

DEVELOPMENT OF A HYDROLOGICAL MODELING FRAMEWORK FOR DATA ASSIMILATION WITH PARTICLE FILTERS

Seong Jin NOH^{1,2}, Yasuto TACHIKAWA³, Michiharu SHIIBA⁴, Kazuaki YOROZU⁵
and Sunmin KIM⁶

¹Student Member of JSCE, Graduate Student, Dept. of Urban and Environmental Eng., Kyoto University
(C1-1, KyotoDaigaku-Katsura, Nishikyoku, Kyoto 615-8540, Japan)
E-mail: seongjin.noh@gmail.com

²Researcher, Water Resources & Environment Research Department, Korea Institute of Construction Technology
(2311 Daewha-dong, Ilsan-gu, Goyang-si, Gyeonggi-do 411-712, Korea)

³Member of JSCE, Associate Professor, Dept. of Civil and Earth Resources Eng., Kyoto University
(C1-1, KyotoDaigaku-Katsura, Nishikyoku, Kyoto 615-8540, Japan)
E-mail: tachikawa@hywr.kuciv.kyoto-u.ac.jp

⁴Member of JSCE, Professor, Dept. of Civil and Earth Resources Eng., Kyoto University
E-mail: shiiba@hywr.kuciv.kyoto-u.ac.jp

⁵Member of JSCE, Assistant Professor, Dept. of Civil and Earth Resources Eng., Kyoto University
E-mail: yorozu@hywr.kuciv.kyoto-u.ac.jp

⁶Member of JSCE, Lecturer, Dept. of Civil and Earth Resources Eng., Kyoto University
E-mail: sunmin@hywr.kuciv.kyoto-u.ac.jp

Data assimilation (DA) methods have received increased attention as a means to accomplish uncertainty assessment and enhancement of prediction capability in various areas. Despite their potential, applicable software frameworks for probabilistic approaches and DA are still limited because most hydrological modeling frameworks are based on a deterministic approach. This paper presents a hydrological modeling framework for DA, namely MPI-OHyMoS. While adapting object-oriented features of the original OHyMoS, MPI-OHyMoS allows users to build a probabilistic hydrological model with DA. In this software framework, sequential DA based on particle filtering (PF) is available for any hydrological models considering various sources of uncertainty originating from input forcing, parameters, and observations. Ensemble simulations are parallelized by the message passing interface (MPI), which can take advantage of high-performance computing (HPC) systems. Structure and implementation processes of DA via MPI-OHyMoS are illustrated using a simple lumped model. We apply this software framework to uncertainty assessment of a distributed hydrological model in both synthetic and real experiment cases. In the synthetic experiment, dual state-parameter updating results in a reasonable estimation of parameters converging into the synthetic true. In the real experiment, dual updating also shows good conformity with the observed hydrograph, having reduced the uncertainty ranges of parameters. Deterministic modeling, based on parameters estimated via PF, shows good performance for extreme events, while dual updating via PF shows improved performance for all events.

Key Words: *data assimilation, MPI-OHyMoS, hydrological modeling framework, particle filtering, uncertainty assessment, dual state-parameter updating*

1. INTRODUCTION

DA has received increased attention from the hydrological research community due to its capability to handle explicitly the sources of uncertainty. Numerous sophisticated DA algorithms have been proposed, from ruled-based, direct-insertion methods to advanced smoothing and sequential techniques, as

well as the variants of these techniques¹⁾. In the hydrological research community, applications of DA have proved promising in improving prediction accuracy and quantifying uncertainty. Despite their potential, general modeling frameworks applicable to probabilistic approaches and DA are still limited because most modeling frameworks are based on a deterministic modeling approach. With the increas-

ing need for DA modeling platforms, a few frameworks such as openDA²⁾ and PCRater applications³⁾ have appeared recently. These approaches seem to provide innovative DA environments to overcome the limitations of conventional deterministic modeling. However, there still remain cumbersome procedures such as the construction of the model wrapper and further steps to be able to use DA in more effective ways.

Meanwhile, over the last couple of decades, there have been improvements in modular modeling approaches to integrate modeling systems, including the modular modeling system (MMS)⁴⁾, object-oriented hydrological modeling system (OHyMoS)⁵⁾, and interactive component modeling system (ICMS)⁶⁾. These kinds of modular approaches provide a flexible platform on which various models and tools are integrated. Thus, modelers can develop various types of models for problem objectives, available data, and spatio-temporal scales of application by organizing registered modules in diverse ways⁷⁾. OHyMoS is a hydrological modeling framework designed on the basis of object-oriented programming concepts. Using OHyMoS as a computational library, users can develop their own element models and easily build a total simulation system model for hydrological simulations⁶⁾. Unlike a process-based modeling framework, OHyMoS benefits from its object-oriented feature to represent hydrological processes flexibly without any change in the main OHyMoS library. However, OHyMoS, like most other modular modeling approaches, is designed on the basis of a deterministic approach. The original version of OHyMoS supports neither probabilistic simulation nor DA.

In this study, MPI-OHyMoS is developed to support stochastic hydrological simulations and DA, while adapting all object-oriented features of the original OHyMoS. Ensemble simulations are computed in parallel via MPI⁸⁾, which can take advantage of the computational power of an HPC system. Among DA methods, PF^{9),10),11),12),13),14)} is selected. The proposed framework is applied for uncertainty assessment of lumped and distributed hydrological models in synthetic and real experiment cases.

The paper is organized in the following way. Section 2 outlines the basic features of MPI-OHyMoS: PF, dual state-parameter updating, and parallelization for ensemble simulation. Section 3 illustrates the DA process in MPI-OHyMoS using a lumped hydrological model. In Section 4, MPI-OHyMoS is implemented for the uncertainty assessment of a distributed hydrological model in both synthetic and real

experiments. Section 5 summarizes the methodology and implementation results.

2. METHODOLOGY

As MPI-OHyMoS is a stochastic and interactive version of OHyMoS, the basic concept of OHyMoS is reviewed briefly. OHyMoS is constructed as a set of dynamic elements communicating with each other based on object-oriented programming⁷⁾.

As illustrated in Fig.1, it provides an operation module, including the common functions required in hydrological simulations, such as initialization of parameters and state variables, setting the computational time steps, and data exchange among element modules through input/output (I/O) ports. Through OHyMoS, users can easily develop their own hydrological modules by connecting them to other modules and transferring data using predefined ports in the system library. Detailed information about OHyMoS and its implementations can be found and downloaded at the developer's website¹⁵⁾.

In MPI-OHyMoS, hydrological modeling is implemented in the stochastic way. Fig.2 shows how model ensembles are interactively assimilated in MPI-OHyMoS. Each ensemble member, representing a probable projection based on different parameters and state variables, is implemented independently. When a new observation arrives, the likelihood of ensemble members is estimated. In the resampling step, the whole information of each ensemble is renewed, depending on its weight. In this way, ensembles can move to the regions of high conditional probability in each time step. Detailed features of MPI-OHyMoS are summarized below.

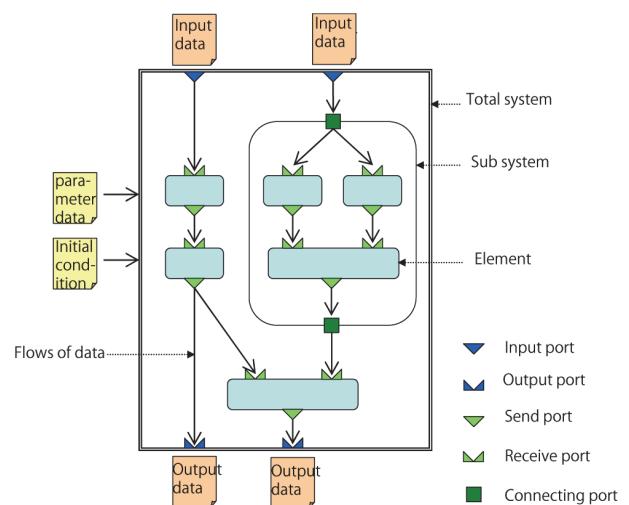


Fig.1 The structure of the original OHyMoS.

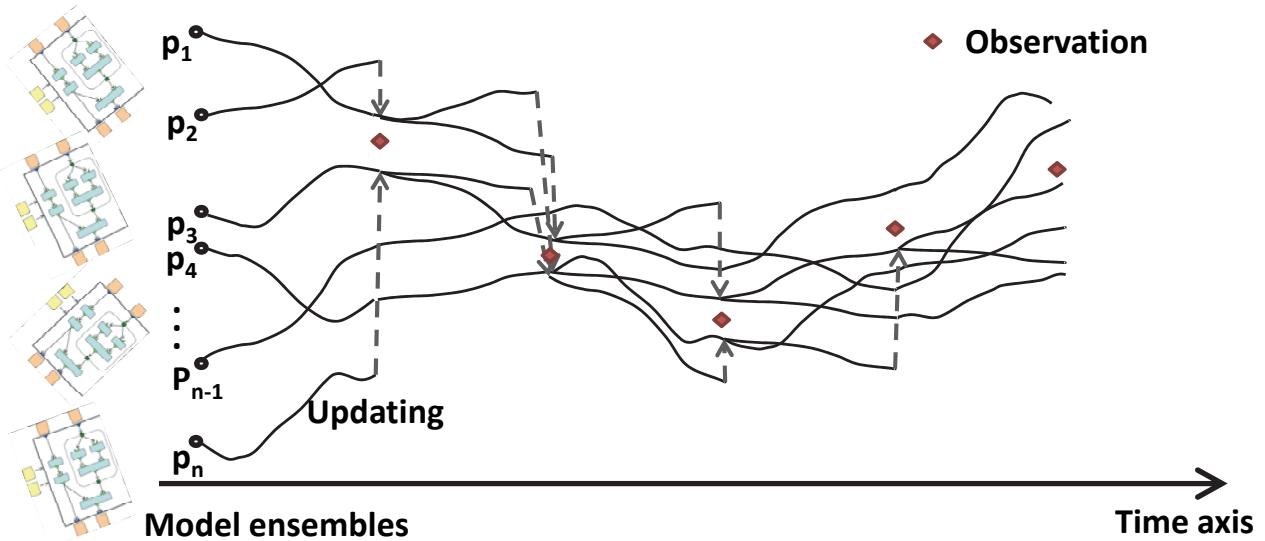


Fig.2 Concept of PF via MPI-OHyMoS.

(1) Particle filtering

PF^{9),10),11),12),13),14)} is a Bayesian learning process that has the capability to handle non-linear and non-Gaussian state-space models.

In hydrological modeling, a particle represents a hydrological model having different parameters and state variables. The key idea of PF is to use point mass representations of probability densities with associated weights¹⁰⁾. To fix the notations, let us introduce x_t , which represents all target states at time t . Then, the posterior filtered probability density $p(x_t | y_{1:t})$ can be approximated as

$$p(x_t | y_{1:t}) \approx \sum_{i=1}^n w_t^i \delta(x_t - x_t^i) \quad (1)$$

where x_t^i and w_t^i denote the i^{th} posterior state ("particle") and its weight, respectively, $\delta(\cdot)$ denotes the Dirac delta function, $y_{1:t}$ denotes all available measurements, and n is the number of particles. In typical circumstances, the recursive weight updating can be derived as follows:

$$w_t^i \propto w_{t-1}^i p(y_t | x_t^i) \quad (2)$$

where $p(y_t | x_t^i)$ is the likelihood of each particle x_t^i

A common problem of original PF is the degeneracy phenomenon: after a few iterations, all but one particle will have negligible weight. The degeneracy phenomenon can be reduced by performing the resampling step whenever a significant degeneracy is observed. Unlike Kalman filter-based methods, PF performs updating on particle weights instead of state variables¹⁶⁾, which has the advantage of reducing numerical instability, especially in process-based models.

In MPI-OHyMoS, a likelihood function is constructed as an independent element model to estimate the likelihood and weight of each particle, and can be combined with any element model and allows any user-defined density function.

(2) Dual state-parameter updating

As mentioned in previous studies^{17),18)}, hydrological models assume that the conditions under which their equations are derived are the same as those in the field, and that the specific characteristics of the dynamics of a process in each location can be reflected by model parameters. These assumptions are limiting because mathematics describing hydrological processes is poorly defined at the model scale and because there is inadequate data to describe the spatial variability of parameters, even in highly instrumented catchments. As a result, hydrological models use "effective" parameters that will provide good values of predicted states and outfluxes at the site of interest. Commonly, "effective" parameters are selected from the calibration process on the basis of very limited measurements: by extrapolation from applications at other sites, or by inference from a comparison of model outputs and observed responses at the site of interest¹⁸⁾. In this respect, there is no guarantee that such parameters calibrated from historical data and extrapolation are the optimum in the current prediction. On the other hand, model states are highly sensitive to the uncertainty of model parameters. Therefore, updating state variables based on inappropriate parameters may increase uncertainty in the prediction of hydrological models¹³⁾. The conventional calibration methods optimizing the objective function for parameters and observations also have significant flaws from the assumption that the mismatch between model and observation originates only from model deficiencies, without considering uncertainties of observation and input data. Recently, probabilistic parameter estimation methods considering various uncertainties (e.g. dual state-parameter estimation) have drawn attention in the hydrological research community due to their applicability to complex and real problems^{12),13)}.

In dual state-parameter updating, information on

different states and parameters is updated simultaneously during the resampling step. In the case of state updating, state variables, which are perturbed in the initial stage, are projected to the next time point by the state-space equation (e.g. hydrological models) and updated in the resampling step. However, in the case of parameter updating, we need additional constraints because there is usually no time-evolution information.

To handle inference with the unknown parameters θ , kernel smoothing^{19),20)} is adapted to improve parameter identifiability. The smooth kernel density can be a mixture of Gaussian densities as follows:

$$P(\theta_t | y_{1:t-1}) \sim \sum_{i=1}^n w_{t-1}^i N(\theta_t | m_{t-1}^i, h^2 V_{t-1}^\theta) \quad (3)$$

where h is the variance reduction parameter and V_{t-1}^θ is the variance of parameter particles at time $t-1$ before resampling. The kernel locations m_{t-1}^i are specified by a shrinkage rule forcing the particles to be closer to their mean:

$$m_{t-1}^i = a\theta_{t-1}^i + (1-a)\bar{\theta}_{t-1} \text{ with } a = \sqrt{1-h^2} \quad (4)$$

where $\bar{\theta}_{t-1}$ is the mean of the parameters at time $t-1$. It is found that a covariance matrix V_{t-1}^θ controlled by Eq. (4) does not increase over time^{19),21)}. MPI-OHyMoS provides the kernel smoothing scheme as a basic option of parameter updating. Statistics of parameters can be estimated in normal and log scales.

(3) Parallelized ensemble simulation

MPI, a parallel computing protocol for a distributed memory system which is common in HPC, is used for the parallelization of the ensemble simulation and DA in MPI-OHyMoS. Among variants of MPI libraries, open MPI⁸⁾ and Boost library²¹⁾ are selected in this study. Note that MPI is different from OpenMP, commonly used in hydrology for loop parallelization in a single model, whose applicability is limited in a shared memory system.

3. ILLUSTRATIVE EXAMPLE OF DA VIA MPI-OHYMOS

In this section, an example of DA is shown to illustrate the basic features and simulation processes of MPI-OHyMoS. A synthetic experiment is implemented using a linear reservoir model with an unknown initial condition and a model parameter.

(1) A lumped model

The linear reservoir model is based on the concept that a catchment behaves as a reservoir in which storage S is linearly related to outflow Q ²²⁾.

The process of a linear reservoir shown in Fig.3 can be described as:

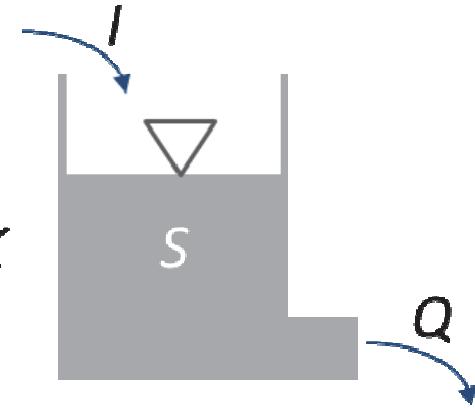


Fig.3 A linear reservoir model.

$$S = KQ \quad (5)$$

$$\frac{dS}{dt} = I - Q \quad (6)$$

where K , called the storage coefficient, is a rate constant, t is the time, and I is inflow. For the computational implementation, a parameter K and the initial state S_0 should be specified. The state S should be stored at each time step as the initial condition of the next time step.

(2) DA process via MPI-OHyMoS

Fig.4 illustrates the process of PF via MPI-OHyMoS using a linear reservoir model. In this case, the total system of MPI-OHyMoS consists of two elements: a linear reservoir model and a likelihood function. At each time step, the linear reservoir model calculates the states S and Q using Eqs. (5) and (6). Each ensemble has a different model parameter and initial storage perturbed by random noises. Thus, n ensembles result in different values of outflow Q . In the likelihood element, the likelihood of simulated outflow is estimated according to measurement outflow. Weights are then calculated by Eq. (2) and normalized such that $\sum_{i=1}^n w_k^i = 1$. Parameters and states are stored in the memory at each time step. In the resampling step, ensembles having large weights are duplicated to others. For example, in Fig.4, Ensemble 2 is duplicated to Ensemble 1 at the end of time step k . After the resampling step, Ensemble 1 has the same parameter and state as Ensemble 2. However, as random noises are added at each time step, Ensemble 1 and 2 result in different outflows at time step $k+1$. In this way, the state and parameter of ensembles are filtered at each time step. Estimated distributions of parameter and state represent posterior distribution. It is worth noting that duplication of both states and parameters in the resampling step is performed on the basis of computer memory without using any temporary files, which has high efficiency and stability especially for complex models (e.g. distributed hydrological models) having numerous spatially distributed states and parameters.

(3) Results of the synthetic experiment

The synthetic experiment is implemented using a linear reservoir model to illustrate basic features of DA in MPI-OHyMoS. Here we assume that the true values of parameters and initial states are unknown. Prior information on the ranges of parameters and initial conditions is shown in **Table 1**. A synthetic observation was produced by adding heteroscedastic errors $N(0, 0.05Q_{sim})$ to the simulation using the synthetic true. $0.05Q_{sim}$ means that the synthetic observation is perturbed by the noise, whose standard deviation is 5% of simulated discharge.

Fig.5 shows two hundred ensemble simulations without PF. As shown in **Fig.5(a)**, the values of parameter K do not change during simulation. Ensemble discharge varies within large uncertainty bounds, shown in **Fig.5(b)**. **Fig.6** shows ensemble simula-

tions with PF. A parameter and a state are updated using the synthetic observation every one-hour interval. The uncertainty bounds of parameter K and outflow are reduced sharply via PF, showing a good agreement with the synthetic true. Note that the simulation converges into the synthetic true quickly because the applied system is linear. A non-linear, non-Gaussian case follows in the next section.

Table 1 Information on parameter and initial state.

Parameter and initial state	Synthetic true values	Ranges of parameter/initial state for ensemble simulation
K (hr)	4	2~10
S_1^{ini} (m^3)	15,000	5,000~20,000

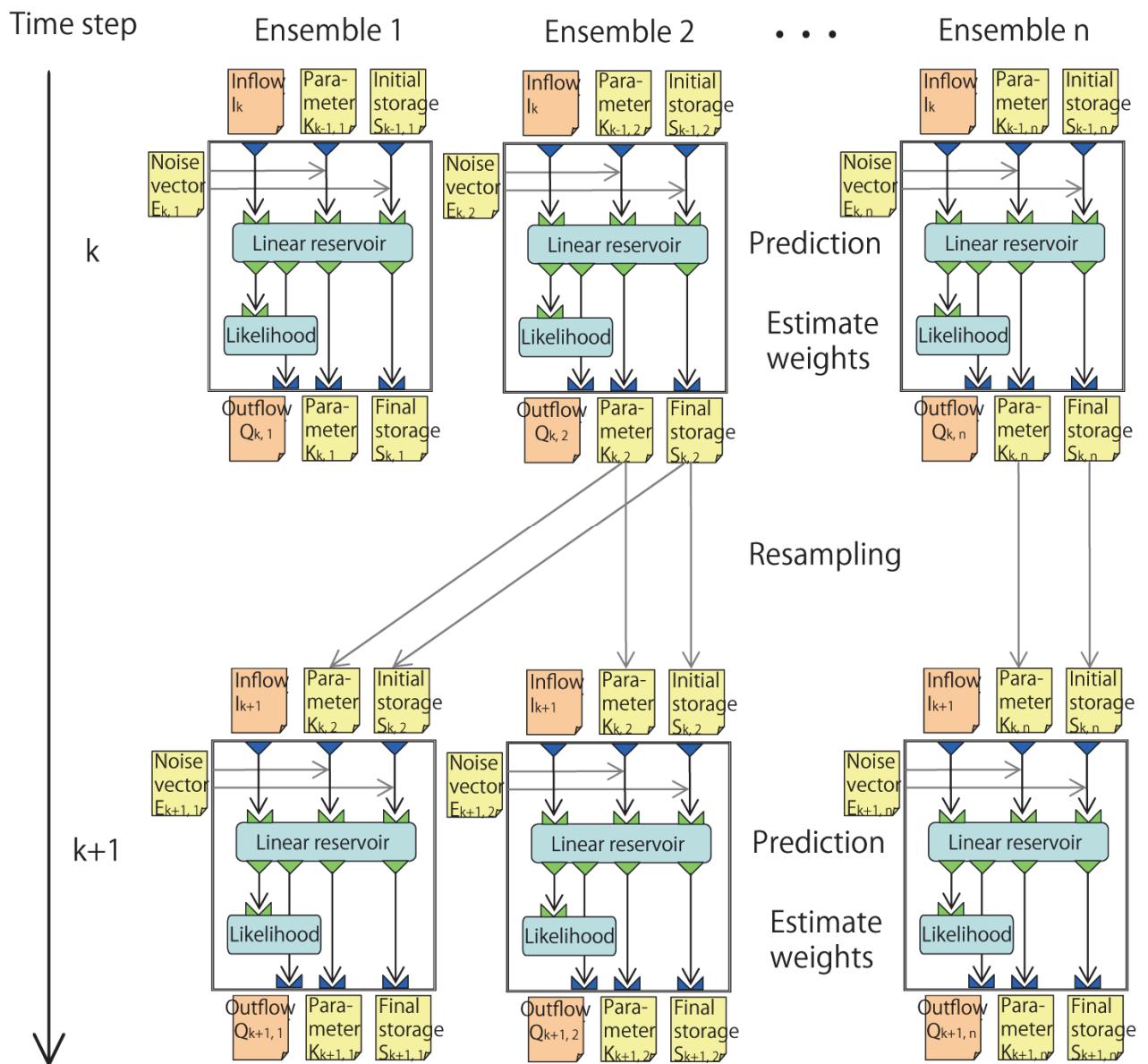


Fig.4 DA process of a linear reservoir model via MPI-OHyMoS.

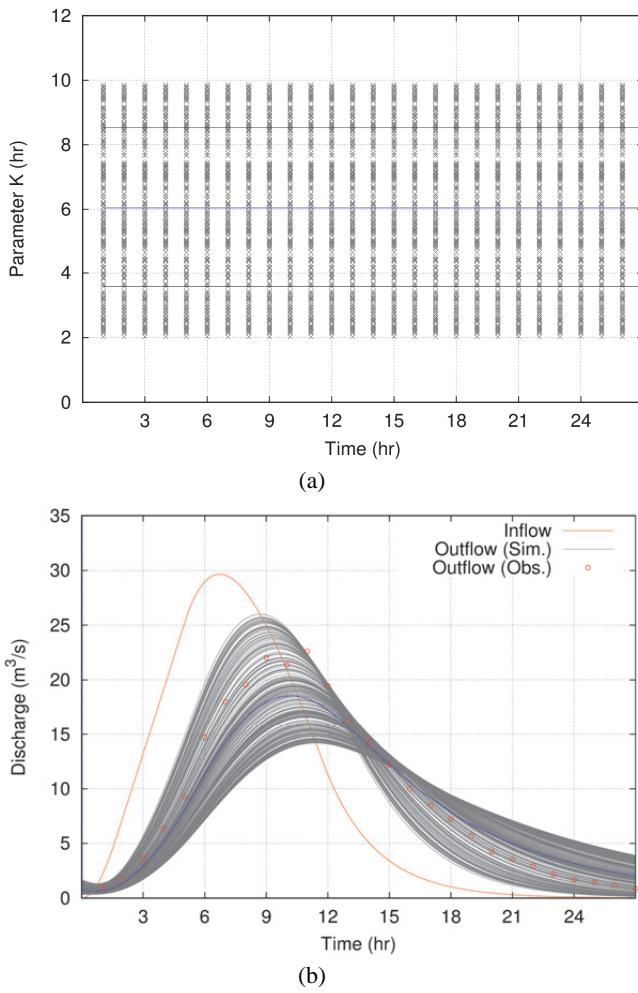


Fig.5 Parallel simulations of the linear reservoir model without PF by 200 ensembles. (a) Traces of parameter K . (b) Traces of inflow and outflow. Blue and black lines represent the mean and 60% confidence intervals of parameter ensembles, respectively. The orange line represents inflow. Red dots represent synthetic observation. Grey lines represent 200 ensembles. The blue line represents the mean of discharge ensembles.

4. UNCERTAINTY ASSESSMENT USING A DISTRIBUTED HYDROLOGICAL MODEL

Synthetic and real experiments are implemented for uncertainty assessment of a fully distributed hydrological model^{23),24)} to illustrate the applicability of MPI-OHyMoS.

(1) Study area

The study area is the Maruyama River catchment in Japan, with an area of about 909 km^2 . **Fig.7** shows the streamflow gauging locations and rainfall stations. Streamflow measurement at Fuichiba is used for DA in both synthetic and real experiments. Land use consists of 37% forest, 10% savannas, and 53% crop land and natural vegetation²⁵⁾.

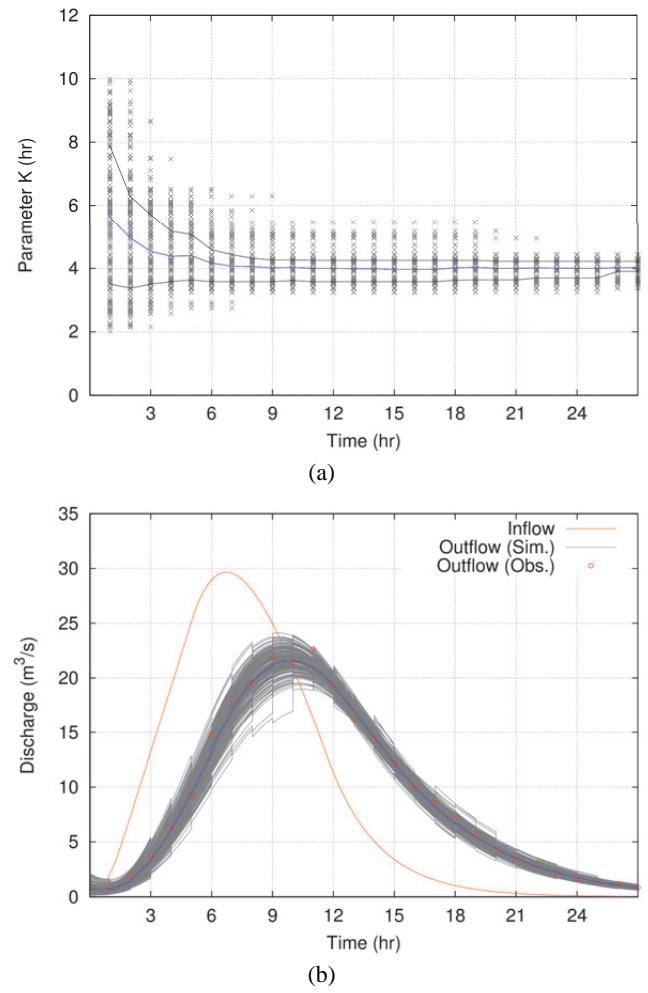


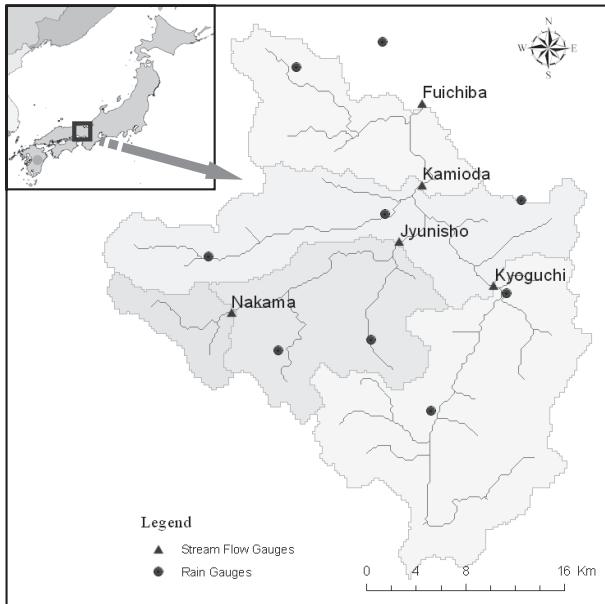
Fig.6 Parallel simulations of the linear reservoir model with PF by 200 ensembles. (a) Traces of parameter K . (b) Traces of inflow and outflow. Blue and black lines represent the mean and 60% confidence intervals of parameter ensembles, respectively. The orange line represents inflow. Red dots represent synthetic observation. Grey lines represent 200 ensembles. The blue line represents the mean of discharge ensembles.

(2) A distributed hydrological model

We construct a probabilistic distributed hydrological model for the Maruyama River catchment based on three element modules: a hillslope runoff generation module, a river routing module, and a likelihood function.

The hillslope and river routing modules were developed as elements of a deterministic distributed hydrological model using the kinematic wave theory in a previous study²⁶⁾. In this model, it is considered that the catchment consists of a number of rectangular slope elements which drain to the steepest gradient of its surroundings, as shown in **Fig.8**.

Fig.9 shows the flow process and the stage discharge relationship used in the hillslope model given in Eq. (7).



$$q = \begin{cases} V_c d_c (h/d_c)^\beta & 0 \leq h \leq d_c \\ V_c d_c + V_a (h-d_c) & d_c < h \leq d_s \\ V_c d_c + V_a (h-d_c) + \alpha (h-d_s)^m & d_s < h \end{cases} \quad (7)$$

$$\frac{\partial h}{\partial t} + \frac{\partial q}{\partial x} = r(t) \quad (8)$$

where t is the time, x is space, and $r(t)$ is the rainfall intensity to the slope element. The discharge per unit width, q , is estimated by Eq. (7) combined with the continuity equation, Eq. (8), where $V_c = k_c i$, $V_a = k_a i$, $k_c = k_a / \beta$, $\alpha = \sqrt{i/n_{slope}}$, $m=5/3$, i is the slope unit gradient, k_c is the hydraulic conductivity of the capillary soil layer, k_a is the hydraulic conductivity of the non-capillary soil layer, n_{slope} is the roughness coefficient of the hillslope component, h is the water stage, V_a and V_c are flow rates, and d_c and d_s are soil depth in the capillary pore and non-capillary pore, respectively. β is a parameter.

For channel routing, we use a one-dimensional kinematic wave equation as follows:

$$\frac{\partial h_{river}}{\partial t} + \frac{\partial q_{river}}{\partial x} = q_L \quad (9)$$

$$q_{river} = \alpha_{river} h_{river}^m \quad (10)$$

where h_{river} is the channel water depth, q_{river} is the channel discharge per unit width, q_L is lateral inflow, $\alpha_{river} = \sqrt{i/n_{river}}$, $m=5/3$, and n_{river} is the roughness coefficient of the river component.

(3) Model setup for DA

The measurement error of the discharge is assumed to be a Gaussian distribution, $N(0, \sigma_{obs}^2)$. The standard deviation of the measurement error is chosen as:

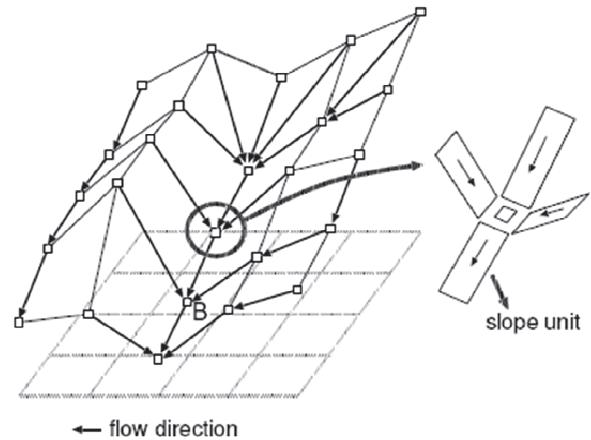


Fig.8 Conceptualization of spatial flow movement.

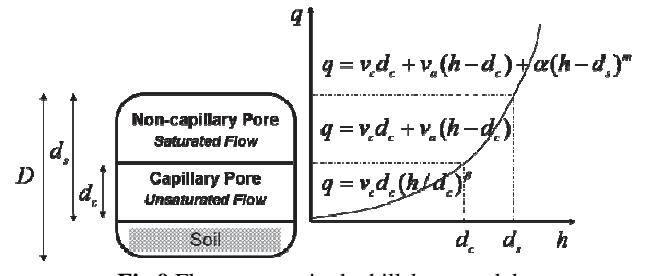


Fig.9 Flow process in the hillslope model.

$$\sigma_{obs_i} = \alpha_{obs} y_i + \beta_{obs} \quad (11)$$

In Eq. (11), y_i is observed discharge at time t , while α_{obs} and β_{obs} are parameters representing uncertainty of observations.

The process noise is generated by a Gaussian distribution, $\varepsilon_{sim} \sim N(0, \sigma_{sim}^2)$. Then, the state variables of slope and river component in each grid are perturbed at each observation time step in a multiplicative manner as follows:

$$\hat{x}_t^j = (1 + \varepsilon_{sim}) x_t^j \quad (12)$$

where x_t^j and \hat{x}_t^j are state variables before and after perturbation, respectively. It is worth noting that the process noise is essential for considering model uncertainties; however, it should be chosen not to make posterior distribution over-dispersed. The standard deviation of the process error σ_{sim} can be selected through sensitivity analysis for open loops (without filtering) and PF.

Among the various parameters in the distributed hydrological model, four are selected for DA. These parameters are d_c , k_a , n_{slope} , and n_{river} .

The model setup uses a 250 m grid resolution. The simulation time steps are six hundred seconds for the hillslope and twenty seconds for river routing. Ensembles are updated hourly by streamflow measurements at the Fuichiba gauging station. We use hourly observed rainfall from nine observation stations organized by the Ministry of Land, Infrastructure, Transport, and Tourism of Japan. Selected flood

events are shown in **Table 2**. Event 3 is used for the synthetic experiment, while all events are used for the real experiment.

(4) Synthetic experiment

The synthetic experiment is implemented using a distributed hydrological model to demonstrate the applicability of MPI-OHyMoS to missing data problems in complex cases and to assess the identifiability of model parameters.

The basic procedures of the synthetic experiment in the distributed hydrological model are the same in the lumped model case. Synthetic observation of streamflow is calculated by synthetic true values of parameters, shown in **Table 3**, adding a small Gaussian noise. For probabilistic modeling, the initial condition of states is perturbed by using noise from the uniform distribution, $\varepsilon_{ini} \sim U(0, \sigma_{ini}^2)$, in a multiplicative manner, shown in Eq. (12). The applied value of σ_{ini} is 0.1. The standard deviation of the process noise of states σ_{sim} is set as 0.01, which accounts for the predictive uncertainty of state variables. The process noise of parameters is controlled by kernel smoothing using the information of ensemble mean and variance at the previous time step shown in Eqs. (3) and (4).

The statistics of parameter k_a are estimated in the log scale to cover wider ranges of uncertainty bounds. Parameters of observation error in Eq. (11), α_{obs} and β_{obs} , are set as 0.05 and 5 (m^3/s), respectively. The size of the ensembles is 1,000. It is worth noting that the proper ensemble size of parameter estimation is quite different from that of state-only updating. It has been found that stabilized prediction accuracy is achieved for estimation of 6 parameters in a lumped model with 1,000 ensembles¹³⁾, whereas state-only updating can be performed with a limited number of ensembles using advanced PF¹⁴⁾. In this study, the size of ensembles is chosen considering the previous studies^{13),14)} and the capacity of computing resources.

The synthetic experiment is implemented in two separate stages. In the preliminary stage, the initial distribution of parameters is selected to cover ranges adopted in previous studies^{7),27)}, while results of the preliminary stage are used as prior information in the second stage simulation. Values of synthetic true and uncertainty ranges of parameters at each stage are shown in **Table 3**.

Table 2 Details of selected flood events in Fuichiba.

Flood Event	Date	Peak flow (m^3/s)	Initial flow of the event (m^3/s)
Event 1	10-20 Sep. 2001	715	42
Event 2	7-10 Sep. 2002	293	6
Event 3	19-24 Oct. 2004	4,782	32

Results of the preliminary stage of the synthetic experiment are shown in **Fig.10**. Simulated streamflow, which is a one-step-ahead prediction, shows good conformity with the synthetic observation in terms of ensemble mean and distributions. Uncertainty of parameters lasts before the flood event as in the initial distributions and reduces sharply around the flood peak. Dual state-parameter updating via PF results in a reasonable estimation of parameters to cover the synthetic true within their posterior distributions. Results of the second stage are shown in **Fig.11**. With reduced uncertainty ranges, traces of parameters show stable variations converging into the synthetic true according to sequential updating. Stabilized spreads of the parameters d_c , n_{river} , and n_{slope} , excluding k_a in the preliminary stage (**Fig.10**), mean that posterior distributions of parameters are properly estimated in this stage, whereas k_a , which is generated from wide uncertainty bounds in the preliminary stage, still needs extra assimilation periods in the second stage (**Fig.11**).

The mean and confidence intervals of the estimated parameters in the second stage are shown in **Table 4**. Identifiability of parameters is increased in the second stage with reduced initial probabilistic distributions. Considering parameter identifiability from the normalized confidence width, n_{slope} shows low identifiability compared to others, excluding k_a , which is estimated in the log scale.

Despite dispersed distributions of parameters in the low flow conditions around the initial ranges of **Figs.10-11**, discharge hydrographs show narrow probabilistic bands. Equifinality and sensitivity can be considered main reasons. Different combinations of parameter values may produce similar discharge hydrographs in low flow conditions (equifinality).

Table 3 Information on parameters in the synthetic experiment.

Parameters	Synthetic true values	Initial ranges of parameters for ensemble simulation	
		Preliminary stage	2nd stage
k_a (m/s)	0.006	0.001~0.1	0.001~0.01
d_c (m)	0.4	0.1~0.5	0.36~0.42
n_{river} ($m^{-1/3}s$)	0.038	0.015~0.055	0.035~0.043
n_{slope} ($m^{-1/3}s$)	0.33	0.15~0.55	0.3~0.4

Table 4 Mean and confidence intervals of estimated parameters in the second stage of the synthetic experiment.

Parameters	Ensemble mean (A)	60% confidence interval (B)	Normalized confidence width (dB/A)
d_c (m)	0.394	0.385~0.404	0.048
k_a (m/s)	0.0052	0.0048~0.0055	0.134
n_{river} ($m^{-1/3}s$)	0.0385	0.0369~0.040	0.081
n_{slope} ($m^{-1/3}s$)	0.343	0.320~0.364	0.128

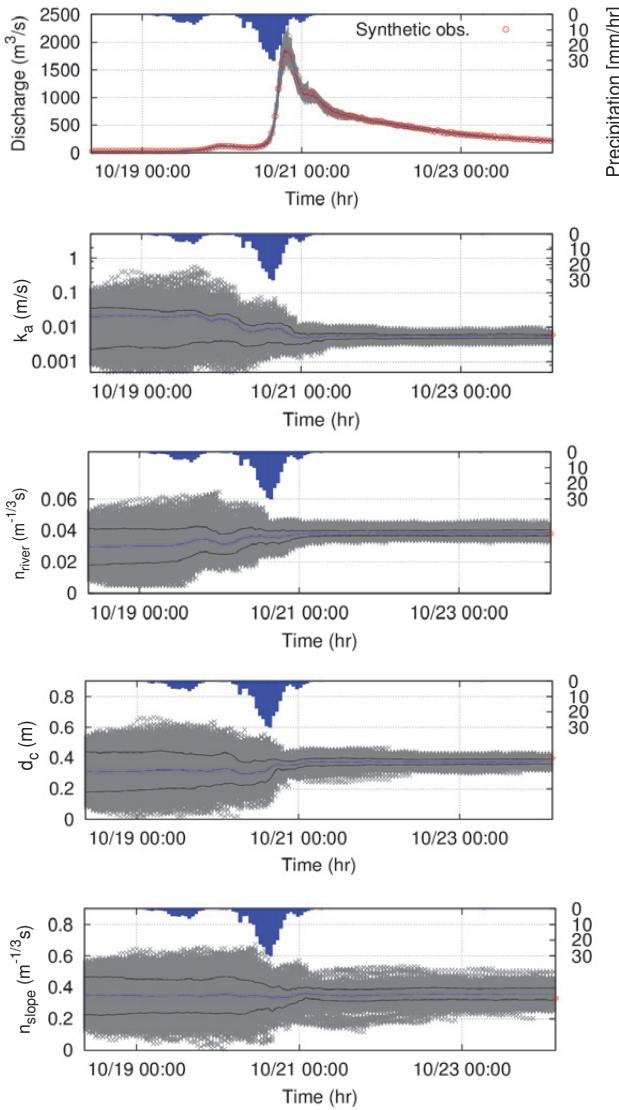


Fig.10 Parallel simulations of the distributed hydrological model with PF in the synthetic experiment's preliminary stage (Event 3). Red dots and squares represent synthetic true observation and parameters, respectively. Grey lines and dots represent streamflow and parameter ensembles, respectively. Blue and black lines represent means and 60% confidence intervals of ensembles, respectively.

However, from their definition and behaviour, each parameter has a different sensitivity for different flow conditions. For example, soil depth d_c and roughness coefficients (n_{river} , n_{slope}) have critical effects on high flow, while k_a has an influence on the entire hydrograph. There are asymptotic bounds where distributions cannot be narrower, because PF considers uncertainties from different sources such as input, models and observations simultaneously.

(5) Real experiment

The real experiment is conducted in two stages using Event 1. The applicability of two-stage estimation is verified in simulations of Events 2 and 3, which are conducted using results of the preliminary

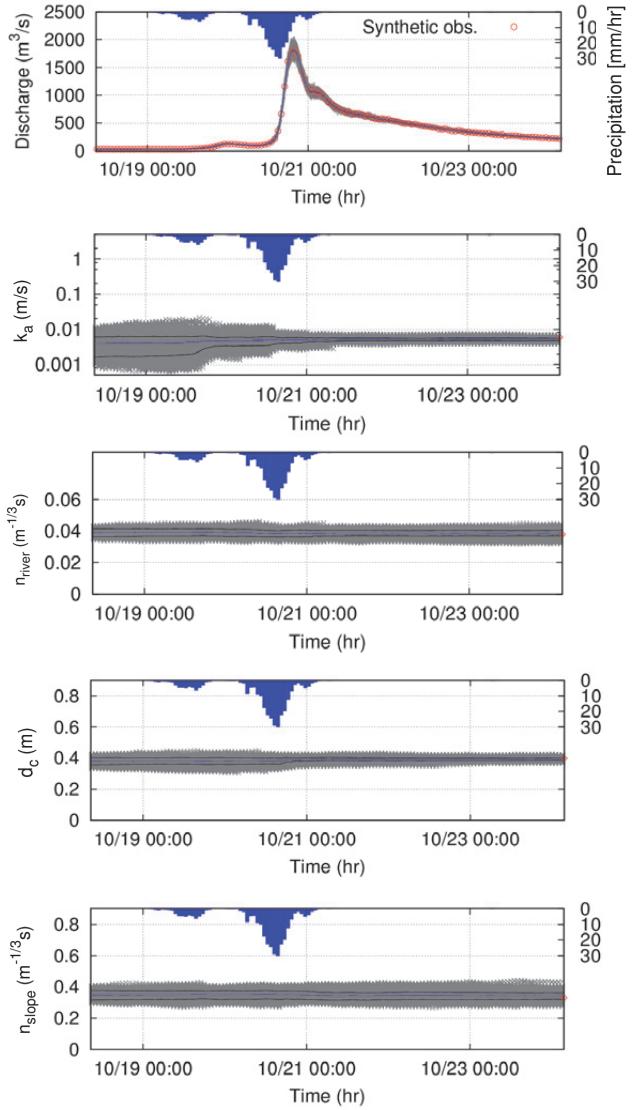


Fig.11 Parallel simulations of the distributed hydrological model with PF in the synthetic experiment's second stage (Event 3). Red dots and squares represent synthetic true observation and parameters, respectively. Grey lines and dots represent streamflow and parameter ensembles, respectively. Blue and black lines represent means and 60% confidence intervals of ensembles, respectively.

stage of Event 1 as initial ranges of parameters. For parameter updating, k_a , d_c , and n_{river} are selected, excluding n_{slope} because the identifiability of n_{slope} is found to be relatively lower in the synthetic experiment.

In the preliminary stage, the initial ranges of parameters k_a , d_c , and n_{river} are the same as in the synthetic experiment shown in **Table 3**, while the value of n_{slope} is set as 0.3, selected from the ranges in previous studies^{7),27)}. In the real experiment, the uncertainty of process and observation is assumed to be larger than that in the synthetic case. The standard deviation of the process noise of states σ_{sim} is set as 0.05 and the parameters of observation error, α_{obs} and β_{obs} , are set as 0.1 and 20, respectively. The standard deviation of initial states σ_{ini} is 0.3.

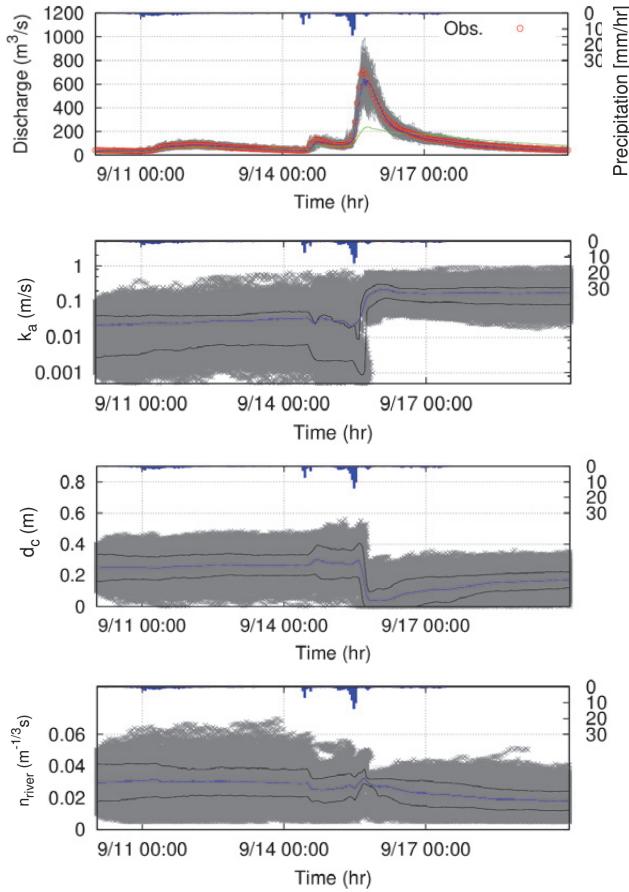


Fig.12 Parallel simulations of the distributed hydrological model with PF in the preliminary stage of the real experiment (Event 1). Red dots represent synthetic observation. Grey lines and dots represent streamflow and parameter ensembles, respectively. Blue and black lines represent means and 60% confidence intervals of ensembles, respectively. Green lines represent streamflow of deterministic modeling with parameters using the mean of the preliminary stage.

Simulation results of preliminary and second stages of Event 1 are shown in **Figs.12~13**. As shown in **Fig.12**, the uncertainty of parameters sharply reduces around the flood peak. Distribution of parameter k_a rapidly becomes narrow around the flood peak, while distribution of n_{river} shows smoother movement. In the second stage, initial distribution of parameters is selected to cover about 60% confidence intervals of the preliminary stage. As shown in **Fig.13**, there is no rapid movement of parameter distribution during the second-stage simulation leading to narrower posterior distribution, compared to the preliminary stage, whose estimated values are shown in **Table 5**.

Streamflow prediction via PF is compared with the deterministic modeling case using the mean of the initial distribution or the posterior of the preliminary stage in **Figs.12~13**. It should also be noted that deterministic streamflow simulation is improved in the second stage using parameters estimated via PF.

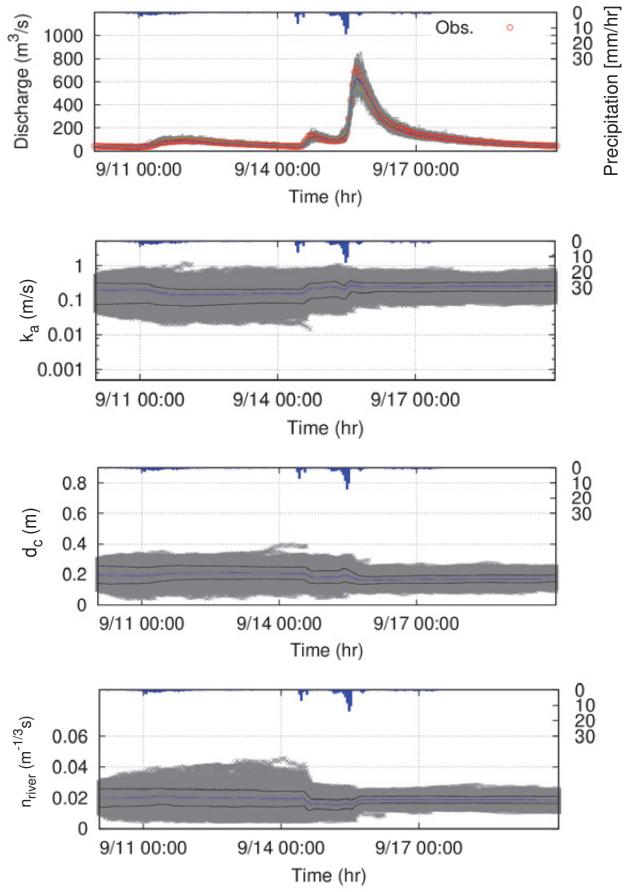


Fig.13 Parallel simulations of the distributed hydrological model with PF in the second stage of the real experiment (Event 1). Red dots represent synthetic observation. Grey lines and dots represent streamflow and parameter ensembles, respectively. Blue and black lines represent means and 60% confidence intervals of ensembles, respectively. Green lines represent streamflow of deterministic modeling with parameters using the mean of the preliminary stage.

Table 5 Mean and confidence intervals of estimated parameters in the second stage of the real experiment (Event 1).

Parameters	Ensemble mean	Confidence interval (60%)
k_a (m/s)	0.262	0.182~0.335
d_c (m)	0.174	0.152~0.196
n_{river} ($m^{-1/3}s$)	0.019	0.016~0.021

Simulation results of Events 2 and 3 are shown in **Figs.14~15**. The initial distributions of the parameters are adopted from the posterior estimated in Event 1 to assess the applicability of the parameters for different flood events. In the results of both cases, the traces of parameter distributions show stable movement reaching narrow posteriors within the initial bounds. Despite large uncertainty bounds of the initial distribution, parameter k_a shows high identifiability in Events 2 and 3. One-step-ahead prediction of streamflow also results in reliable discharge hydrographs in both cases.

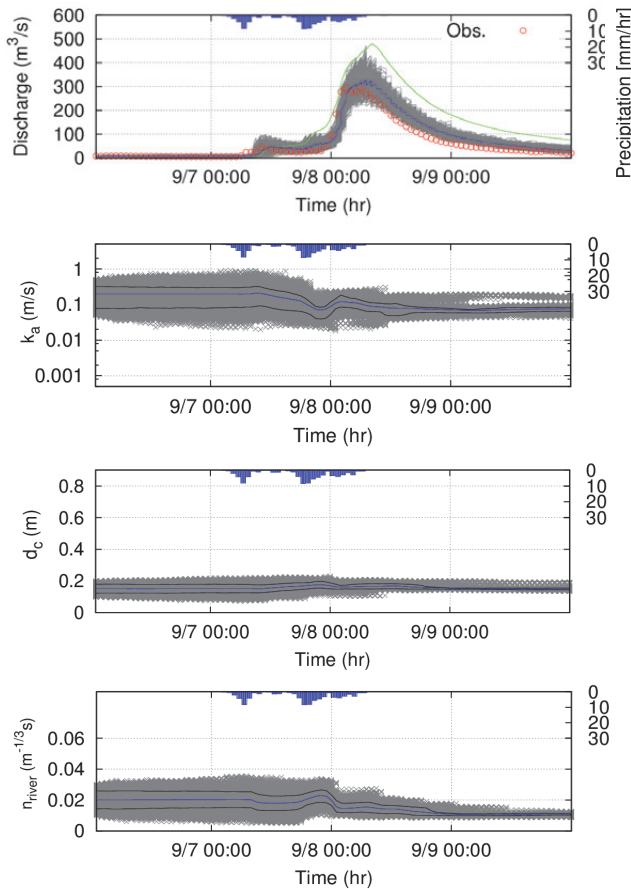


Fig.14 Parallel simulations of the distributed hydrological model with PF in the real experiment (Event 2). Red dots represent synthetic observation. Grey lines and dots represent streamflow and parameter ensembles, respectively. Blue and black lines represent means and 60% confidence intervals of ensembles, respectively. Green lines represent streamflow of deterministic modeling with parameters using results of Event 1.

Note that the magnitudes of the observed flood peaks of Events 2 and 3 are quite different. In keeping up with previous studies^{12),13)}, it is shown that the spread among ensemble members for each parameter drastically decreases around the flood peak for both synthetic and real experiments. The reason for the decrease can be attributed to the role of high flow observations in the updating step. In PF, new discharge observations are assimilated into the hydrological model at every time step. In this sequential procedure, limited combinations of parameters are selected in the high flow conditions, while most of the parameters are accepted in the low flow conditions.

Model performance is summarized in **Table 6** using two indices: Nash-Sutcliffe efficiency (NSE) and root mean square error (RMSE). Statistics show an improvement in the model performance via PF in all events compared to deterministic modeling.

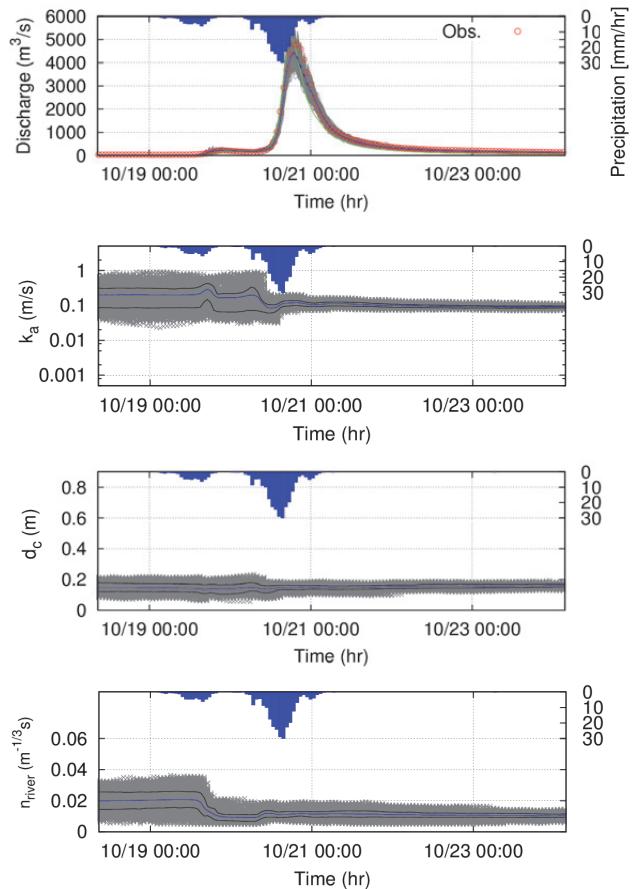


Fig.15 Parallel simulations of the distributed hydrological model with PF in the real experiment (Event 3). Red dots represent synthetic observation. Grey lines and dots represent streamflow and parameter ensembles, respectively. Blue and black lines represent means and 60% confidence intervals of ensembles, respectively. Green lines represent streamflow of deterministic modeling with parameters using results of Event 1.

Parameter distributions estimated by PF at Event 1 result in good performance in Event 3, whose peak flood is about six times higher than that of Event 1. In the case of deterministic modeling, parameters estimated in Event 1 via PF are applied to Events 2 and 3. Deterministic modeling presents improved performance for Event 3, demonstrating transferability of the parameters for an un-experienced high flood. However, application to a smaller flood (Event 2) shows limited performance. Due to uncertainties coming from models and observations, optimal parameters may change according to the magnitude of flood events and initial conditions. The results of deterministic modeling show that parameters estimated at large events (Event 1) may not be appropriate for small events (Event 2) or vice versa. Such a situation is found frequently in hydrological modeling. However, PF with dual state-parameter updating shows improved performance for all cases.

Table 6 Summary of model performance in the real experiment.

	Deterministic		PF	
	NSE	RMSE ($m^3 s^{-1}$)	NSE	RMSE ($m^3 s^{-1}$)
Event 1	0.96	23	0.98	16
Event 2	-0.40	93	0.81	33
Event 3	0.94	217	0.98	117

In the real experiment, there are other uncertainty sources which are not considered in DA, such as uncertainty of rainfall. That might be one of the main reasons that the real experiment shows reduced accuracy with dispersed parameter distributions. However, the real experiment can be improved in combination with auxiliary observations and advanced noise modeling such as remote-sensed soil moisture and the rainfall ensemble generator^{1),28)}.

5. CONCLUSIONS

MPI-OHyMoS was developed as a hydrological software framework for stochastic simulation and DA. This flexible framework provided PF, dual state-parameter updating, and kernel smoothing to consider various sources of uncertainty in hydrological modeling. Ensemble simulation was parallelized by MPI, taking advantage of an HPC system. The applicability of MPI-OHyMoS was demonstrated using different hydrological models, such as lumped and distributed models.

Structure and implementation processes of PF via MPI-OHyMoS were illustrated through the synthetic experiment of a simple lumped model.

Synthetic experiment cases of a distributed hydrological model showed that the dual state-parameter updating scheme of MPI-OHyMoS could be conducted properly for missing data problems. The different parameter identifiability was found in the synthetic experiment of a distributed hydrological model. The roughness coefficient of the slope component showed diffusive probabilistic distribution. However, it is worth noting that identifiability of parameters can be improved with multi-site observations in PF.

In the real experiment of a distributed hydrological model, simulated discharge via PF showed good conformity with the observation. Uncertainty bounds of ensembles were also reduced significantly around the flood peak. The assimilated results could be used to improve streamflow forecasting. Parameters estimated by PF contributed to improvement of deterministic modeling especially in extreme cases.

Despite their potential to estimate and mitigate uncertainty for non-linear, non-Gaussian models, implementation of sequential DA, including PF, has

been limited due to a lack of modeling frameworks. MPI-OHyMoS is expected to make it easy to build a stochastic hydrological model and support DA as a general modeling framework. In the future, we plan to improve MPI-OHyMoS in terms of parameter estimation methods and flexible assimilation control.

ACKNOWLEDGMENT: We thank Dr. Y. Ichikawa for valuable comments and discussions. The research work was supported by the KAKENHI, Grant-in-Aid for Scientific Research (A) 20246082 provided by MEXT and partly by Collaborative Research Program for Large-Scale Computation of ACCMS and IIMC, Kyoto University. The software framework developed in this paper can be obtained from the authors by request via email.

REFERENCES

- 1) Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P. : Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, Vol. 16, pp. 3863–3887, 2012.
- 2) OpenDA : <http://www.opendata.org/joomla/index.php> retrieved 15 Jan., 2013.
- 3) Karssenberg, D., Schmitz, O., Salamon, P., de Jong, K. and Bierkens, M. F. P. : A software framework for construction of process-based stochastic spatio-temporal models and data assimilation, *Environmental Modeling and Software*, Vol. 25, pp. 489-502, 2010.
- 4) Leavesley, G. H., Markstrom, S. L., Restrepo, P. J. and Viger, R. J. : A modular approach to addressing model design, scale and parameter estimation issues in distributed hydrological modeling, *Hydrological Processes*, Vol. 16, pp. 173-187, 2002.
- 5) Ichikawa, Y., Tachikawa, Y., Takara, K. and Shiiba, M. : Object-oriented hydrological modeling system, *Proceedings of 4th International Conference Hydroinformatics 2000*, Iowa, USA, 2000.
- 6) Reed, M., Cuddy, S. M. and Rizzoli, A. E. : A framework for modeling multiple resource management issues - an open modeling approach, *Environmental Modeling and Software*, Vol. 14, pp. 503-509, 1999.
- 7) Lee, G., Kim, S., Jung, K. and Tachikawa, Y. : Development of a large basin rainfall-runoff modeling system using the object-oriented hydrologic modeling system (OHMoS), *KSCE Journal of Civil Engineering*, Vol. 15, pp. 595-606, 2011.
- 8) Open MPI : <http://www.open-mpi.org/> retrieved 15 Jan., 2013.
- 9) Gordon, N. J., Salmond, D. J. and Smith, A. F. M. : Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Proc. Inst. Electr. Eng.*, Vol. 140, pp. 107-113, 1993.
- 10) Ristic, B., Arulampalam, S. and Gordon, N. : *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004.
- 11) Andrieu, C., Doucet, A. and Holenstein, R. : Particle Markov chain Monte Carlo methods, *J. R. Statist. Soc. B*, Vol. 72, Part 3, pp. 269–342, 2010.
- 12) Moradkhani, H., Hsu, K.-L., Gupta, H. and Sorooshian, S. : Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter, *Water Resour. Res.*, Vol. 41, W05012, doi: 10.1029/

- 2004WR003604, 2005.
- 13) Noh, S. J., Tachikawa, Y., Shiiba, M. and Kim, S. : Dual state-parameter updating scheme on a conceptual hydrologic model using sequential Monte Carlo filters, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 55, pp. 1-6, 2011.
 - 14) Noh, S. J., Tachikawa, Y., Shiiba, M. and Kim, S. : Applying sequential Monte Carlo methods into a distributed hydrologic model: lagged particle filtering approach with regularization, *Hydrol. Earth Syst. Sci.*, Vol. 15, pp. 3237-3251, 2011.
 - 15) Hydrology and Water Resources Research Lab., Kyoto University : <http://hywr.kuciv.kyoto-u.ac.jp/ohymos/> retrieved 15 Jan., 2013.
 - 16) Liu, Y. and Gupta, H. V. : Uncertainty in hydrologic modeling: toward an integrated data assimilation framework, *Water Resour. Res.*, Vol. 43, W07401, doi: 10.1029/2006WR005756, 2007.
 - 17) Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. and Ye, M. : Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, Vol. 48, W08301, doi:10.1029/2011WR011044, 2012.
 - 18) Beven, K. J.: *Environmental Modelling: An Uncertain Future?*, Routledge, 2009.
 - 19) Liu, J. and West, M. : Combined parameter and state estimation in simulation-based filtering, *Sequential Monte Carlo Methods in Practice*, Doucet, A., de Freitas, N. and Gordon, N. eds., Springer-Verlag, pp. 197-223, 2001.
 - 20) Moradkhani, H., Sorooshian, S., Gupta, H. V. and Houser, P. R. : Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, Vol. 28, pp. 135-147, 2005.
 - 21) Boost C++ Library : <http://www.boost.org/> retrieved 15 Jan., 2013.
 - 22) US Army Corps of Engineers Hydrologic Engineering Center : Hydrographs by single linear reservoir model, 1980.
 - 23) Takasao, T. and Shiiba, M. : Incorporation of the effect of concentration of flow into the kinematic wave equations and its applications to runoff system lumping, *Journal of Hydrology*, Vol. 102, pp. 301-322, 1988.
 - 24) Ichikawa, Y., Murakami, M., Tachikawa, T. and Shiiba, M. : Development of a basin runoff simulation system based on a new digital topographic model, *J. Hydraulic, Coastal and Environ. Engng., JSCE*, Vol. 691, pp. 43-52, 2001.
 - 25) Hunukumbura, P.B., Tachikawa, Y. and Takara, K. : Improvement of internal behavior of a distributed hydrological model, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 52, pp. 49-54, 2008.
 - 26) Tachikawa, Y., Nagatani, G. and Takara, K. : Development of stage-discharge relationship equation incorporating saturated-unsaturated flow mechanism, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 48, pp. 7-12, 2004.
 - 27) Kim, S., Tachikawa, Y., Lee, G. and Takara, K. : Prediction of the largest ever flood: case study on typhoon Rusa in 2002 at the Gamcheon basin, Korea, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 52, pp. 67-72, 2008.
 - 28) Rakovec, O., Hazenberg, P., Torfs, P. J. J. F., Weerts, A. H. and Uijlenhoet, R. : Generating spatial precipitation ensembles: impact of temporal correlation structure, *Hydrol. Earth Syst. Sci.*, Vol. 16, pp. 3419-3434, 2012.

(Received May 30, 2012)