022-1694(70)90255-6

[28] D. N. Moriasi, J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel and T. L. Veith, "Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations," *Transactions of the American Society of Agriculture and Biological Engineers*, Vol. 50, No. 3, 2007, pp. 885-900.

[29] M. Shahin, H. J. L. Van Oorschot and S. J. De Lange, "Statistical Analysis in Water Resources Engineering," A. A. Balkema, Rotterdam, 1993.

[30] T. Rajaee, "Wavelet and ANN Combination Model for Prediction of Daily Suspended Sediment Load in Rivers," *Science of the Total Environment*, Vol. 409, No. 15, 2010, pp. 2917-2928. doi:10.1016/j.scitotenv.2010.11.028

