

# Variational-Based Data Assimilation to Simulate Sediment Concentration in the Lower Yellow River, China

Hong Wei Fang, M.ASCE<sup>1</sup>; Rui Xun Lai<sup>2</sup>; Bin Liang Lin<sup>3</sup>; Xing Ya Xu<sup>4</sup>; Fang Xiu Zhang<sup>5</sup>; and Yue Feng Zhang<sup>6</sup>

**Abstract:** The heavy sediment load of the Yellow River makes it difficult to simulate sediment concentration using classic numerical models. In this paper, on the basis of the classic one-dimensional numerical model of open channel flow, a variational-based data assimilation method is introduced to improve the simulation accuracy of sediment concentration and to estimate parameters in sediment carrying capacity. In this method, a cost function is introduced first to determine the difference between the sediment concentration distributions and available field observations. A one-dimensional suspended sediment transport equation, assumed as a constraint, is integrated into the cost function. An adjoint equation of the data assimilation system is used to solve the minimum problem of the cost function. Field data observed from the Yellow River in 2013 are used to test the proposed method. When running the numerical model with the data assimilation method, errors between the calculations and the observations are analyzed. Results show that (1) the data assimilation system can improve the prediction accuracy of suspended sediment concentration; (2) the variational inverse data assimilation is an effective way to estimate the model parameters, which are poorly known in previous research; and (3) although the available observations are limited to two cross sections located in the central portion of the study reach, the variational-based data assimilation system has a positive effect on the simulated results in the portion of the model domain in which no observations are available. DOI: 10.1061/(ASCE)HE.1943-5584.0001344. This work is made available under the terms of the Creative Commons Attribution 4.0 International license, <http://creativecommons.org/licenses/by/4.0/>.

**Author keywords:** Variational data assimilation; Yellow River; Suspended sediment concentration; Adjoint equation.

## Introduction

The Yellow River is located in the middle of China, with a drainage area of 795,000 km<sup>2</sup> and a length of 5,464 km (Fig. 1). The Yellow River is well known for its heavy sediment load and frequent shifts in its course. The long-term mean annual sediment concentration is 35 kg/m<sup>3</sup>, and hyperconcentrated floods with sediment concentrations of more than 100 kg/m<sup>3</sup> occur quite often (Wu et al. 2008b). Because of siltation, the lower reach of the Yellow River, downstream the Xiaolangdi Reservoir, has become a suspended river (hanging river) whose bed is higher than the surrounding land surface, in some locations by more than 10 m (Wang et al. 2005).

The heavy sediment load of the Yellow River always results in drastic fluvial processes and avulsion has become a major threat to human life and property. Moreover, climate change and human activities along the river make the sediment transport process more complicated (Lu et al. 2013; Samaras and Koutitas 2014). To ensure the safety of levees, an accurate prediction of sediment concentration and water level becomes extremely important.

Numerical models, which are based on the theory of nonequilibrium sediment transport and address the suspended sediment exchanges in the near-bed layer, have considerable applications in practical river dynamics (e.g., Celik and Rodi 1988; Fang et al. 2008; van Rijn 1986). In applying these models, some key parameters, including the sediment carrying capacity, recovery coefficient, and friction factor, should be identified properly. For a specific water-sediment condition in a flood event, these parameters are often determined on the basis of the data acquired from laboratory experiments or field measurements (Fang and Wang 2000; Hu et al. 2014; Tsai et al. 2014). Some of the numerical models can be used to compute the transport of hyperconcentration sediment occurred in the Yellow River (Fang et al. 2008).

These numerical models are widely used in the pre-design phase for engineering projects. However, there are two fundamental challenges when using numerical models for real-time prediction. One challenge is that the simulated results of classic numerical model usually disagree with the measurements. The errors between the simulations and measurements may be attributed to the imperfection of the model structure and the difficulty in parameter calibration. Moreover, the heavy sediment load in the lower Yellow River can also affect the accuracy of water level and bed deformation predictions. Another challenge is the identification of the parameters in the numerical model. For a natural river, it is difficult to estimate these parameters using existing formulas, as the river characteristic changes with the variation of incoming water flow and sediment concentration from upstream catchments. Consequently,

<sup>1</sup>Professor, Dept. of Hydraulic Engineering, State Key Laboratory of Hydro-Science and Engineering, Tsinghua Univ., Beijing 100084, China. E-mail: fanghw@mail.tsinghua.edu.cn

<sup>2</sup>Senior Engineer, Yellow River Institute of Hydraulic Research, Zhengzhou 450003, China (corresponding author). E-mail: lairuxun@163.com

<sup>3</sup>Professor, Dept. of Hydraulic Engineering, State Key Laboratory of Hydro-Science and Engineering, Tsinghua Univ., Beijing 100084, China; Hydro-Environmental Research Centre, Cardiff Univ., U.K. E-mail: linbl@mail.tsinghua.edu.cn

<sup>4</sup>Ph.D. Research Fellow, Dept. of Hydraulic Engineering, State Key Laboratory of Hydro-Science and Engineering, Tsinghua Univ., Beijing 100084, China. E-mail: xu-xy11@mails.tsinghua.edu.cn

<sup>5</sup>Senior Engineer, Yellow River Institute of Hydraulic Research, Zhengzhou 450003, China. E-mail: 253545768@qq.com

<sup>6</sup>Ph.D. Research Fellow, Dept. of Hydraulic Engineering, State Key Laboratory of Hydro-Science and Engineering, Tsinghua Univ., Beijing 100084, China. E-mail: 81848340@qq.com

Note. This manuscript was submitted on May 11, 2015; approved on November 7, 2015; published online on February 24, 2016. Discussion period open until July 24, 2016; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Hydrologic Engineering*, © ASCE, ISSN 1084-0699.

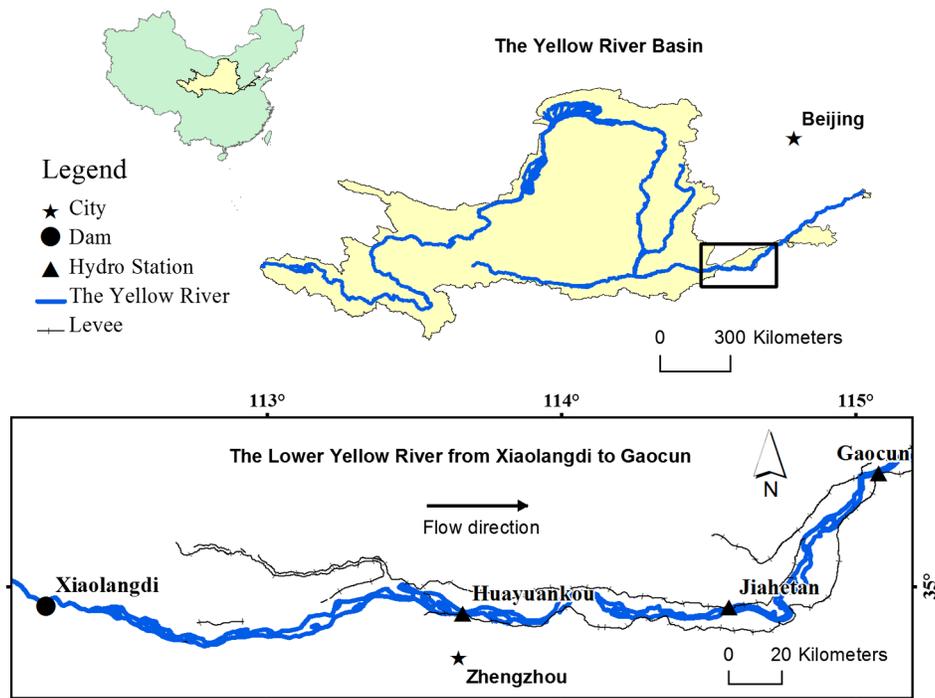


Fig. 1. Lower Yellow River and study area

empirical coefficients and formulas have been adopted and, sometimes, such numerical models fail in practical applications.

Three approaches can be employed to improve the numerical model accuracy. The first approach is to use a series of optimal methods, such as artificial intelligence and neural networks (e.g., Cheng et al. 2005; Chen and Chau 2006; Taormina and Chau 2015; Wu et al. 2009). The second approach is to build up coupled numerical models, in which the flow, sediment transport and morphological evolution processes are strongly coupled with one another (e.g., Cao et al. 2004). The third approach is the data assimilation method, which originates from weather prediction. Compared with other optimization algorithms, the advantage of the data assimilation method is that real-time observation data can be included into the numerical model simulation procedure. Consequently, the simulated variables will be closer to the observations thus improve the predictions in the next time step.

Nowadays, monitoring technology for suspended sediment concentration develops fast and real-time field data can be obtained and rapidly transferred to data center by the internet (Haimann et al. 2014). Therefore, the data can be integrated into the model to improve the accuracy of the prediction, which is the basic principle of data assimilation. The ensemble Kalman filter and variational-based methods are among the most popular data assimilation algorithms. The ensemble Kalman filter is an improved edition of the Kalman filter, which uses Monte Carlo theory to estimate the model error (Evensen 2009). For decades, the ensemble Kalman filter has received considerable attention over a wide range of subjects including ocean circulation, numerical weather forecasting, and hydrology (e.g., Moradkhani et al. 2005; Lai et al. 2013; Lewis et al. 2006). The variational-based algorithm implements the minimum distance between the observations and predictions, while, at the same time, taking an explicit dynamic system as a constraint (Le Dimet and Talagrand 1986).

The variational-based data assimilation method has some advantages comparing with the ensemble Kalman filter. The variational-based data assimilation implements the minimum over a recent time

period whereas the ensemble Kalman filter enhances the prediction using the observation in the latest time step. As a result, the variational-based data assimilation can improve the prediction over the recent time period. This advantage is more useful when the observations are rare. In contrast, the control equation describing the sediment transport can be assumed as a constraint, which indicates that the variational data assimilation considers more physical information than the ensemble Kalman filter. Thus, the obvious theoretical advantage of variational data assimilation is that it can provide consistency between the forecasts and the dynamics system (Kalnay 2003; Le Dimet and Talagrand 1986; Zhang and Zhang 2012).

In the past 10 years, several cases that integrate data assimilation with a sediment transport model have been published (Bélanger and Vincent 2005; Stroud et al. 2009; Thornhill et al. 2012). These cases use observations to enhance the prediction, but none of them are applied to solve the problem with heavy sediment load.

To improve the prediction of heavy sediment load in the lower Yellow River, a one-dimensional numerical model integrated with a variational-based data assimilation method is used to improve the simulated sediment concentration and to estimate coefficients of sediment carrying capacity. In the data assimilation system, a cost function which describes the difference between the observation and the simulation is given first. The adjoint equation and parameters' gradients are then applied to the suspended sediment concentration. The results and capability of the assimilation system are then demonstrated.

## Governing Equations and Discretization

The governing equations of the one-dimensional open channel flows considering the lateral inflow are the de Saint Venant equations

$$B \frac{\partial Z}{\partial t} + \frac{\partial Q}{\partial x} = q_l \quad (1)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left( \frac{Q^2}{A} \right) + gA \frac{\partial Z}{\partial x} + g \frac{Q|Q|}{C^2 AR} = 0 \quad (2)$$

where  $x$  = longitudinal direction along the channel thalweg;  $t$  = time;  $A$  = flow area;  $B$  = width of the cross section;  $Q$  = total volume discharge;  $Z$  = water level;  $q_l$  = lateral inflow discharge per unit channel length;  $g$  = gravitational acceleration;  $R$  = hydraulic radius; and  $C$  = Chézy coefficient.

Traditionally, the sediment transport is divided into a suspended load and a bed load. In the lower reach of the Yellow River, most sediment, with fine-grained size, is transported as the suspended load (Wu et al. 2008b). The bed load is only a small portion of the total load, varying from 0.3 to 0.7% on average at all the hydrologic stations along the lower reaches of the river (Wang et al. 2005). Therefore, no bed load transport is considered in this study. Han (1980) introduced the governing equation for the nonequilibrium transport of uniform sediment and it is widely used for practical applications in the Yellow River

$$\frac{\partial(AS)}{\partial t} + \frac{\partial(QS)}{\partial x} + \alpha\omega B(S - S_*) = S_l \quad (3)$$

where  $S$  = average sediment concentration over a cross section;  $S_*$  = sediment carrying capacity in  $\text{kg/m}^3$ ;  $S_l$  = suspended sediment concentration in the lateral outflow per unit channel length;  $\alpha$  = recovery coefficient; and  $\omega$  = average sediment fall velocity, which is calculated using the Stokes formula.

The sediment carrying capacity represents the maximum amount of sediment that can pass through a given river reach under a certain flow condition (Ch'ien and Wan 1999). If the input sediment concentration is higher than the sediment carrying capacity, the flow is under the condition of extrasaturation and deposition occurs. Otherwise, the flow is of insufficient saturation and erosion occurs (Fang and Wang 2000). For decades, several empirical or semi-empirical formulas, such as Engelund and Hansen (1967), Ackers and White (1973), Yang (1996), van Rijn (2007a, b) and Zhang (1961), have been introduced for practical engineering projects to describe the sediment carrying capacity. Wu et al. (2008a) computed the sediment transport in the Yellow River using the previous formulas, and the best predictions were obtained by Zhang's method and Yang's method because of the highly concentrated and fine-grained sediment. Actually, in China, Zhang's formula has been widely used for practical application, especially in the hyper-concentrated flow. Zhang's formula is on the basis of the energy balance theory in which the amount of energy supplied by the fluid equals frictional energy losses that required to keep sediment in suspension. This equation can be expressed as

$$S_* = k \left( \frac{V^3}{gR\omega} \right)^m \quad (4)$$

where  $k$  and  $m$  = coefficients;  $V$  = average cross sectional velocity. The concepts of sediment carrying capacity,  $S_*$ , and recovery coefficient are widely used in Chinese numerical simulations for nonequilibrium suspended sediment transport, which are reviewed by Fang and Wang (2000) in detail.

The bed deformation caused by the nonequilibrium sediment transport can be written as

$$\rho' \frac{\partial y_0}{\partial t} = \alpha\omega(S - S_*) \quad (5)$$

where  $y_0$  = riverbed deformation caused by erosion and siltation; and  $\rho'$  = dry bulk density of bed materials.

There are several finite difference schemes for the discretization of the governing equations [Eqs. (1) and (2)]. These schemes can be classified as explicit schemes, such as those in leap-frog (Liggett and Cunge 1975), and implicit schemes, such as Abbott and Ionescu (1967) and Preissmann (1961) schemes. In the explicit scheme, the equations are arranged to solve for one point at a time, while a group of advance points are solved simultaneously for the implicit scheme. The leap-frog scheme uses centered differences in both distance and time. In the Abbott scheme, the distinction is made between the storage width in the continuity equation and the so-called computational width in the dynamic equation. Among these discretization schemes, the Preissmann implicit four-point finite difference scheme is one of the widely used schemes. Instead of evaluating space derivatives at one of the grid points as the explicit schemes, the Preissmann method evaluates these derivatives in the middle of the two consecutive grid points, with the advantages of the numerical stability, the computation of discontinuities, and the calculation of the boundary conditions and the variable spatial intervals (Chanson 2004). In the Preissmann method, two commonly used assumptions are made during the process of iteration:  $A_j^{n+1} = A_j^* + \Delta A_j$  and  $Q_j^{n+1} = Q_j^* + \Delta Q_j$ . In the assumptions, the symbol \* denotes variable values at the last iteration step whereas  $\Delta A$  and  $\Delta Q$  are the increments of flow area and discharge, respectively. Substituting these two relations into the discretized Eqs. (1) and (2) to linearize the nonlinear terms, one can obtain the following iteration relations:

$$a_{1j}\Delta Q_j + b_{1j}\Delta Z_j + c_{1j}\Delta Q_{j+1} + d_{1j}\Delta Z_{j+1} = e_{1j} \quad (6)$$

$$a_{2j}\Delta Q_j + b_{2j}\Delta Z_j + c_{2j}\Delta Q_{j+1} + d_{2j}\Delta Z_{j+1} = e_{2j} \quad (7)$$

where  $j$  = number of the current cross section;  $\Delta Z$  = increment of water level;  $a_{1j} \sim e_{1j}$  and  $a_{2j} \sim e_{2j}$  = coefficients related to hydraulic parameters, geometry, and time (Cunge et al. 1980). Then the variables in Eqs. (6) and (7), i.e., water level and discharge, can be mathematically described by a coefficient matrix. There are several methods including the Gauss-Seidel, successive over relaxation method (SOR), and the double sweep method which can be used to solve the matrix. As a point ( $j$ ) is linked only to the adjacent points ( $j - 1$  and  $j + 1$ ), the linear system of equations is described by a sparse and pentadiagonal coefficient matrix. The double sweep method, using the banded matrix structure, is an effective approach to calculate the linear system of equations (Cunge et al. 1980). Here the pentadiagonal matrix described by Eqs. (6) and (7) is solved by the double sweep algorithm to obtain  $\Delta Q$  and  $\Delta Z$  at each iteration step. The water level and discharge are then updated by the assumptions. The iteration is stopped when the  $\Delta Z$  and  $\Delta Q$  are both minimized to a small tolerance.

Using a finite difference scheme, the equations of sediment transport and bed deformation are discretized as

$$\frac{A_j^{t+1} S_j^{t+1} - A_j^t S_j^t}{\Delta t} + \frac{Q_j^{t+1} S_j^{t+1} - Q_{j-1}^{t+1} S_{j-1}^{t+1}}{\Delta x_{j-1}} + [\alpha\omega B(S - S_*)]_j^{t+1} = S_l^{t+1} \quad (8)$$

$$\Delta y_0 = \frac{\Delta t}{\rho'} [\alpha\omega(S - S_*)]_i^{t+1} \quad (9)$$

where  $\Delta y_0$  = thickness of cross averaged bed deformation.

The initial value of sediment concentration and bed deformation are calculated directly by Eqs. (8) and (9). Once the observed sediment concentration is available, the predicted sediment

concentration will be updated using variational assimilation method. Then, the bed deformation can be updated by

$$\Delta y_{0j}^+ = \frac{\Delta t}{\Delta x_j B \rho'} (Q_{j+1} \cdot S_{j+1}^+ - Q_j \cdot S_j^+) \quad (10)$$

where  $S^+$  and  $\Delta y_0^+$  = updated sediment concentration and bed deformation, respectively. This equation describes the balance of the rate of sediment transport between two cross sections.

The model used in this paper has been applied to fine suspended loads and to hyperconcentrated flows in the Yellow River (Fang et al. 2008; Lai et al. 2013), but with some simplifications in sediment carrying capacity.

## Variational-Based Data Assimilation for Sediment Concentration

The general idea of the variational-based data assimilation for suspended sediment concentration is as follows. Firstly, a cost function is introduced to measure the differences between the observations and the calculated values, which consists of a weighted sum of square of the differences between the observations and the calculated values over the entire temporal and spatial domains (Navon 1986). Secondly, a set of adjoint equations are derived to calculate the gradient of the parameters. The adjoint equation of a nonlinear system is firstly used in the area of meteorology and is considered to be a powerful tool for data assimilation, parameter estimation, and stability analysis (Errico 1997). Finally, the cost function and its gradient are used to find the optimal initial conditions. The improved initial conditions are then used to calculate variables in the next time step, which makes the predictions closer to the observations. The steepest descent algorithm is applied for the minimization of the cost function. Compared with other algorithms (e.g., quasi-Newton methods), the steepest descent algorithm is both effective and easy to program.

### Cost Function

For a variational data assimilation system, a cost function, also called the objective function, is introduced first. The cost function is a measure of the magnitude of the discrepancy between observations and predictions (Talagrand 2010). The form of the cost function can be designed according to the needs of a specific variational problem. A general shape of the cost function  $J(S)$ , describing the differences between the observed and calculated sediment concentrations, can be written as

$$J(S) = \sum_{i=1}^l (HS - \tilde{S}_i)^T E^{-1} (HS - \tilde{S}_i) \quad (11)$$

where  $S$  = variable of sediment concentration whose size equals the number of model grid cells  $m$ ;  $\tilde{S}$  = observation whose size equals  $l$ ; and  $i = i$ th observation. Because the number of the observation is much less than the number of grid cell, the size of  $l$  is much less than  $m$ ;  $H$  = linearized operator that transforms model variables to observation space, whose size equals  $l$ . For example,  $H$  can be a simple linear interpolation from the model grid to the nearest location of the observations;  $E$  = observational error covariance whose size is  $l \times l$ . Choosing appropriate values for the covariance matrices is important for weighting correctly. The superscript  $T$  represents the transpose of the matrix whereas  $E^{-1}$  is the symbol for inverse.

The observation error covariance matrix  $E$  is simplified to have a diagonal form, assuming that there is no cross-correlation between the observation errors. According to the observation standard

for suspended sediment in China (Ministry of Water Resources of the PRC 1992), the error in the observed sediment concentration is assumed as following a normal distribution whose mean is 0 and variance is 2.1% of the observed value.

### Adjoint Equation and Gradients

Adjoint method, which integrates the model equations as a constraint, has been found to be an efficient way to obtain the gradients of the cost function with respect to the variables (Yu and O'Brien 1991; Schlitzer 1993). In this case, the model equation is the sediment transport equation. By appending the equation to the cost function as dynamic constraint, a Lagrange function can be constructed to optimize the value of sediment concentration

$$L(\lambda, S, \alpha, k, m) = J(S) + \sum_{t,j} \lambda_j^t \left( \frac{Q_j^{t+1} S_j^{t+1} - Q_{j-1}^{t+1} S_{j-1}^{t+1}}{\Delta x_{j-1}} + \frac{A_j^{t+1} S_j^{t+1} - A_j^t S_j^t}{\Delta t} + \left\{ \alpha \omega B \left[ S - k \left( \frac{V^3}{gR\omega} \right)^m \right] \right\}_j^{t+1} - S_j^t \right) \quad (12)$$

where  $t$  = time;  $j$  = number of the current cross section;  $J(S)$  = cost function described by Eq. (11);  $\lambda$  = Lagrange multipliers, also called adjoint variables (Sanders and Katopodes 2000). The differential equation listed in curly bracket is the constraint condition expressed as Eq. (8).

The problem of minimizing the cost function  $J$  subject to the dynamic constraint in Eq. (12) is now transformed into a problem of minimizing the unconstrained Lagrange function,  $L$ . The sediment transport equation for the best fit solution requires that all first order partial derivative of the Lagrange function with respect to  $\lambda$ ,  $S$ , and the three parameters ( $\alpha$ ,  $k$ ,  $m$ ) are equal to zero

$$\frac{\partial L}{\partial \lambda} = 0 \quad (13)$$

$$\frac{\partial L}{\partial S} = 0 \quad (14)$$

$$\frac{\partial L}{\partial \alpha} = 0 \quad (15)$$

$$\frac{\partial L}{\partial k} = 0 \quad (16)$$

$$\frac{\partial L}{\partial m} = 0 \quad (17)$$

Eqs. (13)–(17) are Euler-Lagrange equations. Differentiating  $L$  in Eq. (13) with respect to  $\lambda$  yields the Eq. (3), whereas differentiating  $L$  in Eq. (14) with respect to  $S$  results in the adjoint equation

$$(S_j^t - \tilde{S}_j^t) + \frac{Q_j^t \lambda_j^t - Q_{j-1}^t \lambda_{j-1}^t}{\Delta x_{j-1}} + \frac{A_j^{t-1} \lambda_j^{t-1} - A_j^t \lambda_j^t}{\Delta t} + [\alpha \omega B \lambda_j^t]_j = 0 \quad (18)$$

Eq. (18) is expressed as a backward difference approximation because the Lagrange multipliers,  $\lambda$ , can only be solved by solving the equation backward in time. Using the computed Lagrange multipliers, the gradients of the cost function ( $J$ ) with respect to the three parameters ( $\alpha$ ,  $k$ ,  $m$ ) can be derived from Eq. (15) to Eq. (17) as follows:

$$\nabla_{\alpha} L = \lambda_j^i \left[ \omega B \left( S - k \left( \frac{V^3}{gR\omega} \right)^m \right) \right]_j^t \quad (19)$$

$$\nabla_k L = -\lambda_j^i \left[ \alpha \omega B \left( \frac{V^3}{gR\omega} \right)^m \right]_j^t \quad (20)$$

$$\nabla_m L = -\lambda_j^i \left[ \alpha \omega B k \left( \frac{V^3}{gR\omega} \right)^m \cdot \ln \left( \frac{V^3}{gR\omega} \right) \right]_j^t \quad (21)$$

The values of the three parameters ( $\alpha$ ,  $k$ ,  $m$ ) can be updated and expressed into a simple form

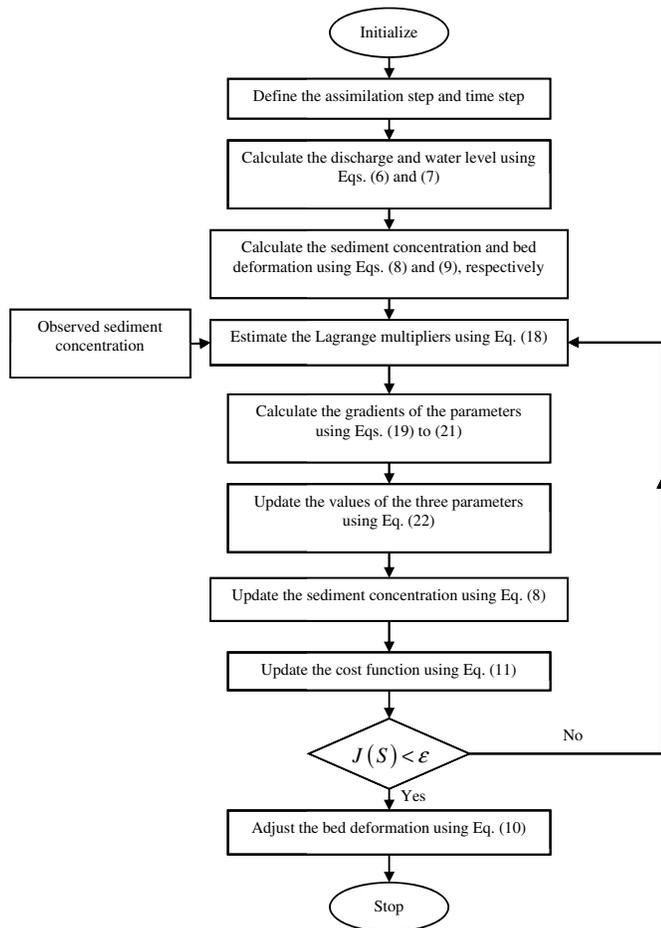
$$\zeta_{ni+1} = \zeta_{ni} - \beta \nabla \quad (22)$$

where  $\nabla$  = gradient of the three parameters;  $\zeta$  = value of the three parameters;  $\beta$  = iterative step; and  $ni$  = number of the iteration.

### Optimal Process

The flow chart of the data assimilation procedure used to optimize the sediment concentration is illustrated in Fig. 2 and can be summarized by the following steps:

1. Define the assimilation step and time step. The assimilation step defines how long a time of the observations will be used in the assimilation system whereas the time step is the discrete set of points covering the temporal field. The combinations of the two steps with different values can determine both the efficiency and accuracy of the data assimilation system.



**Fig. 2.** Flow chart of the data assimilation procedure for suspended sediment concentration

2. Calculate the discharge and water level using Eqs. (6) and (7).
3. Calculate the sediment concentration and bed deformation using Eqs. (8) and (9). The two variables, water level and discharge, will be assumed as initial values for the variational-based data assimilation system.
4. The observed values of sediment concentration are added into the adjoint equation to estimate the Lagrange multipliers. Because the number of observations is far less than the number of points on the computational grid, the calculations are interpolated into the points of observations.
5. Calculate the gradients of the three parameters:  $\alpha$ ,  $k$ , and  $m$ .
6. Update the values of the three parameters using Eq. (22). To find the steepest way to minimize the cost function, the iterative step in Eq. (22) can be varied in each iteration.
7. Calculate the sediment concentration again using the updated parameters.
8. Check whether the cost function satisfies with a convergence criterion  $\varepsilon$ . The symbol  $\varepsilon$  is the criterion specifying the value to determine whether the minimum cost function is reached. Normally, the minimum cannot be reached in the first iteration. Thus, it is necessary to return to Step 4 to continue the process and stop when the value of the cost function is less than  $\varepsilon$ .
9. Now, the optimal values of the sediment concentration are obtained and the bed form can be adjusted using Eq. (10).

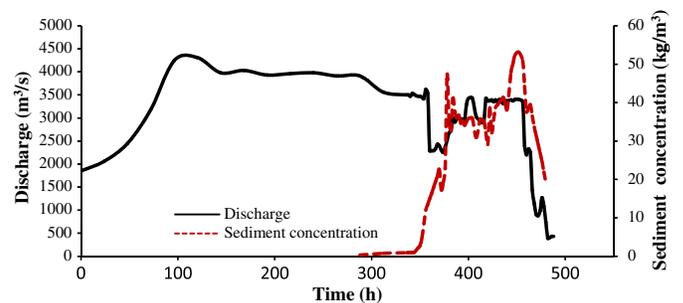
## Results

### Study Area and Boundary Condition

The study area is located in the lower Yellow River from Xiaolangdi to Gaocun, consisting of a distance of approximately 281 km (Fig. 1). There are four hydrological stations along the river: Xiaolangdi, Huayuankou, Jiahetan and Gaocun, with Xiaolangdi and Gaocun being set as the input and output boundaries, respectively.

The data assimilation scheme, previously described, was applied to investigate a flood event that occurred in 2013, from June 19 to July 11, with a total simulation time of 528 h. To make the simulation representing better to the true process of the natural flood, lateral inflow, sand mining, irrigation, evaporation, and infiltration were considered in the numerical model. During the simulation, the outflow for irrigation (388 m<sup>3</sup>/s) and the loss attributable to evaporation and infiltration (120 m<sup>3</sup>/s) were taken into account. The amount of sand loss attributable to mining for construction was set to 0.8% of the total incoming suspended sediment according to the Yellow River sediment bulletin published in 2013 (Yellow River Conservancy 2013).

Fig. 3 shows the input discharge and suspended sediment concentration. As given in the figure, the input discharge from



**Fig. 3.** Input boundary condition of discharge and suspended sediment concentration

**Table 1.** Initial Values of the Model Parameters

Parameter	Description	Unit	Initial value
$n$	Manning's roughness	$m^{-1/3} \times s$	0.008
$\omega$	Settling velocity	m/s	$8.7 \times 10^{-5}$
$\rho'$	Dry bulk density	$kg/m^3$	1,300
$S_l$	Amount of sediment mining	$kg/m^3$	0.8% of the calculated suspended sediment concentration
$\alpha$	Recovery coefficient	—	Erosion: 0.02, deposition: 0.005
$k$	Coefficient	—	0.4
$m$	Coefficient	—	0.6
$\beta$	Iterative step	—	0.002
$\Delta a$	Assimilation step	h	150
$\Delta t$	Time step	s	600

the Xiaolangdi station was around  $4,000 \text{ m}^3/\text{s}$  from the 100th hour to 320th and then dropped dramatically to  $2,000 \text{ m}^3/\text{s}$  in the 360th hour. From the 400th to 480th hour, the input discharge returned to  $3,500 \text{ m}^3/\text{s}$ . Because of the control of the Xiaolangdi dam, the suspended sediment concentration had a low value from the beginning to the 350th hour and then greatly increased from 30 to  $50 \text{ kg/m}^3$  at the 370th hour to the 470th hour.

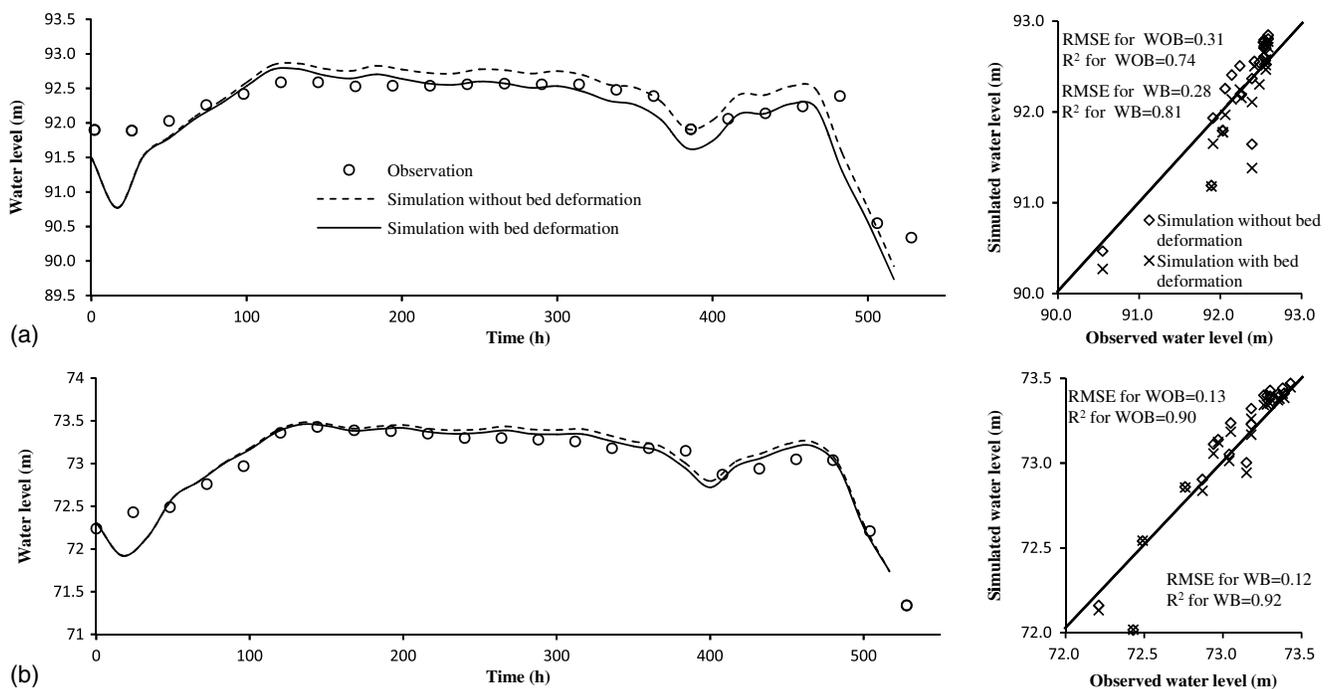
The initial values of parameters used in the numerical model are listed in Table 1. The Manning's roughness coefficient was  $0.008 \text{ m}^{-1/3} \cdot \text{s}$  when the flood was confined to the main channel. The recovery coefficient was  $0.02 \text{ m}^{-1/3} \cdot \text{s}$  when erosion happened, and it became  $0.005 \text{ m}^{-1/3} \cdot \text{s}$  when deposition happened. The initial values of the coefficients  $k$  and  $m$  were 0.4 and 0.6, respectively. The iterative step,  $\beta$ , was assumed as a constant value 0.002. The assimilation step was 150 h and the time step was 600 s. These initial values of the parameters, including Manning's roughness coefficient, recovery coefficient, and sediment carrying capacity, were calibrated using data obtained from historical floods and considered to be reasonable for use in the lower Yellow River (Lai et al. 2013, 2014).

The proposed data assimilation system was performed under the operation system of Linux Ubuntu 14.0. The programming language was GCC Fortran and the compiler was GCC. The computing costs, with and without data assimilation, were approximately 30 and 20 s, respectively. Approximately an extra 50% extra computing time was needed because of data assimilation. However, with regarding the improvement in model accuracy, the increased computing time was considered acceptable.

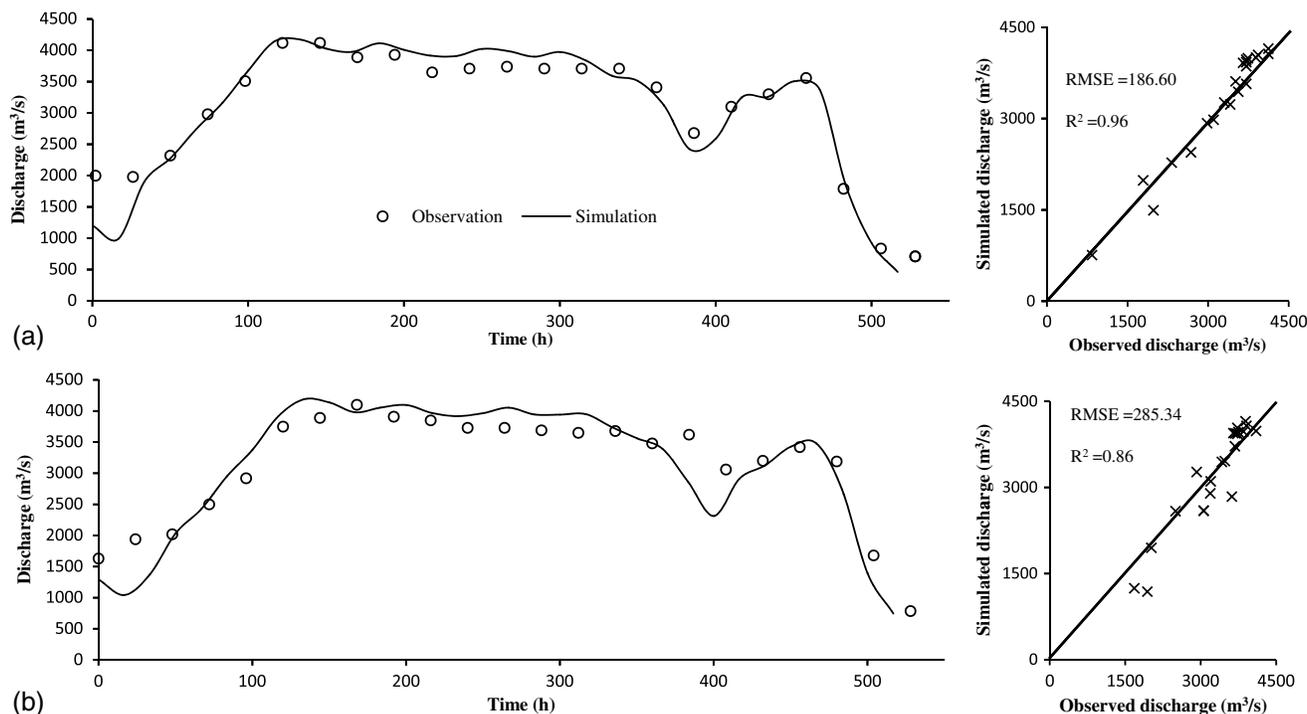
In this paper, the coefficient of determination ( $R^2$ ) and root mean square error (RMSE), which were critically reviewed by Legates and McCabe (1999), were used to evaluate the error of simulated results. These two evaluation criteria are commonly used in the area of both hydrology and hydraulics. Usually, the  $R^2$  parameter defines the relationship between the observations and the predictions, whereas the RMSE measures the residuals between those two sets of values.

### Simulated Water Level and Discharge

Fig. 4 shows the simulated water level with (represented by WB) and without (represented by WOB) the bed deformation calculation. The result demonstrates that the heavy sediment load in the lower Yellow



**Fig. 4.** Measured versus calculated water level for (a) Huayuankou; (b) Jiahetan stations. WB and WOB refer to the results with and without bed deformation, respectively



**Fig. 5.** Measured versus calculated discharge for (a) Huayuankou; (b) Jiahetan stations

River can substantially change the water level. When the bed deformation was considered, the predicted water level was closer to the observations. This improvement is also illustrated by the RMSEs and  $R^2$ s for the two stations. Specifically, at the Huayuankou station, the RMSE value decreased from 0.31 to 0.28 when the bed deformation was considered in the numerical model. At the Jiahetan station, in comparison with the Huayuankou station, the RMSE value reduced slightly from 0.13 to 0.12.

Fig. 5 shows the results of the measured and calculated discharge for the two stations. As is shown in the figure, the simulations produced a similar hydrograph to the observations but the errors were also obvious. The RMSE value for the Huayuankou station was 186.6 whereas for the Jiahetan station the RMSE increased to 285. For the two stations, the simulated values of discharge were generally higher than the observations, especially from the 200th hour to the 300th hour.

### Simulated Sediment Concentration and Bed Deformation

Fig. 6 shows the comparisons between the model simulation and field observed sediment concentrations, with and without data assimilation. Fig. 6(a) reveals two peak values of sediment concentration during the flood event in 2013. The first one appeared at the 410th hour with a value of 27.4 kg/m<sup>3</sup>, whereas the second one appeared at approximately the 482th hour with a value of 29.7 kg/m<sup>3</sup>. Although the numerical model applied without data assimilation could forecast the two peak values, the simulations were substantially higher than the observations. However, at the Jiahetan station, the simulated sediment concentrations were lower than the observations.

The errors between the numerical model and observations can be attributed to the problem of parameter estimation such that the coefficients of recovery and sediment-carrying capacity could not adjust to the change in the water-sediment relation. However, when data assimilation was applied, the simulated accuracy was

improved. For the Huayuankou station, the coefficient of determination increased from 0.97 to 0.98, although the RMSE decreased from 3.1 to 1.19. Similarly, the coefficient of determination at the Jiahetan station increased from 0.93 to 0.99 whereas the RMSE decreased obviously from 3.7 to 0.36.

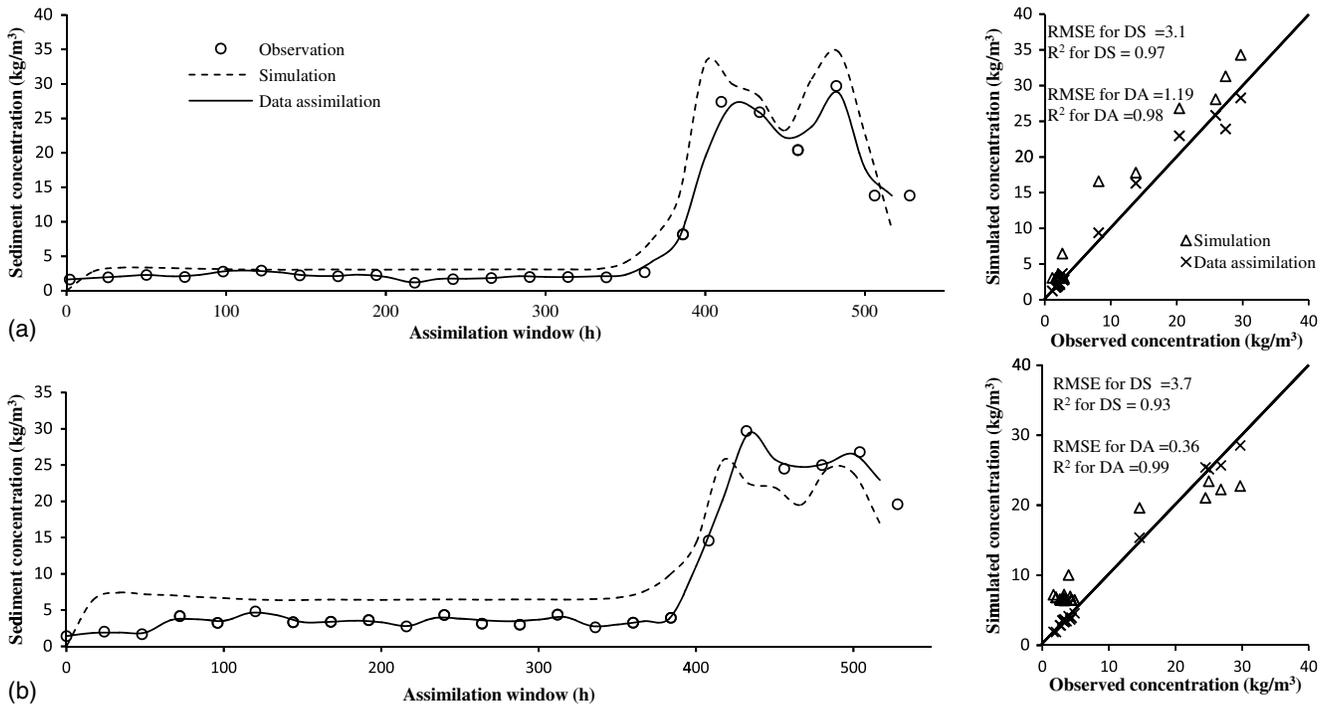
Fig. 7 shows the calculated and measured amount of erosion and deposition. The measured sedimentation was derived from the difference between the sediment loads at the stations along the river. A negative value refers to erosion whereas a positive value refers to deposition. During the flood event in 2013, the reach from Xiaolangdi to Huayuankou experienced deposition whereas from Huayuankou to Gaocun erosion. The simulated bed deformation was calculated using Eq. (9) and then adjusted by Eq. (10). It is concluded that the simulated erosion and deposition were consistent with the observations.

### Discussion

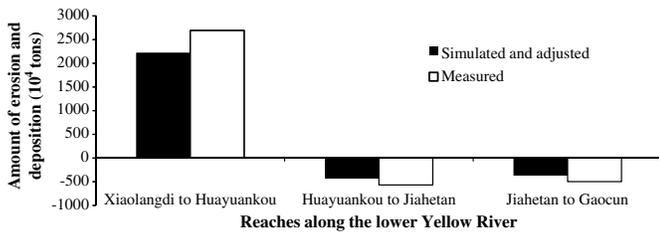
An important advantage of the variational data assimilation method is its ability of parameters estimation. In the area of sediment transport, some parameters, including sediment carrying capacity and recovery coefficient, are generally analyzed on the basis of the data from experimental statistics. However, the values of these parameters and the variation with water-sediment condition are poorly understood. In this section, the variation of these parameters with the changing water-sediment condition is discussed followed by an analysis of the effect of data assimilation system on the model domain in which no observation is available.

### Parameters of Sediment Carrying Capacity

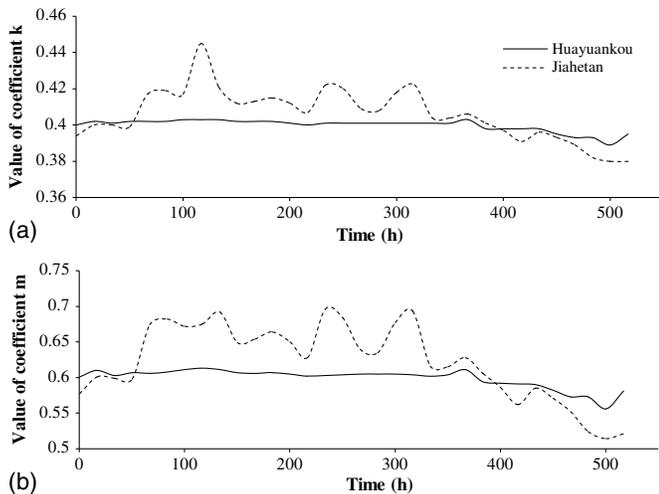
Coefficient  $k$  and exponent  $m$  are the important factors needed to determine the sediment carrying capacity. According to the research of Zhang (1961), coefficient  $k$  and exponent  $m$  are not constant values but have a relation to the value of the function



**Fig. 6.** Simulated sediment concentration with and without data assimilation compared with observations of (a) Huayuankou; (b) Jiahetan stations. DS refers to direct simulation whereas DA represents data assimilation



**Fig. 7.** Measured versus calculated amount of erosion and deposition



**Fig. 8.** Variation of the parameters: (a)  $k$ ; (b)  $m$  for Huayuankou and Jiahetan stations

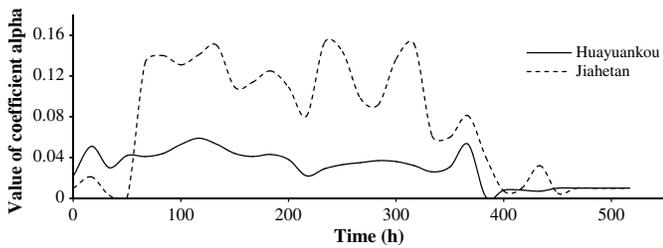
$V^3/gR\omega$ . Specifically, when the value of  $V^3/gR\omega$  increases, the value of the coefficient  $k$  will increase but the value of exponent  $m$  will decrease.

Fig. 8 shows the variation of the two parameters,  $k$  and  $m$ , in the equation for the sediment carrying capacity. For the Huayuankou station, the value of coefficient  $k$  was approximately 0.4 whereas the value of exponent  $m$  was around 0.6, with both being close to the initial values of the two parameters. However, the values of the two parameters for Jiahetan station changed frequently. At Jiahetan station, the value of  $k$  varied from 0.38 to 0.44 whereas  $m$  varied from 0.5 to 0.7. The values at both stations decreased at the 380th hour, which indicated a decrease of the sediment carrying capacity.

### Recovery Coefficient

The recovery coefficient  $\alpha$  is a synthetic empirical parameter, which exerts a great deal of influence on the rate of riverbed deformation. The larger the recovery coefficient is, the greater will be the rate of riverbed deformation, and vice versa (Fang and Wang 2000). The recovery coefficient is used to account for a variety of mechanisms that occur during the deposition and erosion processes. Han (1980) proposed that  $\alpha = 0.25$  or  $0.5$  for a river or reservoir, if deposition occurred, and  $\alpha = 1.0$  or more if erosion took place. However, according to the result of one-dimensional numerical modeling for the lower Yellow River, Wang et al. (2005) considered that the recovery coefficient was much smaller than 0.25, as proposed by Han (1980). Wang et al.'s (2005) result demonstrated that the recovery process in the Yellow River with its heavy sediment concentration was much slower than other rivers.

Fig. 9 shows the variation of the recovery coefficient at the Huayuankou and Jiahetan stations. For the Huayuankou station, the recovery coefficient stabilized near 0.04 from the beginning to the 380th hour. After that, the value decreased dramatically to 0.005, which indicates that the Huayuankou station was transformed from a situation of erosion to deposition. A similar



**Fig. 9.** Variation of the recovery coefficient,  $\alpha$ , at the Huayuankou and Jiahetan stations

phenomenon also happened at the Jiahetan station. During the first 380 h, the recovery coefficient at Jiahetan fluctuated from 0.04 to 0.16, but then declined rapidly.

**Effect of Data Assimilation System on the Model Domain with No Observations**

In the lower Yellow River, there are 121 measured cross sections from Xiaolangdi to Gaocun. However, observation data were only available at the two hydrological stations: Huayuankou and Jiahetan. Compared with the whole computational domain, the distributions of the observation data were non-uniform. Unfortunately, the data assimilation, according to the principle described in section “Variational-based Data Assimilation for Sediment Concentration,” can only work in the portions of the domain in which the observations are available. Therefore, it is not known to what extent the

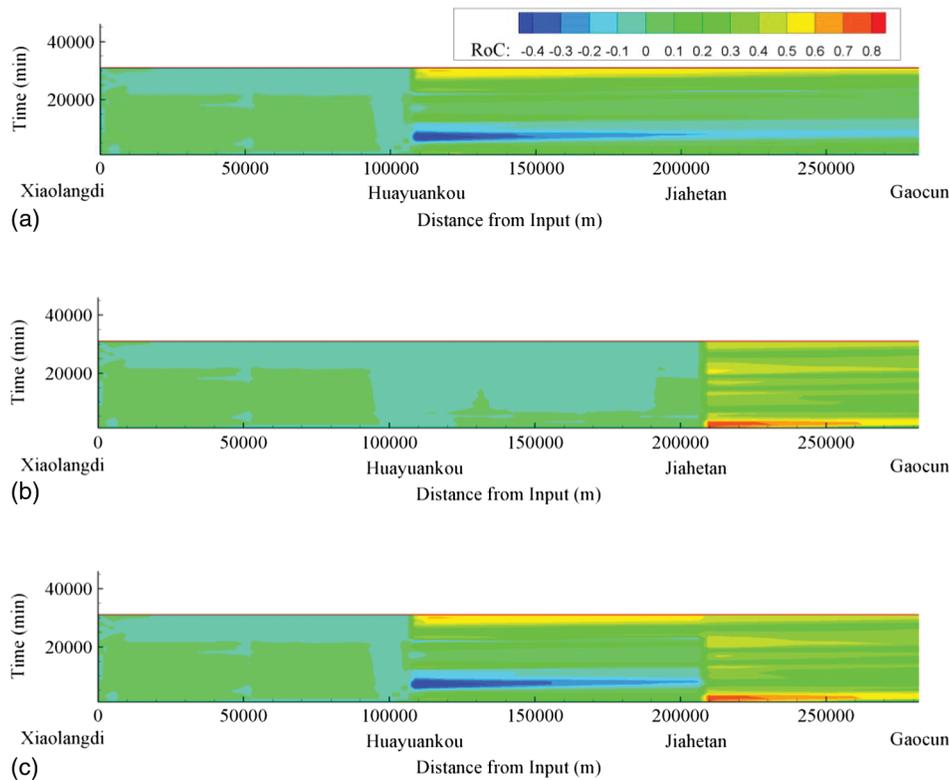
data assimilation system can affect the portion of model domain with no observations.

A scheme has been designed to answer this question. At first, the results of direct simulation without data assimilation were calculated as background data. Then the data assimilation system runs for three scenarios. Scenarios 1 and 2 represented the assimilation system runs considering the observations only from the Huayuankou and Jiahetan stations, respectively. In Scenario 3, the model run with the observation data from both the stations. The changes of the predicted sediment concentration are calculated using Eq. (23)

$$RoC = \frac{S^- - S^+}{S^-} \quad (23)$$

where  $RoC$  = rate of change; and  $S^+$  and  $S^-$  = simulated sediment concentration with and without data assimilation, respectively. The higher the absolute value of  $RoC$  is, the greater effect on the domain with no observations, and vice versa. A positive value of  $RoC$  indicates that the results of direct simulation are larger than the results of the data assimilation, and vice versa.

Fig. 10 illustrates the results produced with the data assimilation system and its effect on the model domain with no observations. In the figure, the horizontal axis represents the distance of cross sections from the flow input whereas the vertical axis represents the assimilation time. The units for the horizontal and the vertical axis are meters and minutes, respectively. The result from Scenario 1 shows that the location of the Huayuankou station is an obvious boundary to distinguish whether the influence of the data assimilation is great or small. Specifically, the data assimilation system has a distinct effect on the domain downstream of the Huayuankou



**Fig. 10.** (Color) Effects of the data assimilation system on the computational domain with no observations: (a) result of Scenario 1: the data assimilation system runs with the observations only from Huayuankou station; (b) result of Scenario 2: the data assimilation system runs with the observations only from Jiahetan station; (c) result of Scenario 3: the data assimilation system runs with the observations both from Huayuankou and Jiahetan stations. The horizontal axis represents the distance of cross sections from the flow input whereas the vertical axis represents the assimilation time. The units for the horizontal and the vertical axis are meters and minutes, respectively

station. However, for the domain upstream of the Huayuankou station, the influence is not noticeable. A similar phenomenon was found in Scenarios 2 and 3. The reason for this phenomenon can partly be attributed to the numerical equations which control the value of sediment concentration and the sediment transportation to the model domain with no observations.

## Conclusions

In this paper, a variational-based data assimilation system was presented and applied to simulate the sediment transport process in the lower Yellow River, China. In the data assimilation system, the variables of the water level and discharge were calculated directly whereas the suspended sediment concentration was optimized using variational-based data assimilation method. A form of cost function was used to describe the difference between the calculated and observed data. The constraints that are the equations of the physical model are combined into the cost function to construct a functional-linear function on vectors. The variable of sediment concentration and the parameters for sediment carrying capacity are represented by the gradients of the functional. The adjoint equations and the steepest descent algorithm were used to minimize the functional and to obtain the optimized variables.

The data assimilation system was applied to the lower Yellow River to reproduce the 2013 flood. It was found that the data assimilation system efficiently reduced the error in the suspended sediment concentration. Using the gradient of the functional, the parameters for sediment carrying capacity can be successfully estimated. The optimal values of the parameters were considered to be reasonable according to earlier research. Furthermore, the data assimilation system, under the control of the numerical model, can affect the model domain with no observations available.

In the future, a more stable and practical variational-based data assimilation system will be established. As the available data are mostly sparse and scattered in several hydrostations, some optimal methods should be applied to make the data assimilation system more stable. In addition, more historical flood data should be acquired to capture the trajectory of the parameters variation with the changing water-sediment condition. According to the results of the inverse parameter problem, the physical meanings of these parameter variations help one to better understand the identification of parameters. However, it is not enough to deeply understand the parameters' characteristics with only a collection of the parameters variation from a single river and single flood data. The authors expect to use more flood data to establish a function between these parameters and the variables (e.g., water level and discharge).

## Acknowledgments

This project was supported by the National Natural Science Foundation of China (Grant Nos. 51409113 and 51479081), the National Basic Research and Development Program of China (973 Program, Grant No. 2011CB403306), and the Ministry of Water Resources' Special Funds for Scientific Research on Public Causes (201301062).

## References

Abbott, M. B., and Ionescu, F. (1967). "On the numerical computation of nearly horizontal flows." *J. Hydraul. Res.*, 5(2), 97–117.  
Ackers, P., and White, W. R. (1973). "Sediment transport: New approach and analysis." *J. Hydraul. Div.*, 99(11), 2041–2060.

Bélanger, E., and Vincent, A. (2005). "Data assimilation (4D-VAR) to forecast flood in shallow-waters with sediment erosion." *J. Hydraul. Eng.*, 130(1–4), 114–125.  
Cao, Z., Pender, G., Wallis, S., and Carling, P. (2004). "Computational dam-break hydraulics over erodible sediment bed." *J. Hydraul. Eng.*, 130(7), 689–703.  
Celik, I., and Rodi, W. (1988). "Modeling suspended sediment transport in nonequilibrium situations." *J. Hydraul. Eng.*, 114(11), 1157–1191.  
Chanson, H. (2004). *Environmental hydraulics of open channel flows*, Elsevier Butterworth Heinemann, Amsterdam, Netherlands.  
Chen, W., and Chau, K. W. (2006). "Intelligent manipulation and calibration of parameters for hydrological models." *Int. J. Environ. Pollut.*, 28(3–4), 432–447.  
Cheng, C., Chau, K., Sun, Y., and Lin, J. (2005). "Long-term prediction of discharges in manwan reservoir using artificial neural network models." *Advances in Neural Networks—ISNN 2005*, J. Wang, X.-F. Liao, and Z. Yi, eds., Springer, Berlin, 1040–1045.  
Ch'ien, N., and Wan, C. (1999). *Mechanics of sediment transport*, ASCE, Reston, VA.  
Cunge, J. A. (1980). *Practical aspects of computational river hydraulics*, Pitman Advanced Publications Program, Boston.  
Engelund, F., and Hansen, E. (1967). *A monograph on sediment transport in alluvial streams*, Teknisk Forlag Technical Press, Copenhagen, Denmark.  
Errico, R. M. (1997). "What is an adjoint model?" *Bull. Am. Meteorol.*, 78(11), 2577–2591.  
Evensen, G. (2009). *Data assimilation: The ensemble Kalman filter*, Springer, Dordrecht, Holland.  
Fang, H., Chen, M., and Chen, Q. (2008). "One-dimensional numerical simulation of non-uniform sediment transport under unsteady flows." *Int. J. Sediment Res.*, 23(4), 316–328.  
Fang, H. W., and Wang, G. Q. (2000). "Three-dimensional mathematical model of suspended-sediment transport." *J. Hydraul. Eng.*, 126(8), 578–592.  
Haimann, M., Liedermann, M., Lalk, P., and Habersack, H. (2014). "An integrated suspended sediment transport monitoring and analysis concept." *Int. J. Sediment Res.*, 29(2), 135–148.  
Han, Q. W. (1980). *A study on the non-equilibrium transportation of suspended load*, Guanghai, Beijing, 793–802.  
Hu, P., Cao, Z., Pender, G., and Liu, H. (2014). "Numerical modelling of riverbed grain size stratigraphic evolution." *Int. J. Sediment Res.*, 29(3), 329–343.  
Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability*, Cambridge University Press, New York.  
Lai, R., Fang, H., He, G., Yu, X., Yang, M., and Wang, M. (2013). "Dual state-parameter optimal estimation of one-dimensional open channel model using ensemble Kalman filter." *J. Hydrodyn., Ser. B*, 25(4), 564–571.  
Lai, R. X., Fang, H. W., and Xu, X. Y. (2014). "Dynamic numerical model for the prediction of water and sediment transport." *Chin. J. Hydraul. Eng.*, 45(8), 930–937 (in Chinese).  
Le Dimet, F. X., and Talagrand, O. (1986). "Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects." *Tellus A*, 38A(2), 97–110.  
Legates, D. R., and McCabe, G. J. (1999). "Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation." *Water Resour. Res.*, 35(1), 233–241.  
Lewis, J. M., Lakshminarayanan, S., and Dhall, S. K. (2006). "Dynamic data assimilation: A least squares approach." *Encyclopedia of mathematics and its applications*, Cambridge University Press, Cambridge, U.K.  
Liggett, J. A., and Cunge, J. A. (1975). "Numerical methods of solution of the unsteady flow equations." *Unsteady flow in open channels*, Water Resources Publications, CO, 484.  
Lu, X. X., Ran, L. S., Liu, S., Jiang, T., Zhang, S. R., and Wang, J. J. (2013). "Sediment loads response to climate change: A preliminary study of eight large Chinese rivers." *Int. J. Sediment Res.*, 28(1), 1–14.  
Ministry of Water Resources of the PRC. (1992). "Specification for the measurement of suspended sediment load." *GB 50159-92*, China (in Chinese).

- Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). "Dual state-parameter estimation of hydrological models using ensemble Kalman filter." *Adv. Water Resour.*, 28(2), 135–147.
- Navon, I. M. (1986). "A review of variational and optimization methods in meteorology." *Variational methods in geosciences, developments in geomathematics*, Elsevier, Amsterdam, Netherlands, 29–34.
- Preissmann, A. (1961). "Propagation des intumescences dans les canaux et Les Rivieres." *1 Congres de l'Association Francaise de Calcule*, Grenoble, France.
- Samaras, A. G., and Koutitas, C. G. (2014). "Modeling the impact of climate change on sediment transport and morphology in coupled watershed-coast systems: A case study using an integrated approach." *Int. J. Sediment Res.*, 29(3), 304–315.
- Sanders, B. F., and Katopodes, N. D. (2000). "Adjoint sensitivity analysis for shallow-water wave control." *J. Eng. Mech.*, 10.1061/(ASCE)0733-9399(2000)126:9(909), 909–919.
- Schlitzer, R. (1993). "Determining the mean, large-scale circulation of the atlantic with the adjoint method." *J. Phys. Oceanogr.*, 23(9), 1935–1952.
- Stroud, J. R., Lesht, B. M., Schwab, D. J., Beletsky, D., and Stein, M. L. (2009). "Assimilation of satellite images into a sediment transport model of Lake Michigan: Satellite data assimilation in lake Michigan." *Water Resour. Res.*, 45(2), W02419.
- Talagrand, O. (2010). "Variational assimilation." *Data assimilation*, W. Lahoz, B. Khattatov, and R. Menard, eds., Springer, Berlin, 41–67.
- Taormina, R., and Chau, K. (2015). "Neural network river forecasting with multi-objective fully informed particle swarm optimization." *J. Hydroinf.*, 17(1), 99–113.
- Thornhill, G. D., Mason, D. C., Dance, S. L., Lawless, A. S., Nichols, N. K., and Forbes, H. R. (2012). "Integration of a 3D variational data assimilation scheme with a coastal area morphodynamic model of Morecambe Bay." *Coastal Eng.*, 69(11), 82–96.
- Tsai, C. W., Man, C., and Oh, J. (2014). "Stochastic particle based models for suspended particle movement in surface flows." *Int. J. Sediment Res.*, 29(2), 195–207.
- van Rijn, L. C. (1986). "Mathematical modeling of suspended sediment in nonuniform flows." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1986)112:6(433), 433–455.
- van Rijn, L. C. (2007a). "Unified view of sediment transport by currents and waves. I: Initiation of motion, bed roughness, and bed-load transport." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(2007)133:6(649), 649–667.
- van Rijn, L. C. (2007b). "Unified view of sediment transport by currents and waves. II: Suspended transport." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(2007)133:6(668), 668–689.
- Wang, G., Wu, B., and Wang, Z.-Y. (2005). "Sedimentation problems and management strategies of Sanmenxia reservoir, Yellow River, China: Management strategies of Sanmenxia reservoir." *Water Resour. Res.*, 41(9), W09417.
- Wu, B., van Maren, D. S., and Li, L. (2008a). "Predictability of sediment transport in the Yellow River using selected transport formulas." *Int. J. Sediment Res.*, 23(4), 283–298.
- Wu, B., Wang, G., Xia, J., Fu, X., and Zhang, Y. (2008b). "Response of bankfull discharge to discharge and sediment load in the Lower Yellow River." *Geomorphology*, 100(3–4), 366–376.
- Wu, C. L., Chau, K. W., and Li, Y. S. (2009). "Methods to improve neural network performance in daily flows prediction." *J. Hydrol.*, 372(1–4), 80–93.
- Yang, C. T. (1996). *Sediment transport: Theory and practice*, McGraw-Hill, New York.
- Yellow River Conservancy. (2013). "The Yellow River sediment bulletin." Zhengzhou, China (in Chinese).
- Yu, L., and O'Brien, J. J. (1991). "Variational estimation of the wind stress drag coefficient and the oceanic Eddy viscosity profile." *J. Phys. Oceanogr.*, 21(5), 709–719.
- Zhang, M., and Zhang, F. Q. (2012). "E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model." *Mon. Weather Rev.*, 140(2), 587–600.
- Zhang, R. J. (1961). *River dynamics*, China Industry Press, Beijing (in Chinese).