Predicting precipitation on nonpoint source pollutant exports in the source area of the Liao River, China
Y. Wang, J. M. Bian, S. N. Wang and S. Y. Nie

ABSTRACT
The source area of the Liao River is an important grain growing area in China which experiences serious problems with agricultural nonpoint source pollution (NPS) which is impacting the regional economy and society. In order to address the water quality issues it is necessary to understand the spatial distribution of NPS in the Liao River source area. This issue has been investigated by coupling a wavelet artificial neural network (WA-ANN) precipitation model with a soil and water assessment tool (SWAT) model to assess the export of nonpoint source pollutants from the Liao River source area. The calibration and validation of these models are outlined. The WA-ANN models and the SWAT model were run to generate the spatial distribution of nonpoint source nutrient (nitrogen and phosphorus) exports in the source area of the Liao River. It was found that the SWAT model identified the sub-catchments which not only receive high rainfall but are also densely populated with high agricultural production from dry fields and paddy fields, which are large users of pesticides and chemical fertilizer, as the primary source areas for nutrient exports. It is also concluded that the coupled WA-ANN models and the SWAT model provide a tool which will inform the identification of NPS issues and will facilitate the identification of management practices to improve the water environments in the source area of the Liao River.

Key words | Jilin Province, nonpoint source pollution, precipitation, source area of the Liao River, SWAT model, wavelet artificial neural networks

INTRODUCTION
Nonpoint source pollution (NPS) is increasingly a great concern due to its adverse impacts on receiving waters and resulting impacts on society. From the beginning of the 1970s, the study of the mobilization, transport and fate of NPS and of its management using mitigation actions and measures has progressed significantly (Lee et al. 2010; Ma et al. 2011). However, the factors which influence NPS processes still require further research. NPS includes two processes. The first process is the mobilization and transport of NPS to a drainage line of watercourse, while the second process relates to the transportation and transformation process within a channel or watercourse (Li et al. 2014). It is well known that precipitation, which is the driving force behind all hydrologic processes, can mobilize and transport nonpoint source pollutants to receiving waters.

A number of models of NPS have been developed based on mathematical methods that describe the physical processes to predict runoff, erosion and transport of sediments and nutrients from watersheds under different conditions. The soil and water assessment tool (SWAT) is one of the most complete models because it is able to predict runoff, erosion and the transport of sediments and associated pollutants over long periods of time in complex watersheds subject to variations in soil type, land use, fertilizer and pesticide application (Arnold & Allen 1995). Many studies have used SWAT to evaluate the impacts of precipitation on NPS (Muhammed & Atilla 2012; Wu & Liu 2012; Liu & Lu 2015). Hao studied the nitrogenous NPS loads exported from various land use types under different precipitation levels in the Hei River basin. Cheng & Hao (2006) found that the export rates of particulate (adsorbed) nitrogen in descending order were from farmlands, towns, grasslands, bushes and forests. In the case of dissolved nitrogen, the export rates in descending order of magnitude were from farmlands, grasslands, bushes, forests and towns (Cheng & Hao 2006).
Accurate and reliable precipitation forecasting methods can help ensure the accuracy of the SWAT inputs. Methods of precipitation forecasting have been widely researched. Traditional methods include the autoregressive integrated moving average model, the time series model, the support vector machines model, and the Markov chain model. These models limit the complexity of the evolution of precipitation, which can result in lower forecast accuracy. In the last 5–6 years the application of wavelet artificial neural networks (WA-ANN) for forecasting of precipitation has gained attention (Hou & Lu 2013). Examples include Kisi (2009), who explored the use of WA-ANN models for daily flow forecasting of intermittent rivers, while Wang et al. (2013) developed a WA-ANN model to forecast the inflow at the Three Gorges Dam on the Yangtze River. The WA-ANN approach is able to tolerate the presence of chaotic components better than most alternative methods. The high performance of artificial neural networks (ANN) is due to the ability to take into account all variables, internal and external, as well as the relationships between each other (Adamowski 2011).

The upper catchment (which is also termed the source area) of the Liao River, which is an important grain growing area in China, experiences serious problems with agricultural NPS, which is impacting the regional economy and society (Bian et al. 2014). In this paper, the coupling of a WA-ANN precipitation model with a SWAT model to assess the export of NPS from the Liao River source area is outlined.

**STUDY AREA AND DATA**

**Study area**

The Liao River source area is located in the southwest of Jilin Province (123° 31’–125° 42’ E, 34° 44’–42° 08’ N), China (see Figure 1). This source area covers the Liao River basin, the Zhaosutai River basin and the Tiaozi River basin. The study area has a temperate continental monsoon climate. The spatial and temporal distribution of precipitation across the study area is affected by atmospheric circulation, topography and other factors. The average annual precipitation is 545 mm. Approximately 80% of the annual precipitation falls during the months of June, July, August and September.

Dry fields occupy around 70% of the total land area. The descending order of the land use area is woodlands, towns and paddy fields. Grassland, water and others are less than 5% (see Figure 2). In recent years, human agricultural activity has degraded grasslands, and desertification is becoming a serious problem which is adversely impacting on the local environment and degrading the water quality in the Liao River source area. In order to address the water quality issues it is necessary to understand the spatial
distribution of NPS in the Liao River source area (Zhao 2014).

Digital elevation model, land use and soil data

The digital elevation model (DEM) used in this study was obtained from the international scientific data service platform (https://wist.echo.nasa.gov/) which provides digital elevation data at a 90 m by 90 m resolution. The DEM was used to delineate the watersheds and to analyze the drainage patterns within the study area.

The types and spatial distribution of land use were obtained from Landsat TM (Thematic Mapper) and ETM (Enhanced Thematic Mapper) image data mosaics acquired in 2008 using geometric corrections, image enhancement, cutting and preprocessing using erdas9.0.

The soil data were obtained from the second soil survey in Jilin Province and was processed using the same steps which were followed when preparing land use maps.

Climate data

The model is run using daily precipitation, maximum/minimum temperatures, solar radiation, wind speed and relative humidity input from observed data or generated during a simulation. Meteorological data from eight stations were obtained from the China Meteorological data sharing service network (https://cdc.cma.gov.cn/home.do/) and/or the Siping City Bureau of Hydrology. The meteorological stations were located at Changchun, Siping, Shuangliao, Liaoyuan, Quantai, Erlongshan Reservoir, Wangben and Lishu. The period of record was from 2005 to 2015. Future precipitation data were predicted using the WA-ANN approach.

METHODOLOGY

The WA-ANN model

The WA-ANN model combines wavelet transformation theory (WA) with an ANN (Wang et al. 2009; Adamowski & Sun 2010). One of the advantages of the WA-ANN method compared with the ANN method is its ability to identify irregular components in a data series with multi-level wavelet decomposition. WA-ANN consists of:

(i) an input layer with neurons representing input variables to the problem,
(ii) an output layer with neurons representing the dependent variables, and
(iii) one or more hidden layers containing neurons to help capture the nonlinearity in the data.

A typical three-layer feed forward WA-ANN structure is shown in Figure 3. The input signal is transmitted through the network in a forward direction, layer by layer. The connections between neurons in different layers are supplied by adjusted weighting values. Neuron output is calculated as:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + \varepsilon\right)$$

$$f(y) = \cos\left(1.75 \frac{x - b}{a}\right) \exp\left[-0.5 \left(\frac{x - b}{a}\right)^2\right]$$
where \( w \) is the weight matrix, \( x \) is the input matrix, \( e \) is the bias, \( n \) is the number of inputs, \( f \) is the activation function (a morlet function was used as the activation function in the hidden layer neurons in this study), and \( a \) and \( b \) are respectively the stretch factor and the shift factor in the wavelet function.

Weights are calculated by minimizing the error function \( (E) \). According to Maier & Dandy (2000) the error function is calculated as:

\[
E = \frac{1}{2p} \sum_{p=1}^{P} \sum_{i=1}^{n} (d_i - y_i)^2 \tag{3}
\]

where \( p \) is the number of training patterns; \( n \) is the number of output neurons; \( d_i \) and \( y_i \) are measured and predicted outputs.

This is undertaken using an iterative optimization algorithm such as back propagation (BP) that computes the derivatives of the training error with respect to those weights and biases. The BP algorithm is essentially a gradient descent technique that minimizes the network error. In this study, the gradient descent algorithm was used to train the WA-ANN model when forecasting precipitation because it is fast, accurate and reliable.

**SWAT model**

The SWAT was developed by Arnold at the United States Department of Agriculture (Arnold & Allen 1993). SWAT is a continuous, spatially distributed simulation of the hydrologic cycle and agricultural pollutant mobilization and transport at a catchment scale (Kim & Park 2015). SWAT has been widely used to predict the impact of agriculture and land management practices on water, sediment, and agricultural chemical yields in large complex watersheds over extended periods of time. The major components of SWAT include routines for weather, hydrology, soil temperature, plant growth, nutrients, pesticides and land management practices. SWAT takes into account surface runoff, percolation, lateral subsurface flow, groundwater return flow, evapotranspiration and channel transmission losses. The runoff volume is estimated using a modified SCS curve number method. SWAT uses a storage routing algorithm to estimate flow through each soil layer in the root zone. Downward flow occurs when the field capacity of a soil layer is exceeded and the layer below is not saturated. SWAT partitions groundwater into two aquifer systems: a shallow unconfined aquifer which contributes return flows to streams within a watershed and a deep, confined aquifer that, except for pumping, is disconnected from the system. SWAT simulates nitrogen and phosphorus washes off by runoff and leaching through the soil profile, while a fixed nitrate concentration is adopted for the shallow aquifer when estimating the groundwater contribution to the in-stream nitrogen load. The in-stream transformation of nitrogen and phosphorus is estimated using the QUAL2E model, which includes the major interactions of the nutrient cycles, algae production, and benthic oxygen demand. Runoff, sediment and chemical exports are calculated for each hydrological response unit and then aggregated at the sub-basin level before being routed to the watershed outlet. The time of concentration for a watershed is calculated from the flow velocities calculated for overland flow and channel flow using Manning’s formula. A detailed description of SWAT can be found in Neitsch et al. (2002).

**MODEL DEVELOPMENT**

**Training and testing the WA-ANN model**

WA-ANN models were used to forecast future precipitation at eight locations. Eleven years of monthly average daily precipitation data (from January 2005 to December 2015) collected at eight rainfall gauges located at Changchun, Siping, Shuangliao, Liaoyuan, Quantai, Erlongshan Reservoir, Wangben and Lishu were used to train and test the WA-ANN models. The data series were each divided into a 10-year training record (from January 2005 to December 2014) and a 1-year testing record (January 2015 to December 2015). WA-ANN models were trained using the MATLAB neural network toolbox.

For all of the rainfall gauges, the best WA-ANN models were a function of the monthly average daily precipitation from the previous 2 months in the current year and the same month in the previous year, 2 years before and 3 years before. These observed five monthly average daily precipitations were input and then passed into the hidden layer after multiplying with connection weights, before generating the output, which is the monthly average daily precipitation in the current year. For example, the monthly average daily precipitation recorded at a gauge in January 2005, January 2006, January 2007, November 2007, and December 2007 were the input used to estimate the monthly average daily precipitation in January 2008. Then at each step the first vector of the input matrix was omitted and the predicted output vector from the previous step was substituted into
the input matrix. Each model was then tested on a trial and error basis to determine the optimum number of neurons in the hidden layer based on different combinations of variables in the model’s input layer. The optimum number of hidden layer neurons was found to be eight for all the rainfall gauges.

In this paper, the values recorded at Shuangliao station in 2015 are compared with the calculated values of monthly average daily precipitation in Table 1. It can be seen that the best WA-ANN model for the Shuangliao station had a maximum relative error of 8.7% (in November 2015) for the predicted monthly average daily precipitation. A similar level of accuracy was obtained from each of the WA-ANN models at the seven other stations (the relative error is less than 10%). It was concluded that the WA-ANN models achieved a good level of agreement between observed and predicted monthly average daily precipitation at all eight stations.

Sensitivity analysis, calibration and validation of the SWAT model

The sensitivity analysis method implemented in SWAT is called the Latin Hypercube One-factor-At-a-Time (LH-OAT) approach proposed by Morris (1991). The LH sampling approach is a stratified sampling approach that better covers the sampling hypercube with fewer samples (McKay et al. 1979). However, the OAT approach needs many samples to cover the full range of parameter values (Van Griensven et al. 2006).

The sensitivity analysis found that the SCS runoff curve number (CN2), the soil saturated hydraulic conductivity (SOL_K), the soil conservation measures factor (USLE_P) and the average slope length (SLSUBBSN) are the most sensitive parameters.

The SWAT model was calibrated by determining the set of model parameters which achieved the best goodness-of-fit between observed and predicted values. In this study, the Nash-Sutcliffe (NS) efficiency coefficient and the correlation coefficient ($R^2$) were used to evaluate the goodness-of-fit of the SWAT model parameters. It was considered that when $R^2 \geq 0.6$ and $NS \geq 0.5$, then the model simulation results are reliable and could be used for prediction. NS and $R^2$ coefficients are defined as follows:

$$NS = 1 - \frac{\sum_{i=1}^{n} (Q_o - Q_p)^2}{\sum_{i=1}^{n} (Q_o - Q_o)^2}$$

$$R^2 = \frac{\left[\sum_{i=1}^{n} (Q_p,i - Q_o,i)(Q_o,i - Q_o)\right]^2}{\sum_{i=1}^{n} (Q_o,i - Q_o)^2 \sum_{i=1}^{n} (Q_p,i - Q_o)^2}$$

where $Q_o$ is the predicted value; $Q_o$ is measured value; $Q_p$ is mean predicted value; $Q_p$ is mean measured value and $n$ is number of observations.

A SWAT model was established for the three small watersheds in the source area of the Liao River, the Dongliang River watershed, the Zhaosutai River watershed and the Tiaozi River watershed, respectively. Due to the lack of observed flow data in the Tiaozi River, only the Quantai gauging station (on the Dongliang River) and the Lishu gauging station (on the Zhaosutai River) were selected to calibrate and validate the model. The calibration period was 2006–2008 while the verification period was 2009–2010. A comparison of the measured and predicted monthly average daily flows for the calibration period and the validation period are shown in Figures 4 and 5, respectively.

During the calibration period, the $R^2$ and NS coefficient values at the Quantai station were 0.82 and 0.79, respectively, while the $R^2$ and NS coefficient values at the Lishu station were 0.81 and 0.77, respectively. During the validation period, the $R^2$ and NS coefficient values at the Quantai station were 0.82 and 0.78, respectively, while the $R^2$ and NS coefficient values at the Lishu station were 0.79 and 0.75, respectively. Based on the adopted evaluation criteria, it was considered that the level of accuracy of the predicted monthly average daily flows was acceptable.

The only sediment data recorded during the calibration and validation periods were collected at the
Quantai station. A comparison of the measured and predicted monthly total sediment discharge in tonnes for the calibration period and the validation period is shown in Figure 6.

During the calibration period the $R^2$ and NS coefficient values for monthly total sediment discharge at the Quantai station were 0.83 and 0.79, respectively, while the $R^2$ and NS coefficient values at the Quantai station during the
validation period were 0.80 and 0.75, respectively. The predicted results are satisfactory given the level of accuracy of the predicted monthly total sediment discharge.

The only nonpoint source nutrient data recorded during the calibration and validation periods were collected at the Quantai station. A comparison of the measured and predicted monthly total nonpoint source nutrient (nitrogen and phosphorus) exports in tonnes for the calibration period and the validation period is shown in Figures 7 and 8, respectively.

During the calibration period, the \( R^2 \) and NS coefficient values for monthly total nonpoint source nutrient (nitrogen) exports at the Quantai station were 0.83 and 0.80, respectively, while the \( R^2 \) and NS coefficient values for monthly total nonpoint source nutrient (phosphorus) exports at the Quantai station were 0.79 and 0.73, respectively. During the validation period, the \( R^2 \) and NS coefficient values for monthly total nonpoint source nutrient (nitrogen) exports at the Quantai station were 0.81 and 0.77, respectively, while the \( R^2 \) and NS coefficient values for monthly total nonpoint source nutrient (phosphorus) exports were 0.78 and 0.71, respectively. The predicted results are satisfactory given the level of accuracy of the predicted monthly total nonpoint source nutrient (nitrogen and phosphorus) exports.

**RESULTS AND DISCUSSION**

**Modeled nonpoint source pollutant exports under different precipitation levels**

In the paper, the main purpose is based on the calibration and validation of the SWAT model and analysis of the main influencing factors of nonpoint source nutrient (nitrogen and phosphorus) exports. In China, hydrological regions are based on the annual precipitation, which is divided into five categories in 200 mm intervals, namely, 0–200 mm (dry), 200–400 mm (less water), 400–800 mm (flat water), 800–1,600 mm (more water) and >1,600 mm (wet). Analysis of observed data in the source area of the Liao River found that when precipitation is less than 200 mm there is very little surface runoff and the export of nonpoint source pollutants is negligible, so these areas can be excluded from further consideration. It was also found that when annual precipitation is more than 1,200 mm, then nonpoint source pollutant exports are broadly constant. Consequently, the assessment of the variation in nonpoint source nutrient exports under conditions where the annual rainfall varies from 200 to 1,200 mm was investigated. The calculated average annual nitrogen and phosphorus unit exports...
as a function of annual rainfall and land use are Tables 2 and 3, respectively. Figure 9 implies that the export values represent the ‘middle’ of the range, given that the values are plotted on a line, which simply reinforces that the export rates should vary within each precipitation range and land use.
It can be seen that nitrogen and phosphorus loads increase with precipitation when it is more than 200 mm; this is due to the increase of surface runoff, resulting in the ability to enhance the transport of pollutants. The results of the study also show that the export of phosphorus is less than the export of nitrogen for all land uses. For nitrogen and phosphorus, the descending order of pollution exports is dry fields, paddy fields, towns and woodlands. This is due to the use of pesticides and fertilizers, resulting in nonpoint source nutrient exports in dry fields and paddy fields being higher than in other land uses. Nonpoint source nutrient exports in towns are mainly affected by human activities, i.e. garbage piles and wastewater discharges. The vegetation coverage is higher in woodlands, and has a certain retention influence on the amount of soil erosion, greatly reducing the nonpoint source nutrient exports. It describes that farm land is the major land use type, and agricultural activities have become the main components of NPS in the Liao River source area.

Forecasting runoff and nonpoint source pollutant exports in the period 2005 to 2020

The WA-ANN model was run to estimate monthly average daily precipitation for all eight meteorological stations within

| Table 2 | Average annual nitrogen exports as a function of precipitation and land use |
|-----------------|--------|--------|--------|--------|
| Precipitation (mm) | Dry field (kg/yr/ha) | Paddy field (kg/yr/ha) | Town (kg/yr/ha) | Woodland (kg/yr/ha) |
| 200–400 | 1.66 | 0.60 | 0.50 | 0.11 |
| 400–600 | 2.30 | 1.08 | 0.86 | 0.19 |
| 600–800 | 5.44 | 2.04 | 1.73 | 0.37 |
| 800–1,000 | 6.41 | 2.64 | 2.01 | 0.45 |
| 1,000–1,200 | 7.27 | 3.00 | 2.88 | 0.56 |
| >1,200 | 7.90 | 3.84 | 3.31 | 0.67 |

| Table 3 | Average annual phosphorus exports as a function of precipitation and land use |
|-----------------|--------|--------|--------|--------|
| Precipitation (mm) | Dry field (kg/yr/ha) | Paddy field (kg/yr/ha) | Town (kg/yr/ha) | Woodland (kg/yr/ha) |
| 200–400 | 0.25 | 0.07 | 0.06 | 0.04 |
| 400–600 | 0.40 | 0.12 | 0.11 | 0.06 |
| 600–800 | 0.51 | 0.18 | 0.14 | 0.07 |
| 800–1,000 | 0.72 | 0.24 | 0.22 | 0.09 |
| 1,000–1,200 | 0.97 | 0.30 | 0.25 | 0.11 |
| >1,200 | 1.08 | 0.36 | 0.32 | 0.15 |

Figure 9 | Average annual nitrogen and phosphorus as a function of precipitation and land use.
the source area of the Liao River for the years 2005 to 2020 inclusive. Shuangliao station, Changchun station and Liaoyuan station, respectively, represent the lowest, average and highest annual average precipitation at a larger scale. The predicted monthly total precipitation for these three meteorological stations is shown in Figure 10.

According to WA-ANN model results, trends of precipitation change in each weather station are cyclical, and precipitation is mainly concentrated from June to September. It was found that the annual average precipitation in 2016–2020 years is more than the annual average precipitation in 2005–2015 years at eight rainfall gauges. In general, the future precipitation presented an increasing tendency, though there were variations in different periods.

The SWAT model was run to estimate monthly total flow and nonpoint source pollutant exports within the source area of the Liao River for the years 2005 to 2020 inclusive. The predicted monthly total flow is plotted in Figure 11 for the period 2005–2020. It can be seen that the variation of flow is consistent with precipitation, caused by the hydrological process of rainfall runoff generation and confluence.

The SWAT model was also able to generate the spatial distribution of nonpoint source nutrients (nitrogen and phosphorus) exports in the source area of the Liao River. The spatial distributions of predicted nutrient and phosphorus exports for the year 2020 are plotted in Figure 12.

The spatial distribution of the total nitrogen (TN) and total phosphorus (TP) exports identifies which sub-catchments are the primary contributors to nonpoint source pollutant exports, including the east of the Dongliao River and Siping, Gongzhuling, and Lishu in the upper reaches.
of Zhaosutai River. These areas do not only receive high rainfall but are also densely populated and with high agricultural production from dry fields and paddy fields, which are large users of pesticides and chemical fertilizer, which exacerbes nutrient exports.

CONCLUSION

The source area of the Liao River is an important grain growing area in China which experiences serious problems with agricultural NPS, which is impacting the regional economy and society. In order to address the water quality issues, it is necessary to understand the spatial distribution of NPS in the Liao River source area. This issue has been investigated by coupling a WA-ANN precipitation model with a SWAT model to assess the export of nonpoint source pollutants from the Liao River source area.

WA-ANN models were used to forecast precipitation at eight rainfall gauges located at Changchun, Siping, Shuangliao, Liaoyuan, Quantai, Erlongshan Reservoir, Wangben and Lishu. The WA-ANN models were trained over a 10-year period (from January 2005 to December 2014) and tested over a 1-year period (January 2015 to December 2015) and were run to predict precipitation from January 2016 to December 2020. It was concluded that the WA-ANN models achieved a good level of agreement between observed and predicted monthly average daily precipitation at all eight stations.

A SWAT model was also assembled and run to estimate the runoff and nonpoint pollutant exports in the source area of the Liao River. The calibration period was 2006–2008, while the validation period was 2009–2010. Based on the adopted evaluation criteria, it was considered that the level of accuracy of the predicted monthly average daily flows at Quantai and Lishu stations was acceptable. The only sediment data and nonpoint source nutrient data recorded during the calibration and validation periods was collected at the Quantai station. It was concluded that the predicted results are satisfactory, given the level of accuracy.
of the predicted monthly total sediment discharge and non-point source nutrient (nitrogen and phosphorus) exports.

The WA-ANN models and the SWAT model were run to generate the spatial distribution of nonpoint source nutrient (nitrogen and phosphorus) exports in the source area of the Liao River. It was found that the SWAT model identified the sub-catchments which not only receive high rainfall but are also densely populated with high agricultural production from dry fields and paddy fields, which are large users of pesticides and chemical fertilizer, as the primary source areas for nutrient exports.

It is also concluded that the coupled WA-ANN models and the SWAT model provide a tool which will inform the identification of management practices to improve the water environment in the source area of the Liao River.

ACKNOWLEDGEMENTS

This research was financially supported by the Department of Science and Technology of Jilin Province and the National Key S&T Special Project of Control and Management of the Water Pollution under the agreement 2012ZX07202-00. The financial support of the Graduate Innovation Fund of Jilin University, Scientific Research Foundation for Returned Scholars of Jilin University and the Seed Science Foundation of the Ministry of Education of China are also greatly appreciated.

REFERENCES