Variational assimilation of remotely sensed flood extents using a two-dimensional flood model

X. Lai\textsuperscript{1}, Q. Liang\textsuperscript{2}, and H. Yesou\textsuperscript{3}

\textsuperscript{1}State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography & Limnology, CAS, Nanjing 210008, China
\textsuperscript{2}School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
\textsuperscript{3}SERTIT, Université de Strasbourg, Bd Sébastien Brant, BP 10413 67412 Illkirch, France

Received: 23 July 2013 – Accepted: 18 August 2013 – Published: 27 August 2013

Correspondence to: X. Lai (xjlai@niglas.ac.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

A variational data assimilation (4D-Var) method is proposed to directly assimilate flood extents into a two-dimensional (2-D) dynamic flood model, to explore a novel way of utilizing the rich source of remotely sensed data available from satellite imagery for better analyzing or predicting flood routing processes. For this purpose, a new cost function is specially defined to effectively fuse the hydraulic information that is implicitly indicated in flood extents. The potential of using remotely-sensed flood extents for improving the analysis of flood routing processes is demonstrated by applying the present new data assimilation approach to both idealized and realistic numerical experiments.

1 Introduction

Flooding poses a significant threat to human society. Nowadays, floods are becoming more frequent as a result of intensive regional human activities and environmental change. Hydraulic or hydrodynamic models have become reliable and cost-effective tools to analyze and predict flood routing through catchments, rivers and floodplains. These models can provide dynamic outputs, e.g. inundation area, water depth, and/or flow velocity, for flood warning and risk assessment. Nevertheless, models are not perfect and uncertainties and computational errors may arise from various sources, including the uncertainties associated with hydrological parameters, initial and boundary conditions, as well as numerical errors as a result of numerical discretization and mathematical approximations. In order to eliminate errors and reduce uncertainties, field measurements are usually used to verify and calibrate a model before applying it to make predictions. Traditional trial and error approaches are commonly used in model calibration but they are well-known to be subjective and tedious (Ding, 2004). Therefore, in order to make a better prediction, it would be more beneficial to have more intelligent calibration methods achieved by fusing a dynamic flood model with observed information to obtain an optimal estimate of model states and parameters.
Data and model fusion methods are termed data assimilation, which stems from meteorology and oceanography (McLaughlin, 2002; Reichle, 2008; Wang et al., 2000). The variational data assimilation method, also called the 4D-Var method, is based on the optimal control theory of partial differential equations, which offers a powerful tool for data assimilation (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987). This method has been widely applied to improve the predictive capability of one-dimensional (1-D) and two-dimensional (2-D) hydraulic models (Atanov et al., 1999; Bélanger and Vincent, 2005; Ding, 2004; Honnorat et al., 2007, 2009; Roux and Dartus, 2006).

In river hydraulics, the available measurements commonly include water stage (level) and discharge at hydrological stations, and velocity at gauging points. These measurements are generally sparse even for those study areas with decent monitoring systems and therefore likely to be insufficient to support reliable model calibration. During a flood event, the available measurements may be even scarcer due to malfunctioned operation of some monitoring systems under extreme flow conditions and the difficulty in performing field surveys. Fortunately, rich sources of remote sensing data with different spatial and temporal coverage now become increasingly available. Remote sensing imagery provides spatially distributed information about flood states which is hard to obtain from the traditional point-based field measuring approaches (Hostache et al., 2010). As a whole, due to their low-cost and large coverage, remotely sensed data are now becoming an important source of measurements and widely applied to flood monitoring and loss evaluation for flood hazards (Pender and Néelz, 2007). Furthermore, recent intensive researches, such as the direct estimation of hydraulic variables (e.g. water discharge and stage) from satellite imagery, the use of remote sensing data to calibrate and validate model, the fusion of these data with dynamic model using data assimilation method and among others, have significantly contributed to the advances of the integrating remotely sensed data from space with flood models (e.g. Schumann et al., 2009; Smith, 1997).
Substantial efforts have been made using the 4D-Var and Bayesian-updating methods to demonstrate the potential of assimilating remotely sensed data from space for improving flood prediction (Andreadis et al., 2007; Durand et al., 2008; Komma et al., 2008; Roux and Dartus, 2006) attempted to determine flood discharge from remotely sensed river width using a 1-D hydraulic model. In 2-D river hydraulic modeling, 4D-Var methods have been developed to assimilate spatially distributed water stage (Lai and Monnier, 2009) and Lagrangian-type observations, e.g. remotely sensed surface velocity (Honnorat et al., 2009, 2010). Hostache et al. (2010) employed a 4D-Var method to assimilate the water stage derived from a RADARSAT-1 image of the 1997 Mosel River flood event in France into a 2-D flood model to improve model calibration. However, since the water stage from satellite imagery is indirectly retrieved by estimating the elevation of waterline from the remotely observed flood extents and topographic maps or digital elevation models (DEMs) the accuracy is generally low and typically in a range of 40–50 cm (Alsdorf et al., 2007; Hostache et al., 2010; Matgen et al., 2010). Simple overlay analysis of DEM and flood extent map may lead to high errors to the order of meter even when a 30 m-resolution ERS ASAR image is used (Brakenridge et al., 1998; Oberstadler et al., 1997; Schumann et al., 2011). Generally, additional steps must be performed in order to obtain an acceptable estimation of water levels for the use with hydrodynamic modeling. The complexity of these steps varies with the methods being applied (Matgen et al., 2007, 2010; Raclot, 2006; Schumann et al., 2007). For instance, Raclot (2006) and Hostache et al. (2010) used a hydraulic coherence constraint to minimize the estimation errors. Schumann et al. (2007) proposed a Regression and Elevation-based Flood Information eXtraction model (RFIX) for water depth estimation and later suggested an alternative for deriving water level from river cross-section data (Schumann et al., 2008). Therefore, the derivation of water level from flood extent with acceptable accuracy is not a straightforward procedure. Additional steps, i.e. the complex procedure of retrieving water stage, are required when conducting the assimilation of remotely sensed water stage into a hydraulic model.
In contrast to water stage, the remotely sensed flood extent can be directly derived from satellite imagery without affecting the original resolution (for example 30 m for Envisat ASAR and 250 m for MODIS data), which is comparable to the mesh size normally adopted in flood modeling. Various simple and mature approaches are available for rapid and automatic extraction of flood extent map from optical and SAR imageries (Matgen et al., 2011; Smith, 1997). However, to the best of our knowledge, there has been no attempt at the direct assimilation of flood extent data into a 2-D dynamic flood model using a 4D-Var method to date.

Herein, we attempt to use a 4D-Var method to assimilate remotely sensed flood extent data into a dynamic flood model based on the numerical solution to the 2-D shallow water equations (SWEs). For this purpose, a new cost function is specifically constructed to effectively fuse the hydraulic information available implicitly in flood extents. The numerical results show that the proposed 4D-Var method can effectively assimilate the flood extent data and improve the prediction accuracy of flood routing. The rest of the paper is organized as follows. First, a short description is given in Sect. 2 to introduce the 2-D flood model coupled with a 4D-Var method. In order to implement the assimilation of the observed flood extent into the 2-D flood model, Sect. 3 proposes a cost function that measures the discrepancy between observed data and modeling results. The new approach is validated by an idealized test in Sect. 4 before being applied to a realistic case in Sect. 5. Finally, summary and brief conclusions are drawn in Sect. 6.

2 Two-dimensional dynamic flood model with variational data assimilation

2.1 Overview of variational data assimilation

4D-Var is a method based on the optimal control theory of a physical system governed by partial differential equations. It allows us to perform flow state analysis or prediction for a system by combining a physically based dynamic model with observations. To
implement a 4D-Var, a cost function must be firstly defined to measure the discrepancy between the computational results and observations. A cost function \( J \) without regularization terms may be given as

\[
J(p) = \frac{1}{2} \int_0^T \|H U - O\|^2 dt = \frac{1}{2} \int_0^T (H U - O)^T W^{-1} (H U - O) dt
\]

(1)

where \( p \) is the control vector, \( \| \cdot \| \) is the Euclidean norm, \( H \) is the observation operator that maps the space of the state variables to the space of observations, \( U \) is the vector of state variables, \( W \) is the error covariance matrix, and \( O \) is the observed data. Herein, the statistical information can be incorporated into the norm through the error covariance matrix \( W \).

4D-Var can be considered as an unconstrained optimization problem that seeks an optimal control vector \( p^* \) to minimize the cost function \( J(p) \) in Eq. (1). According to the optimal control theory, optimum conditions are reached if the gradient \( J = 0 \), which means that an optimal control vector is obtained and the optimal flow analysis results are closest to the true (measured) state. This optimization problem may be solved by a descent-type algorithm and the quasi-Newton minimization subroutine M1QN3 developed by (Gilbert and Lemaréchal, 1989) is adopted in this work. The algorithm calculates the gradient of the cost function, i.e. the vector of its partial derivatives with respect to each of the control variables, which may be efficiently performed using the adjoint method as described in Sect. 2.3.

### 2.2 Two-dimensional shallow water equations

The 2-D SWEs are widely used to approximate flood routing over a floodplain. They can be written in a conservative form as follows:

\[
\frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} + \frac{\partial G(U)}{\partial y} = B(U)
\]

(2)
where \( x \) and \( y \) represent the Cartesian coordinates, \( t \) is the time, \( \mathbf{U} = (h, hu, hv)^T \) is a vector containing the flow variables with \( h \) being the water depth and \( u \) and \( v \) the two velocity components, \( \mathbf{F} = \left(hu, hu^2 + 0.5gh^2, hv\right)^T \) and \( \mathbf{G} = \left(hv, huv, hv^2 + 0.5gh^2\right)^T \) are the flux vectors in the \( x \) and \( y \) directions, \( g \) is the gravitational acceleration, \( \mathbf{B} = [0, gh(S_{0x} - S_{fx}), gh(S_{0y} - S_{fy})]^T \) is the vector of the source terms, \( S_{0x} = -\partial Z_b/\partial x \) and \( S_{0y} = -\partial Z_b/\partial y \) are the two bottom slopes with \( Z_b \) denoting the bed elevation, and \( S_{fx} = n^2 q_x h^{-7/3} \sqrt{q_x^2 + q_y^2} \) and \( S_{fy} = n^2 q_y h^{-7/3} \sqrt{q_x^2 + q_y^2} \) are the two friction slopes in \( x \) and \( y \) directions, respectively, with \( n \) being the Manning roughness coefficient. Given initial and boundary conditions, the flood routing process over a floodplain may be numerically predicted on different temporal and spatial scales by solving the above governing equations.

### 2.3 Adjoint governing equations

Although several methods are available for evaluating the gradient of the cost function, the adjoint method is favorable, owing to its relatively low computational burden for large-scale problems (Cacuci, 2003). The adjoint equations for the 2-D SWEs can be derived for the cost function Eq. (1) as follows:

\[
\frac{\partial \mathbf{U}^*}{\partial t} + \frac{\partial \mathbf{F}^T}{\partial \mathbf{U}} \frac{\partial \mathbf{U}^*}{\partial x} + \frac{\partial \mathbf{G}^T}{\partial \mathbf{U}} \frac{\partial \mathbf{U}^*}{\partial y} = -\frac{\partial \mathbf{B}^T}{\partial \mathbf{U}} \mathbf{U}^* + \mathbf{H}^T \mathbf{W} (\mathbf{O} - \mathbf{HU})
\]
Where the adjoint variable $U^* = (h^*, q^*_x, q^*_y)$ and the coefficient matrices are given by

$$
\frac{\partial F}{\partial U}^T = \begin{pmatrix}
0 - u^2 + c^2 - uv \\
1 2u v \\
0 0 u
\end{pmatrix}, \quad \frac{\partial G}{\partial U}^T = \begin{pmatrix}
0 -uv -v^2 + c^2 \\
v 0 0 \\
1 u 2v
\end{pmatrix},
$$

$$
\frac{\partial B}{\partial U}^T = \begin{pmatrix}
0 gS_{0x} + \frac{7}{3} gS_{fx} & gS_{0y} + \frac{7}{3} gS_{fy} \\
0 - gS_{ fx} \frac{2u^2+v^2}{u(u^2+v^2)} & -gS_{fy} \frac{u^2+2v^2}{u^2+v^2} \\
0 - gS_{ fx} \frac{v}{u^2+v^2} & -gS_{fy} \frac{v u^2}{v(u^2+v^2)}
\end{pmatrix}.
$$

The partial derivative of the cost function $J$ corresponding to the control vector $p$ is a simple function of the adjoint variables $U^*$, which can be found in Lai and Monnier (2009).

Adopting the adjoint equations in gradient computation significantly reduces the computational cost because evaluation of the adjoint variables requires only one backward integral in time. Once the adjoint variables are known, the partial derivatives of the cost function with respect to the control variables can be computed in a straightforward way.

### 2.4 Forward model and adjoint model

The 2-D SWEs Eq. (2) are discretized using a finite volume Godunov-type scheme with the inter-cell mass and momentum fluxes evaluated using the HLLC approximate Riemann solver (Toro, 2001). The scheme has first-order accuracy in space but provides physically based representation of flow discontinuities. Time discretization is achieved using an explicit Euler scheme. Readers may consult Honnorat et al. (2007) for a more detailed description of the shallow flow model, which is referred to as forward model herein.

The adjoint model is developed by directly differentiating the source codes of the forward model that solves the 2-D SWEs in Eq. (2). The automatic differentiation tool TAPENADE (Hascoët and Pascual, 2004) is adopted in this work to generate the
reverse codes. This method, based on source codes, helps to build a consistent adjoint model corresponding to the forward solver.

3 Cost function for flood extent assimilation

As mentioned previously in the introduction, the flood extent can be derived from satellite imagery more directly and easily than the water stage. However, the flood extent is not a state variable in the 2-D SWEs but basically the union of pixels where water depth is not zero. Therefore it has no explicit relationship to the state variables. As a consequence, it is difficult to define a cost function to implement the assimilation of flood extent in the framework of 4D-Var. In this work, a new cost function is constructed to facilitate the assimilation of flood extent into a 2-D dynamic flood model.

If we assume a function $f$ as an observable quantity, the cost function may be defined as:

$$J(p) = \frac{1}{2} \int_{0}^{T} \| f - f^{\text{obs}} \|^2 dt$$  \hspace{1cm} (4)

In this work, the regularization terms are neglected from the above cost function to facilitate simplified but more informative verification and validation of the proposed method and allow direct investigation of the potential benefit of assimilating flood extent data.

A specific form of $f$ should be introduced to determine the cost function for assimilation of the hydraulic information included implicitly in the remotely sensed flood extent data. If $f$ is defined as the total volume of water, it may be applied to model the gradually-varied flow, e.g. in a seasonal lake using a known area-volume curve. Inspired by this, we propose a method based on the volume of water over each computational cell, which facilitates the computational requirements of more challenging cases of rapidly-varying flow, including the dyke breach induced flood waves considered in this paper.
The definition of the cost function is illustrated in Fig. 1. Also demonstrated in Fig. 1 are the predicted and observed flood extents, and obviously, difference commonly exists between the observed flood extent and the predicted one that results from the initially guessed model parameters. The water depth along the boundary line of the flood extent should be theoretically zero. In practice, however, a finite threshold of water depth, $h_c$, must be introduced in order for the flood extent to be extracted from the remotely sensed data. This is essential in order to minimize the effects of the disturbances from different land covers, the resolution of the image, and other sources of uncertainty as suggested by Aronica et al. (2002). Taking into account the uncertainty of the observed inundation areas, a confidence coefficient $w$ is introduced to represent the wet-dry status of each pixel or the degree of certainty of a pixel being wet in a remotely sensed image. As shown in Fig. 1, $w = 1$ indicates a pixel being definitely wet and $w = 0$ denotes a pixel being absolutely dry, where the values in between is given according to the level of certainty of a pixel being wet. The observed flood extent map can then be depicted in 2-D raster format with pixel values equal to $w$. When observations are used, they should be mapped into the model space by an observation operator.

Water depth $h$, representing the unit water volume of a computational cell, is adopted herein to facilitate the definition of $f$. However, the difficulty lies on how to determine the observed water depth. Herein it is assumed that the observed and predicted water depths are the same in those overlapping areas between the predicted and observed extents. In those non-overlapping regions, different assumptions have to be made, depending on the specific location under consideration. The computational domain is separated into two parts as illustrated in Fig. 1, i.e. $\Omega_1$ represents the region with predicted water depth $h > h_c$ while $\Omega_2$ is the area outside of $\Omega_1$. Inside $\Omega_1$, the observed water depth is defined to be “zero” (essentially $h_c$) if the cell under consideration is outside the area covered by the remotely sensed flood extent. As a result, the cost function in $\Omega_1$ may be defined as $J_1 = 0.5(1 – w)^2 h^2$, where $w$ is the certainty of flooding as described in the above paragraph. Obviously, $J_1$ decreases to zero when the
predicted and observed extents coincide. Inside $\Omega_2$, an observed water depth, $h_{\text{obs}}$, is required to construct the cost function in those areas covered by the remotely sensed flood extent. Numerical experiments show that it is feasible to set $h_{\text{obs}} = 2h$ to keep a similar gradient along the boundary, which leads to a cost function $J_2 = 0.5w^2h^2$ in $\Omega_2$. $J_2$ will also decrease to zero when the predicted and observed extents coincide.

Taking into account all of above considerations, the cost function measuring the discrepancy of observations and predictions over computational domain may be written as:

\[
J(p) = 0.5\alpha \left( \sum_{\Omega_1} (1 - w_i)^2h_i^2 + \sum_{\Omega_2} w_i^2h_i^2 \right)
\]  

This form of cost function carries a physical dimension. $\alpha$ is the scaling parameter. When the water stage or velocity observations are assimilated together, the scaling parameter should be imposed in Eq. (5) to respect an initial balance between different terms of cost function.

4 Idealized test case

In order to verify the performance of the current model in data assimilation, a series of numerical experiments are carried out using a laboratory-scale dyke breach test, where the synthetic data generated from model are used to eliminate the disturbances of numerical and measured errors encountered in an actual case.

4.1 Description of the test case

We consider a flood routing process induced by a dyke break over a $10 \times 8$ m rectangular floodplain with a flat bottom, i.e. $Z_0 = 0$. As shown in Fig. 2a, the left boundary represents a river bank with a breach of 0.4 m in the middle. The floodplain consists
of five types of land covers corresponding to Manning’s $n$ 0.03, 0.04, 0.05, 0.06 and 0.07 respectively, from left to right. The computational domain has been discretized into a uniform mesh of 0.2 m $\times$ 0.2 m resolution. During the simulation, a fixed time step of 0.01 s is used. The boundary discharge hydrograph $Q_i(t)$ (half of total discharge through dyke breach to floodplain) is shown in Fig. 2b and imposed on each of the two breach cells. The other three lateral boundaries of the floodplain are assumed to be solid walls. The floodplain is initially dry.

With the aforementioned “accurate” $n$ set for each land cover, the dyke-break flow routing process is firstly simulated by the forward model for 5 s over the floodplain. Synthetic binary maps of the flood extent and the time history of water stage at the middle of the domain are generated and will be used as observed data during the following numerical experiments. Five groups of observations are obtained, as listed in Table 1, with different combinations of synthetic flood extents and/or the stage hydrograph at the central point. The assimilation window is set to be 5 s, the same as the duration of the forward simulation. Three series of numerical experiments are carried out by controlling $n$, $Q_i(t)$ or both of them, respectively.

### 4.2 Experiment series A

The control variable of the experiment series A is the distributed Manning coefficient $n$. Five assimilation experiments are run with the same first guess of $n_0 = 0.02$ over whole floodplain, but with different groups of synthetic data being assimilated. In each run, the optimal analysis of flood routing over the floodplain is undertaken and the distributed $n$ is retrieved, as provided in Table 2.

Table 3 lists the root-mean-square (RMS) errors of water depth over the whole computational domain at different output times. For the runs involving the observations of Groups A and B which just assimilate flood extents, the RMS errors decrease by 78% and 94%, respectively. This is also clearly demonstrated by comparing the flood extents obtained from different runs that assimilate different observations (Fig. 3a). After data assimilation, the predicted flood extents are significantly improved and agree

---

**HESSD**

10, 11185–11220, 2013

Variational assimilation of remotely sensed flood extents

X. Lai et al.
much more closely with the “observed” extents. The more observed flood extent data being assimilated, the closer the results become to the “true” state. In the numerical experiment involving water stage observations (Group C), only the stage hydrograph is assimilated and the RMS errors decrease by 82% on average. However, the predicted results at \( t = 3–5 \) s are significantly different to the “true” states, which can be also seen evidently from the difference between the predicted and “true” flood extents (Fig. 3a). The results from simulations using Groups D and E observations show that the RMS errors are further decreased by about 95% after assimilating both the time series of water stage and spatial flood extents.

As a whole, by assimilating different synthetic data, different level of improvement in flood prediction has been achieved during the numerical experiments, which leads to the assimilated predictions that are always much closer to the “true” state. It confirms that the current assimilation analysis of fusing observed flood extent and relevant information improves the accuracy of flood prediction in both space and time (Fig. 5a). The quality of the assimilated results can also be confirmed from the identified \( n \), as listed in Table 2. The value of \( n \) for the first land block can be accurately identified in all of the experiments, regardless of whether flood extent or stage hydrograph is assimilated. However, since the stage hydrograph only provides upstream information, it cannot optimize the values of \( n \) for the downstream land blocks 4 and 5. Therefore the \( n \) values remain to be their initial guess in the numerical experiment using the Group C observations, which leads to apparent difference between the simulated and “true” extents after \( t = 3–5 \) s (Fig. 3a).

### 4.3 Experiment series B

Taking the inflow discharge as a control variable, we carried out further numerical experiments using the five given groups of observations. The initial guesses of discharge calculated by \( Q_i^0 = Q_i(1 + 0.6R) \) with \( R \) being a random number between 0 and 1 are imposed through the inflow boundary. With the help of the minimization algorithm, the initial guesses of the discharge boundary condition are corrected and the corresponding
analysis results after data assimilation are computed. The hydrographs of inflow discharge for numerical experiments using the Groups B, C and E observations are shown in Fig. 4a. They are slightly corrected to minimize the cost function.

The RMS errors of each run at $t = 1, 2, 3, 4$ and $5$ s are listed in Table 3. They decrease by $28 \sim 32\%$ for those simulations assimilating the flood extents, but only $5\%$ for runs just assimilating point-based data provided as the stage hydrograph. Figure 3b compares the predicted and “true” flood extents.

In this experiment series, it is interesting to note that better prediction over the whole duration and spatial extent (Table 3 and Fig. 3b) is produced by assimilating flood extent, even though poor prediction of water stage hydrograph at the central gauge station is found (Fig. 5b). This may be attributed to the fact that flood extent data covering the whole study area include inflow information from inlets over the whole time scale. Assimilation of these data can help to estimate the inflow hydrograph and then increase the assimilation accuracy. On the contrary, point-based time series data only imply part of the inflow discharge information prior to the propagation time from the inlet to the given points. The inclusion of point-based measurements helps to improve the accuracy of the stage hydrograph at the central station but has no obvious benefit for prediction for the whole duration and spatial extent.

### 4.4 Experiment series C

In the experiment series C, both the Manning coefficient and the inflow discharge hydrograph are controlled. The same initial guesses of $n$ and discharge are used. After running the assimilation model, $Q_i(t)$ and the distributed $n$ are corrected to minimize the cost function. Although the discharge hydrograph (Fig. 4b) and $n$ (Table 2) of each run are not well identified, the predictions (Fig. 3c) obtained after assimilating the flood extents are much closer to the “true” one than those just assimilating point-based measurements. The RMS errors of the runs assimilating the observations of Group A, B, C, D and E decrease by $50\%$, $64\%$, $45\%$, $48\%$ and $41\%$, respectively, as listed in Table 3. It is encouraging to observe that almost half of the RMS errors decrease for
5 Assimilation of an actual remotely sensed flood extent

Based on the findings of the previous numerical experiments, this section intends to investigate further the potential of the proposed data assimilation method using actual satellite remote sensing data (here MODIS). The study area, Mengwa Flood Detention Area (MFDA), is located at Fuyang, Anhui Province of China, the middle reach of the Huaihe River. It is the most important region for flood control within this river basin. MFDA covers a narrow and elongated area of 180 km$^2$ (Fig. 6a), with a population of 148,000 farming 120 km$^2$ of cropland. The domain is discretized using an unstructured grid (Fig. 6b) consisting of 1222 nodes and 1136 quadrilateral and triangular cells. The size of the cell edges ranges from 200 m to 400 m. The bed elevation at each cell is extracted from a digital elevation model (DEM) of 100 m resolution, which is generated from a 1:2500 topographic map.

The data assimilation experiments are carried out based on the flood routing process over MFDA induced by the flood diversion event happened in the summer of 2007. From 29 June to 15 July 2007, persistent heavy rain was experienced in the Huaihe River basin. To reduce the risk of severe flooding that might cause significant economic and human loss downstream, MFDA was operated by opening the Wangjiaba gate to receive flood water from the Huaihe River starting from 04:28 UTC 10 July, with an order from the Chinese central government. Until 12 July 2007, the total diverted volume reached about $0.25 \times 10^9$ m$^3$, which effectively stored and retained flood water and hence reduced flood risk. Figure 6 plots the 45 h inflow hydrograph to MFDA through the flood gate, from 04:28 UTC 10 July 2007.

Two MODIS instruments on the Terra and Aqua spacecraft platforms have provided daily measurements with the global coverage since 1999. The 250 m resolution with each run. As in the experiment series B, although the inclusion of point-based measurements improves the accuracy of the stage hydrograph at the central station, no obvious improvement is detected in terms of overall RMS errors.
daily revisits makes them particularly suitable for monitoring the changes of flooding over a floodplain. Herein, we downloaded one scene of Aqua-MODIS Level-1B and Geo-location data covering the whole MFDA from the Level 1 and Atmosphere Archive and Distribution System (LAADS). The MODIS data acquired at 06:00 UTC with 250 m resolution capturing the flood routing during the flood diversion event. Although MFDA was partly covered by light cloud at that moment, the image is of sufficient quality to identify the flood extent.

A simple method is adopted to extract the flood extent based on the luminance of the composite image from the Band 7-2-1 combination. The luminance $L$ of each pixel is firstly calculated using the following formula (Gonzales and Woods, 2002)

$$L = 0.299b_7 + 0.587b_2 + 0.114b_1$$  \hspace{1cm} (6)

where $b_7$, $b_2$ and $b_1$ are the digital values of Band 7, Band 2 and Band 1. The luminance image is shown in Fig. 7a, after setting the pixel to null value where heavy cloud covered. The flood extent is then easily extracted over MFDA by setting a critical value of luminance as a threshold to separate the water area from image. However, due to the fact that the extraction of flood extent may be affected by the land surface, such as trees and vegetation cover (Smith, 1997), and the current image is in relatively low resolution of 250 m, there exists certain uncertainties in the boundary water line. In light of this, the concept of membership degree from the Fuzzy Set Theory (Huang, 2000; Nguyen and Walker, 2006) is introduced as an indicator to determine the flood extent. The degree of membership $w$ quantifies the grade of membership of an element to a fuzzy set, which is herein the possibility of a pixel being wet. A membership function may be written as (Huang, 2000)

$$w = \begin{cases} 
1, & L_i \leq a \\
0.5 + 0.5 \sin \left( \frac{\pi}{b-a} \cdot \left( L_i - \frac{a+b}{2} \right) \right), & a < L_i \leq b \\
0, & L_i \geq b
\end{cases}$$  \hspace{1cm} (7)
where $L_i$ is the luminance of pixel $I$, $a$ and $b$ are the upper and lower bounds of the luminance to separate the water and land. The degree of membership $w = 0$ and $w = 1$ mean that pixel $i$ is completely dry and wet, respectively. A value between 0 and 1 characterize fuzzy members that are only partially wet/dry. For those areas covered by heavy clouds, null values are given to the corresponding pixels and these cells are excluded from the evaluation of cost function.

From visual interpretation, it may be clearly demonstrated that those areas with luminance $L_i$ less than 110 are covered by water and hence $a = 110$. The upper bound $b$ is more difficult to determine owing to the effects of complicated land cover. In this paper, $b = 121$ and 126 are respectively examined. The flood extents retrieved from fixed thresholds 110, 121 and 126 are shown in Fig. 7b–d.

Taking the membership degree computing from Eq. (7) as a weighting factor $w$ and substituting it into Eq. (5), the cost function $J$ is obtained as

$$J = \frac{1}{2} \left[ 0.5(1 - \frac{h - h_c}{|h - h_c|}) - w \right]^2 h^2$$  \hspace{1cm} (8)

Based on this cost function, data assimilation experiments are conducted with a computational time step of 12 s. The simulation time is set to 36 h, starting from the gate opening at 04:28 UTC 10 July 2007. The actual discharge hydrograph for flood diversion to MFDA as shown in Fig. 6c is imposed through the inflow discharge boundary. Simulation starts from an originally dry floodplain. The critical water depth to derive the boundary line of flood extent from remote sensing data, $h_c$, is set to 0.2 m.

The Manning roughness coefficient, $n$ was assumed to be constant over the whole computational domain because of little knowledge about land use or cover. The control variable of the numerical experiments is the lumped Manning's $n$, namely the control vector contains only one element. Giving different $n_0$ (Table 4), we carried out six numerical simulations, assimilating one single remotely sensed flood extent from MODIS data at $t = 25.5$ h with $b = 121$ and 126. The minimized cost functions of the
experiments with \( b = 126 \) are less than those with \( b = 121 \), but the values are close to their minimum for an independent \( b \) (Table 4).

Figure 8 shows the computed flood extents before and after data assimilation. It can be observed that consistent flood extents are obtained in the assimilation experiments with different \( n_0 \) by assimilating the flood extent information from MODIS data. Also, it is obvious that the computed flood extents are improved after data assimilation has been performed in both experiments. The estimated flood extents are much closer to the one extracted from MODIS (Fig. 7). The findings are encouraging, which indicate that hydraulic information from satellite imagery can be directly assimilated into a 2-D dynamic flood model via the flood extent using the cost function as suggested in this work.

We also identify a consistent \( n \) in the assimilation experiments with different \( n_0 \), as listed in Table 4. The identified \( n \) is about 0.2 \( \sim \) 0.25, partly depending on \( n_0 \). It is greater than the empirical value of a normal floodplain, which may be caused by the loss of accuracy from the low resolution MODIS data and uncertainties in the domain topography, etc. In addition, minimization procedure of the 4D-Var method seems to be trapped into the local minimal value for different \( n_0 \) in our experiments. Taking the experiments with \( b = 126 \) as an example, the optimized \( n \) is 0.208 if \( n_0 = 0.025 \) or 0.030; but it is close to 0.24 if \( n_0 = 0.5 \) or 0.8. After checking the relationship between the cost function and \( n \) (Fig. 9), two local minimal values of cost function exist when \( n \) is close to 0.20 or 0.24. This leads to different estimations of \( n \) in our experiments. The double minima may originate primarily from the assumption of a constant \( n \) over the study area with heterogeneous landscapes, which is inconsistent with the actual situation. Furthermore, insufficient data (a single low resolution flood extent) may also lead to the appearance of double minima in the cost function.
6 Summary and conclusions

To the best of our knowledge, no attempt has been reported to directly assimilate the flood extent data into a 2-D flood model in the framework of 4D-Var. In this work, a 4D-Var method incorporated with a new cost function is introduced to advance this research topic. The new approach has been validated using a series of numerical experiments undertaken for an idealized test case before applying to a realistic simulation in MFDA. The main results of this study are summarized as follows:

- A new cost function is defined to facilitate assimilation of flood extent data directly using a 4D-Var method. While it can efficiently help the 2-D flood model to assimilate the spatially distributed flood dynamic information of the flood extent data from remote sensing imagery, the current approach does not require those additional steps of retrieving water stage (Hostache et al., 2010). Since the flood extent is much easier to map from remote sensing image than water stage and gradients (Schumann et al., 2009), the present scheme provides a more promising way of data assimilation for flood inundation modeling. However, as a new data assimilation method for flood modeling, an interesting research question to answer is whether the direct assimilation of flood extent data can improve the assimilation accuracy comparing with the assimilation of water level observations retrieved from the same data sources of satellite imagery. This is worth a comprehensive comparative study in the future, which may then provide useful guideline for the practical applications of remote sensing data assimilation.

- Flood extent is a type of spatially distributed data and implicitly implies hydraulic information of flood routing. The observed flood extent data may provide an alternative to obtaining a denser time series as stated by Roux and Dartus (2006) and to compensating for unavailable field measurements during a flood event (Lai and Monnier, 2009). The assimilation of flood extent data is suitable for improving flood modeling in the floodplains or similar areas (e.g. seasonal lakes with significant wetting and drying processes) with slowly varying bed-slopes. However, it
should be noted that this approach has its own limitation. If the flood extent does not contain enough hydraulic information, the assimilation exercise may fail. For example, in the case of flood inundation in a domain constrained by steep slopes, the water stage but not the flood extent varies evidently with time. Since the extent data do not actually represent the physical evolution of such a flood event, they are not suitable for assimilation. Therefore, the correlation between extent and flood dynamics must be established before applying the current data assimilation scheme.

- The results of flood modeling are much improved by successfully estimating the roughness parameter over a floodplain even though the low-resolution MODIS data (250 m) is adopted in the application of MFDA. This implies that the proposed method may extend the usable data sources for assimilation to the imageries from most of currently in-orbit satellites that provide large spatial and temporal coverage.

Overall, this study shows that the assimilation of the flood extent data is effective in improving flood modeling practice. Future work should be carried out to understand the full potential of this new way of making use of spatially distributed data from various existing satellites in data assimilation.

Acknowledgements. The research was supported by National Key Basic Research Program of China (973 Program) (2012CB417000) and the National Natural Science Foundation of China (Grant No. 50709034, No. 41071021 and No. 41001348).

References


Table 1. The five groups of observations used in the idealized test.

<table>
<thead>
<tr>
<th>Group</th>
<th>Description of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Flood extent at $t = 5$ s</td>
</tr>
<tr>
<td>B</td>
<td>Flood extents at $t = 1, 3$ and $5$ s</td>
</tr>
<tr>
<td>C</td>
<td>$Z(t)$, time history of water stage at central position of floodplain (time interval of measurement is $0.2$ s)</td>
</tr>
<tr>
<td>D</td>
<td>Flood extent at $t = 5$ s and $Z(t)$</td>
</tr>
<tr>
<td>E</td>
<td>Flood extents at $t = 1, 3$ and $5$ s and $Z(t)$</td>
</tr>
</tbody>
</table>
Table 2. The identified manning roughness coefficients, $n$ in experiment series A and C.

<table>
<thead>
<tr>
<th>observations</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>True value</td>
<td>–</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>First guess</td>
<td>–</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>0.031</td>
<td>0.053</td>
<td>0.053</td>
<td>0.028</td>
<td>0.042</td>
</tr>
<tr>
<td>Group B</td>
<td>0.030</td>
<td>0.038</td>
<td>0.054</td>
<td>0.036</td>
<td>0.074</td>
</tr>
<tr>
<td>Group C</td>
<td>0.030</td>
<td>0.040</td>
<td>0.050</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>Group D</td>
<td>0.030</td>
<td>0.040</td>
<td>0.050</td>
<td>0.042</td>
<td>0.070</td>
</tr>
<tr>
<td>Group E</td>
<td>0.030</td>
<td>0.04</td>
<td>0.05</td>
<td>0.038</td>
<td>0.072</td>
</tr>
<tr>
<td>Series C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>0.024</td>
<td>0.061</td>
<td>0.118</td>
<td>0.099</td>
<td>0.220</td>
</tr>
<tr>
<td>Group B</td>
<td>0.031</td>
<td>0.069</td>
<td>0.057</td>
<td>0.032</td>
<td>0.046</td>
</tr>
<tr>
<td>Group C</td>
<td>0.020</td>
<td>0.052</td>
<td>0.040</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>Group D</td>
<td>0.052</td>
<td>0.047</td>
<td>0.052</td>
<td>0.039</td>
<td>0.049</td>
</tr>
<tr>
<td>Group E</td>
<td>0.047</td>
<td>0.077</td>
<td>0.026</td>
<td>0.023</td>
<td>0.030</td>
</tr>
</tbody>
</table>
Table 3. The RMS errors of water depth at $t=1,2,3,4$ and 5 s in three series of numerical experiments.

<table>
<thead>
<tr>
<th>Control Variables</th>
<th>Time(s)</th>
<th>Guess</th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Group D</th>
<th>Group E</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0040</td>
<td>0.0009</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0183</td>
<td>0.0063</td>
<td>0.0010</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0313</td>
<td>0.0074</td>
<td>0.0017</td>
<td>0.0039</td>
<td>0.0009</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0396</td>
<td>0.0073</td>
<td>0.0026</td>
<td>0.0081</td>
<td>0.0018</td>
<td>0.0022</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0436</td>
<td>0.0076</td>
<td>0.0032</td>
<td>0.0117</td>
<td>0.0022</td>
<td>0.0028</td>
<td></td>
</tr>
<tr>
<td>$Q_i(t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0064</td>
<td>0.0051</td>
<td>0.0044</td>
<td>0.0037</td>
<td>0.0049</td>
<td>0.0049</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0097</td>
<td>0.0071</td>
<td>0.0076</td>
<td>0.0082</td>
<td>0.0083</td>
<td>0.0086</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0109</td>
<td>0.0081</td>
<td>0.0082</td>
<td>0.0125</td>
<td>0.0091</td>
<td>0.0091</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0132</td>
<td>0.0074</td>
<td>0.0076</td>
<td>0.0128</td>
<td>0.0069</td>
<td>0.0071</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0102</td>
<td>0.0067</td>
<td>0.0068</td>
<td>0.0103</td>
<td>0.0065</td>
<td>0.0064</td>
<td></td>
</tr>
<tr>
<td>$n$ and $Q_i(t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0095</td>
<td>0.0065</td>
<td>0.0050</td>
<td>0.0086</td>
<td>0.0073</td>
<td>0.0062</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0245</td>
<td>0.0102</td>
<td>0.0108</td>
<td>0.0142</td>
<td>0.0178</td>
<td>0.0196</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0373</td>
<td>0.0217</td>
<td>0.0149</td>
<td>0.0136</td>
<td>0.0198</td>
<td>0.0239</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0410</td>
<td>0.0191</td>
<td>0.0137</td>
<td>0.0260</td>
<td>0.0196</td>
<td>0.0230</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0514</td>
<td>0.0243</td>
<td>0.0142</td>
<td>0.0272</td>
<td>0.0201</td>
<td>0.0235</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. The identified $n$ and the final cost functions in the application to MFDA.

<table>
<thead>
<tr>
<th>Upper bound of luminance, $b$</th>
<th>Final cost function, $J$</th>
<th>Decrease rate of cost function (%)</th>
<th>Initial guess Manning coefficients, $n_0$</th>
<th>Identified Manning coefficients, $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>126</td>
<td>28.118</td>
<td>81.2</td>
<td>0.025</td>
<td>0.208</td>
</tr>
<tr>
<td></td>
<td>28.127</td>
<td>78.0</td>
<td>0.030</td>
<td>0.208</td>
</tr>
<tr>
<td></td>
<td>28.432</td>
<td>14.6</td>
<td>0.500</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>28.319</td>
<td>24.2</td>
<td>0.800</td>
<td>0.240</td>
</tr>
<tr>
<td>121</td>
<td>49.071</td>
<td>70.6</td>
<td>0.030</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>48.937</td>
<td>18.6</td>
<td>0.800</td>
<td>0.240</td>
</tr>
</tbody>
</table>
Fig. 1. Definition of a cost function. The predicted flood extent is enclosed by solid line. The area inside the dashed line is covered by the observed extent, part of which is shaded by oblique lines and has $0 < w < 1.0$.

$$
\Omega_1, J_1 = 0.5(1-w)^2h^2 \quad \Omega_2, J_2 = 0.5w^2h^2
$$
Fig. 2. Idealized test of flood routing over a rectangular floodplain induced by dyke breach: (a) computational domain; and (b) hydrograph of the inflow discharge $Q_i(t)$. 

11213
Fig. 3. Comparison of the predicted and “true” flood extents for different simulations at $t = 1, 2, 3, 4$ and $5$ s: (a) experiments series A; (b) experiments series B; and (c) experiments series C. The solid and dashed lines mark, respectively, the predicted and “true” flood extents.
Fig. 4. Identified discharge hydrograph from (a) experiment series B; and (b) experiment series C.
Fig. 5. Water stage validation at the gauge point in (a) experiment series A; (b) experiment series B; and (c) experiments series C.
Fig. 6. (a) Mengwa Flood Detention Area (MFDA); (b) unstructured grid; (c) inflow discharge hydrograph.
Fig. 7. (a) Luminance of MODIS image with Band 7-2-1; (b) flood extent extracted from the fixed digital number threshold 110; (c) flood extent extracted from the fixed digital number threshold 121; and (d) flood extent extracted from the fixed digital number threshold 126.
Fig. 8. Comparing flood extents obtained before and after assimilation of the remotely sensed flood extent from MODIS image specified by $b = 126$ when (a) $n_0 = 0.025$; and (b) $n_0 = 0.8$. The solid line represents the boundary of the flood extent after assimilation where water depth is equal to $h_c$. The filled area is the flood extent computed by forward model with $n_0$. 
Fig. 9. The relationship between the cost function and the Manning’s $n$. 

Cost function, $J$

Manning roughness coefficient, $n$