Runoff and sediment yield simulation in a large basin using GIS and a distributed hydrological model

FANGHUA HAO, XUESONG ZHANG, HONGGUANG CHENG, CHANGMING LIU & ZHIFENG YANG
School of Environment, State Key Laboratory of Water Environment Simulation, Beijing Normal University, Beijing 100875, China
fanghua@bnu.edu.cn

Abstract A GIS-based distributed SWAT (Soil and Water Assessment Tool) model was used to simulate the runoff and sediment yield in the upper basin of the Luohe River, a tributary of the Yellow River. Firstly, the basic GIS database integrating DEM, soil, land-use map, climate, and land management data, is established. To simplify input data preparation and interpretation without compromising simulation accuracy, the paper delineates the basin into 63 sub-basins to perform the simulation. In the process of calibration, the automated digital filter technique is used to separate the surface runoff and base flow. The surface runoff, the base flow, the total runoff, and the sediment yield are then calibrated sequentially. The simulated results demonstrate that the GIS-based SWAT model could be successfully used to simulate long-term runoff and sediment yield in large river basins such as the Yellow River basin where soil erosion is a serious problem.

Key words GIS; large basin; runoff; sediment; SWAT; Yellow River

INTRODUCTION

The development of distributed hydrological models has been greatly improved with the application of GIS (Geographic Information System) technology. In China, while several studies on runoff and sediment yield simulation have been conducted with the combination of distributed hydrological model (e.g. ANSWERS2000) and GIS (e.g. Niu et al. 2001; Liu et al. 2003) in recent years, most of these investigations are mainly concentrated on small basins. Under such a background, the present research aims to explore the suitability of similar modelling techniques in large river basins.

In the Yellow River basin soil erosion is a serious problem, while runoff and sediment yield simulation has not been extensively studied on the basis of the distributed hydrological model. In this study, the Lushi basin, which is located upstream of the Lushi hydrological station in the Luohe River—the largest tributary of the Yellow River below Xiaolangdi dam—was selected as the study area. The level of soil erosion in the Lushi basin is moderate in the Yellow River basin—the rate of soil erosion is about 2000–4000 t km\(^{-2}\) year\(^{-1}\) and sediment runoff modulus reaches 770 t km\(^{-2}\) year\(^{-1}\) (Guo & Zheng, 1995).

SWAT (Soil and Water Assessment Tools), a distributed hydrological model, was selected to simulate long-term runoff and sediment yield in the study area with the support of GIS technology. The model has been used in several projects by the
USEPA, NOAA, NRCS and others to estimate the off-site impacts of climate and management on water use, non-point source loads, and has been extensively validated across the United States for stream flow and sediment yields (Arnold et al., 1998).

DATA AND METHODS

SWAT model description

SWAT is a hydrologic/water quality model developed by the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) (Santhi et al., 2001). It is a continuous time model that operates on a daily time step. SWAT uses a modified version of the SCS CN method for predicting surface runoff yield (USDA-SCS, 1972):

\[ Q = \frac{(R - 0.2S)^2}{(R + 0.8S)} \quad R > 0.2S \]
\[ Q = 0 \quad R \leq 0.2S \]  

where \( Q \) is the daily surface runoff (mm), \( R \) is the daily rainfall (mm), and \( S \) is a retention parameter. \( S \) varies among basins under various soil, land-use, management, and slope conditions, and over time responding to changes in soil water content. The parameter \( S \) is related to curve number (CN) by the SCS equation:

\[ S = 25.4 \left( \frac{1000}{CN} - 10 \right) \]

Erosion and sediment yield are estimated for each sub-basin with the Modified Universal Soil Loss Equation (MUSLE) (Williams, 1975):

\[ Y = 11.8(Vq_p)^{0.56}(K)(C)(PE)(LS) \]

where \( Y \) is the sediment yield from the sub-basin, \( V \) is the surface runoff column for the sub-basin in \( \text{m}^3 \), \( q_p \) is the peak flow rate for the sub-basin in \( \text{m}^3 \text{ s}^{-1} \), \( K \) is the soil erodibility factor, \( C \) is the crop management factor, \( PE \) is the erosion control practice factor, and \( LS \) is the slope length and steepness factor.

Channel routing consists of flood and sediment routing. The flood routing model uses a variable storage coefficient method developed by Williams (1969). Channel inputs include the reach length, channel slope, bankfull width and depth, channel side slope, flood plain slope, and Manning’s \( n \) for channel and floodplain. Flow rate and average velocity are calculated using Manning’s equation and travel time is computed by dividing channel length by velocity. Outflow from a channel is also adjusted for transmission losses, evaporation, diversions, and return flow. The channel sediment routing equation uses a modification of Bagnold’s sediment transport equation that estimates the transport concentration capacity as a function of velocity (Bagnold, 1977):

\[ CY_n = SPCON \times V^{SPEXP} \]
where, $CY_n$ is sediment transport concentration capacity in g m$^{-3}$; $SPCON$ is the concentration capacity in g m$^{-3}$ at a velocity of 1 m s$^{-1}$; $V$ is flow velocity in m s$^{-1}$; and $SPEXP$ is a constant in Bagnold’s equation. The SWAT model either deposits excess sediment or re-entains sediments through channel erosion depending on the sediment load entering the channel.

**Study area description**

Lushi basin (Fig. 1) with an area of 4623 km$^2$ is characterized by mountainous landscape. Qinling Mountain is located on the south of the basin, Huashan Mountain and Yaoshan Mountain on the north. This area belongs to the warm temperate climate zone, and the annual average precipitation is about 600–800 mm.

Land uses in this basin are mostly forest and cropland in the upper reaches while cropland and pasture are widely spread in the lower reaches. The major soil types are zonal from the low elevation to the high elevation area, changing from Calcic Cinnamon Soils (27%, the percentage of the type of soil in the studied basin), Typic Cinnamon Soils (34%), Typic Burozems (38%) to Clay Pan Yellow-brown Earths (1%).

**Development of the database for Lushi basin**

The basic database for the Lushi basin was established using ArcView GIS, which mainly includes topography, soil and land-use maps, as well as climate and land management data (Table 1). Initially, the basin was delineated into sub-basins using the digital elevation map. The delineated sub-basin map was then overlaid with land-use and soil maps. The SWAT model simulates under different land-use in each sub-basin. Winter wheat and summer maize in rotation was simulated on the cropland.

![Fig. 1 Map of the study area.](image_url)
Table 1 Data sources for the Lushi basin.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Source</th>
<th>Scale</th>
<th>Data Description/Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography</td>
<td>National Geomatics Center of China</td>
<td>1:250 000</td>
<td>Elevation, overland, and channel slopes, lengths, etc.</td>
</tr>
<tr>
<td>Soil</td>
<td>Institute of Soil Science, Chinese Academy of Sciences (CAS)</td>
<td>1:4 000 000</td>
<td>Soil classifications and physical properties like bulk density,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>texture, saturated conductivity, etc.</td>
</tr>
<tr>
<td>Land-use</td>
<td>Institute of Geographical Sciences and Natural Resources Research, CAS</td>
<td>1:1 000 000</td>
<td>Land-use classifications such as cropland, pasture, forest, etc.</td>
</tr>
<tr>
<td>Weather</td>
<td>Water Resources Conservancy Committee of the Yellow River basin</td>
<td>–</td>
<td>Daily precipitation, air temperature, relative humidity, solar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>radiation and wind speed, etc.</td>
</tr>
<tr>
<td>Land Management</td>
<td>On-site survey</td>
<td>–</td>
<td>Tillage, planting and harvesting dates for different crops.</td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion on the sub-basin number

Distributed hydrological models use spatially distributed inputs and parameters to predict basin response commonly by aggregating input data on the basis of grid or sub-basin. Naturally, model output could be affected by the spatial extent over which input data are aggregated to produce parameters. The impact of sub-basin scaling on a basin simulation is directly related to the sources of heterogeneity in the basin (Arnold et al., 1998), including channel network, topography, soil, land use, and climate conditions. While studies have shown that the prediction of runoff and sediment was affected by the sub-basin number (Mamillapalli et al., 1996; Bingner et al., 1997; FutzHugh & MacKay, 1999), no reasonable subdivision level has been proposed so far. In this study, it is assumed that, while holding the precipitation constant, the output of the model predictions will not change greatly once the number of the sub-basins exceeds a specific level; we also assume that such level of subdivision then well represent the heterogeneity of the basin except precipitation. To test this assumption, the Lushi basin was partitioned into seven different basin delineations (Fig. 2). The maximum number of sub-basin used is 127, because more detailed delineations were found to contain an increased percentage of spurious (small or highly elongated) sub-basins. The streamflow and sediment change against sub-basin number is plotted in Fig. 3. As shown by the figure, while we hold precipitation constant, when the sub-basin number exceeds 24 and 37, the changes of model (we used rainfall data from Sanyao station in the centre of the basin) outputs on sediment load and streamflow against the increase of the number of sub-basins, respectively, reduced to minimal. This result indicates that spatial variations of the ground conditions are fairly represented with the 37 sub-basin division scheme. The sub-basin number of the study area is however taken as 63 for the final simulation after considering other factors, including precipitation, municipal boundary and so on. It is worth noting that the spatial variability of precipitation affects the results of a distributed hydrological model (Lopes, 1996; Chaubey et al., 1999), thus capturing the spatial precipitation distribution is a critical concern in determining the level of sub-basin division. Nevertheless, it is also meaningful to explore the level of variations of the ground conditions and its relationship to the model outputs while holding the precipitation constant.
Runoff and sediment yield simulation using GIS and a distributed hydrological model

Fig. 2 Seven different sub-basin delineations of the study basin.

Fig. 3 Runoff and sediment yield change against sub-basin number.

Evaluation of model output

Mean, Relative error ($Re$), coefficient of determination ($R^2$), Nash-Sutcliffe efficiency ($E_{ns}$) (Nash & Sutcliffe, 1970) are used to evaluate model performance. The $R^2$ is an indicator of strength of relationship between the observed and simulated values. $E_{ns}$ indicates how well the plot of the observed value vs the simulated value fits the 1:1 line. If the $R^2$ and $E_{ns}$ values are less than or very close to zero, the model performance is considered “unacceptable or poor”. If the values are equal to one, then the model prediction is considered to be “perfect”.

Model calibration

SWAT is not a “parametric model” with a formal optimization procedure (as part of the calibration process) to fit any data. Instead, a few important variables that are not
well defined physically such as runoff curve number and Universal Soil Loss Equation's cover and management factor, or C factor, may be adjusted to provide a better fit. A two stage “Brute Force” optimization procedure is used to find the optimum parameter values (Allred & Haan, 1999). This “brute force” optimization procedure, despite of its lower computational efficiency than other methods, has the advantage of being insensitive to local minimums in the objective function.

The procedure for calibrating the SWAT model for flow and sediment is shown in Fig. 4 (Santhi et al., 2001). Originally, base flow was separated from surface flow for both observed and simulated streamflow using an automated digital filter technique (Arnold & Allen, 1999). Calibration parameters for various model outputs are constrained within the ranges shown in Table 2. Model outputs are calibrated to fall within a percentage of average measured values and then regression statistics ($R^2$ and $E_{ns}$) are evaluated for monthly data. If all parameters were pushed to the limit of their ranges for a model output (i.e. flow or sediment) and the calibration criteria were still not met, then calibration would be terminated for that output.

![Flowchart Diagram](image_url)

Fig. 4 Calibration procedure for flow and sediment in SWAT model.

$E_{ns}$: Nash Sutcliffe Efficiency

$R^2$: Coefficient of Determination
Table 2 Inputs used in model calibration.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Processes</th>
<th>Description</th>
<th>Range</th>
<th>Value/Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN2</td>
<td>Flow</td>
<td>Curve Number</td>
<td>±8</td>
<td>Pasture: +2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Forest: −4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cropland: +1</td>
</tr>
<tr>
<td>REVAPC</td>
<td>Flow</td>
<td>Ground water revap coefficient</td>
<td>0.00 to 1.00</td>
<td>0.10</td>
</tr>
<tr>
<td>ESCO</td>
<td>Flow</td>
<td>Soil Evaporation compensation factor</td>
<td>0.00 to 1.00</td>
<td>0.4</td>
</tr>
<tr>
<td>EPCO</td>
<td>Flow</td>
<td>Plant uptake compensation factor</td>
<td>0.00 to 1.00</td>
<td>0.2</td>
</tr>
<tr>
<td>SMFMN</td>
<td>Flow</td>
<td>Melt factor for snow on December 21</td>
<td>0 to 10</td>
<td>5.5</td>
</tr>
<tr>
<td>C Factor</td>
<td>Sediment</td>
<td>Cover or management factor</td>
<td>0.003 to 0.45</td>
<td>Pasture: 0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Forest: 0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cropland: 0.20</td>
</tr>
<tr>
<td>SPCON</td>
<td>Sediment</td>
<td>Linear factor for channel sediment routing</td>
<td>0.0001 to 0.01</td>
<td>0.0006</td>
</tr>
<tr>
<td>SPEXP</td>
<td>Sediment</td>
<td>Exponential factor for channel sediment routing</td>
<td>1.0 to 1.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Flow With the combination of the streamflow from 1992 to 1997, obtained from the Water Conservancy Committee of the Yellow River’s monitoring station (Lushi hydrological station), the SWAT model is calibrated (Fig. 4). The runoff curve number (CN2) is adjusted using surface runoff data to allow a range of ±8 from the tabulated curve numbers to reflect the impact of conservation tillage practices and soil residue cover conditions of the basin (Table 2). The initial area-weighted CN2 value of pasture, forest and cropland is 69.2, 59.3 and 67.1, respectively. For base flow, related model parameters such as re-evaporation coefficient (REVAPC) for groundwater that represents the water that moves from the shallow aquifer back to the soil profile/root zone and plant uptake from deep roots, soil evaporation compensation factor (ESCO), and plant evaporation compensation factor (EPCO) are adjusted from the initial estimates to match the simulated and observed base flow (Table 2). Finally, in order to match the streamflow, minimum melt factor for snow (SMFMN) is adjusted for snow-melt periods. The simulation started from 1991 to reduce errors in initial estimates of variables such as the soil water content and surface residue at the beginning of 1992.

Sediment The cover, or C factor, of the Universal Soil Loss Equation is adjusted to match observed and simulated sediment loads (Fig. 4). The C factor is adjusted to better represent the surface (Table 2). Channel sediment routing variables such as the linear factor (SPCON) and the exponential factor (SPEXP) for calculating the maximum amount of sediment resentered during channel sediment routing are also adjusted (Table 2) in the process of sediment calibration. These two variables are adjusted to represent the cohesive nature of the channels.

Model validation

In the validation process, the model is operated with parameters obtained in the process of calibration without any change and the results are compared with the remaining observational data (from January 1998 to December 1999) to evaluate the model performance. The same statistical measures are used to assess the model performance.
RESULTS AND DISCUSSION

Calibration

Flow The measured and simulated monthly flow at Lushi hydrological station match well (Fig. 5(a)). According to the filter technique, the base flow accounts for 30% of the observed flow, while it is 26% for simulated flow. Means of the observed and simulated streamflow are within a difference of 15% (Table 3). Further agreement between observed and simulated flows are shown by the $R^2$ and $E_{ns}$, both are larger than 0.8 (Table 3). These results show that the hydrological processes in SWAT are simulated realistically in the study area.

Sediment The temporal variations of sediment load at Lushi station are represented in Fig. 5(b). Means of observed and simulated sediment are within a difference of 20% (Table 3). The values of $R^2$ and $E_{ns}$ are both 0.70 (Table 3), which indicates that the simulated sediment is close to the observed sediment and this model is able to predict sediment loads well.

Validation

Flow The observed and simulated flows at Lushi hydrological station matched well (Fig. 6(a)). The base flow accounts for 28% of the measured flow, and 26% for simulated one. $Re$ is 14.6%, $R^2$ and $E_{ns}$ are all greater than 0.80. The model

<table>
<thead>
<tr>
<th>Variable (units)</th>
<th>Annual mean observed</th>
<th>Annual mean simulated</th>
<th>$Re$</th>
<th>$R^2$</th>
<th>$E_{ns}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow volume ($10^8$ m$^3$)</td>
<td>4.15</td>
<td>3.88</td>
<td>-6.5%</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>Sediment ($10^4$ t)</td>
<td>96.6</td>
<td>106.44</td>
<td>10.2%</td>
<td>0.70</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Fig. 5 Observed and simulated monthly flow, and sediment loads during calibration period.

Table 3 Calibration results at Lushi hydrological station for the period from 1992 to 1997.
Runoff and sediment yield simulation using GIS and a distributed hydrological model

![Figure 6](image_url)

**Fig. 6** Observed and simulated monthly flow, and sediment loads during validation period.

<table>
<thead>
<tr>
<th>Variable (units)</th>
<th>Annual mean</th>
<th>Re</th>
<th>$R^2$</th>
<th>$E_{ns}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow Volume ($10^6$ m$^3$)</td>
<td>4.87</td>
<td>5.58</td>
<td>14.6%</td>
<td>0.84</td>
</tr>
<tr>
<td>Sediment ($10^4$ t)</td>
<td>189.24</td>
<td>169.92</td>
<td>-15.5%</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 4 Validation results at Lushi hydrological station for the period from 1998 to 1999.

overestimated the flow in some months such as September 1998, from April to August 1999, and slightly underestimated in May, August, and December 1998 (Fig. 6(a)). The difference might result from the spatial variability of precipitation. However, the prediction statistics are acceptable (Table 4).

**Sediment** The observed and simulated sediment load match well, although in August 1998 the sediment is underestimated and in May 1998 and September 1999 the sediment is overestimated (Fig. 6(b)). The values of $R^2$ and $E_{ns}$ are both above 0.9, which indicate that the model is able to predict sediment reasonably. The reason for high values of $R^2$ and $E_{ns}$ may be that the sediment yield in 1998 is much greater than the sediment yield in 1999. As the “goodness-of-fit” of observed and simulated data in 1998 is good ($R^2$ and $E_{ns}$ are 0.989 and 0.944, respectively), the results are acceptable even though the results doesn’t match well in 1999 ($R^2$ and $E_{ns}$ are 0.50 and 0.51, respectively). The results to some extent indicate that SWAT model is suitable for high flow yeasr (the precipitation in 1998 is at 10% of occurrence frequency) than low flow years (the precipitation in 1999 is at 75% of occurrence frequency).

**CONCLUSIONS**

The validation results demonstrate that the SWAT model based on GIS can be successfully used to simulate long-term runoff and sediment yield in a large basin in Yellow River basin: Relative error (Re) is within 20%, determination coefficient ($R^2$) and Nash-Sutcliffe efficiency ($E_{ns}$) are all above 0.70 during calibration and validation periods. It should be noted that SWAT model is more suitable for high flow years than
low flow years. With its spatial analytical capability, the SWAT model based on GIS could be used as a useful tool for water resources and soil erosion simulation in the Yellow River basin.

Acknowledgements The study was supported by the Major State Basic Research Development Programme, “Evolving Law and Maintaining Mechanism of Renewable Capacity of Water Resources in the Yellow River Basin” (no. G1999043605). The authors thank Prof. Zongxue Xu at the Beijing Normal University, P.R. China for providing detailed advice that contribute to improvements and refinement in the paper.

REFERENCES

Williams, J. R. (1969) Flood routing with variable travel time or variable storage coefficients. Trans. ASAE 12(1), 100–103.