

# An Assimilation Simulation Approach to Suspended Sediment Concentration in Inland Lakes Using Hybrid Perturbation Generation Method

Fei Guo (✉ [guofei@njnu.edu.cn](mailto:guofei@njnu.edu.cn))

Nanjing Normal University School of Geography <https://orcid.org/0000-0002-5115-7249>

Jingjia Zhang

Nanjing Normal University School of Geography

A-xing Zhu

University of Wisconsin-Madison Department of Geography

Zhuo Zhang

Nanjing Normal University School of Geography

Hong Zhang

Nanjing Normal University School of Geography

---

## Research Article

**Keywords:** Hydrodynamic, perturbation, RMSE(root mean square error)

**Posted Date:** March 11th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-307235/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Water Resources Management on April 10th, 2021. See the published version at <https://doi.org/10.1007/s11269-021-02827-1>.

## **An assimilation simulation approach to suspended sediment concentration in inland lakes using hybrid perturbation generation method**

**Abstract:** Suspended sediments, as one of the most important factors affecting the water environment of inland lakes, are closely related to the various pollutants migration and interaction. Thus, the simulation and prediction of suspended sediment concentration is important. Existing studies show that the prediction accuracy of suspended sediment concentration can be effectively predicted based on assimilation methods coupled with hydrodynamic models. However, in the process of assimilation of hydrological simulation, the existing perturbation generation methods consider that the perturbation error is a random Gaussian distribution, which does not consider the spatial variation characteristics of errors. In this paper, a new method to generate the perturbation field for assimilation simulation was proposed. This method uses hybrid error to generate perturbation field for assimilation simulation instead of using random error. The proposed approach was validated through its application to assimilation simulation of suspended sediment concentration of Taihu Lake in China, and five assimilation experiments were conducted. The proposed method was compared with existing methods for perturbation field generation. After three days and 72 time steps of assimilation simulation based on hybrid perturbation generation, we found that the proposed assimilation method provided results that were more consistent with buoy-measured data. The accuracy of the two assimilation methods based on hybrid perturbation is improved. Compared with the assimilation method based on random perturbation, the mean values of RMSE (root mean square error) decreased from 9.56 to 8.70 and from 9.55 to 8.68, respectively. The results show that the proposed hybrid perturbation generation method has higher simulation accuracy than other methods. This study shows that the method is effective and provides a new idea for the assimilation simulation of suspended sediment concentration in inland lakes.

### **1. Introduction**

Due to the increasing anthropogenic pressure, the expansion of urban areas and the unreasonable exploitation, the deterioration of inland lake water environments is

becoming more and more serious (Chen et al., 2006; Friese et al., 2010; Wu et al., 2012; Ghale et al., 2018). The main influencing factors of inland lake water environments include suspended sediments, plankton and dissolved organic matter. As one of the most important factors affecting the water environment, suspended sediments are the carriers of various pollutant migration and interaction (Wu et al., 2012; Wei et al., 2019; Zhang et al., 2016, 2019). Because of the obvious effect of wind and wave on sediments, suspended matter is easy to appear resuspension, which provides the living environments for algae and promotes their growth. Then, this phenomenon in turn will lead to frequent outbreaks of algal blooms. Especially when the concentration of suspended sediment is too high, the light attenuation will cause the death of aquatic plants and cause great ecological damage and water pollution (Wu et al., 2012; Zhang et al., 2019). Therefore, the simulation and prediction of suspended sediment concentration is one of the key issues in the study of water environments of inland lakes. At present, the hydrodynamic modeling method has become one of the main methods for the prediction of suspended sediment concentration. Based on the coupled hydrodynamic characteristics, the hydrodynamic model comprehensively considers the interaction and genetic mechanism of multi factors such as lake bathymetry, wind speed, water depth, sedimentation rate and resuspension, so as to describe the multi-dimensional spatial-temporal distribution characteristics and dynamic evolution process of suspended sediments (Pang et al., 2008; Kouakou et al. 2013; Barasa and Wanyama, 2020; Song et al., 2020). However, due to the complicated hydrodynamic conditions and evolution mechanism, it is difficult to accurately simulate the concentration of suspended sediments for long-term forecast processes. In addition, the uncertainty of data and parameters will bring the initial error, measurement error and model error, and the simulation accuracy has also been affected.

In order to overcome the uncertainty of model simulation and to improve model accuracy, ensemble Kalman filter (EnKF) data assimilation method can provide an effective solution for the model simulation of suspended sediment concentration (Kalman et al., 1961; Evensen, 1994, 2003; Zhang et al., 2002; Vrugt et al., 2006; Clark et al., 2008).

Ensemble Kalman filter (EnKF) is based on the Bayesian theory and introduces ensemble forecast idea. Monte Carlo method is used for ensemble forecast and subsequent filtering matrix calculation. It overcomes the defect that other sequential assimilation methods cannot be applied to complex nonlinear and non-Gaussian distribution error systems (Evensen, 1994; Burgers, 1998; Bishop et al., 2001; Kazuyuki and Miyazaki, 2009; Szunyogh et al., 2010). Ensemble forecast uses initial fields within a certain error range to represent the initial state distribution, and then obtains the analysis field through the model prediction. Then, the analysis field is used as the initial field of the next simulation period, and the simulation is carried out iteratively. Ensemble forecast includes initial perturbation, model perturbation and boundary condition perturbation. The initial perturbation method is the first step of assimilation simulation, and its accuracy will directly affect the result quality of numerical simulation (Evensen, 1994; Burgers, 1998; Bishop et al., 2001).

At present, the generation methods of initial perturbation for hydrodynamic model simulation include Monte Carlo method, lagged average forecast (LAF) method, breeding of growing modes (BGM) method and ensemble transform Kalman filter (ETKF) method. Monte Carlo method is based on the viewpoint of ensemble prediction and considers that the main source of observation error is random error (Leith, 1974; Hollingsworth, 1980). As a part of ensemble Kalman filter, the random perturbation of Gaussian distribution is superimposed on the initial observation field to carry out the model calculation. Then, the standard Kalman filter method based on prediction update-process is proposed. However, due to the control of macro laws and the influence of accidental factors in the process of geographical model simulation, the spatial variability of errors is neither pure random nor pure structural, but has the dual characteristics of local randomness and overall structure. Therefore, the Monte Carlo method does not consider the spatial variation characteristics of errors, so it is difficult to generate a reasonable error field.

For LAF method, the initial perturbation is superimposed on the initial field of the model for ensemble forecast, and then the average value of the simulation results at the same time is taken as the initial value field with appropriate time lags (Hoffman and

Kalnay, 2010). In this ensemble forecast, the influence of spatial variability of errors is not considered in the generation of initial perturbation field.

The BGM method is based on the existing model and generates the error field from the repeated prediction of the model (Toth and Kalnay, 1997, Chen et al., 2018). Firstly, random perturbation is added to the initial analysis field, and the prediction model is calculated with and without perturbation. The results of perturbation prediction and control prediction are used to generate the difference field. Then, the proportion of the attenuation error in the difference field is reduced, and the modified difference field is used as the perturbation field for the next prediction. The above-mentioned process can generate reasonable initial perturbation field after many times of model prediction. In general, this method is better than Monte Carlo method. However, the error is regarded as random distribution in the process of model prediction.

For the ETKF initial perturbation field generation method, the transformation matrix is calculated from the analysis error covariance containing the observation information, and the size and distribution of the perturbation are adjusted by the transformation matrix (Bishop, et al., 2001; Majumdar 2010, 2015). Compared with Monte Carlo method and BGM method, this method can improve the generation of error field by introducing the concept of transformation matrix and expansion factor. In the process of generating the initial perturbation field, it is also considered that the model error and observation error are random and static distribution, and the influence of spatial structure of model error is not considered.

The above analysis shows that EnKF assimilation method coupled with hydrodynamic model can effectively improve the prediction accuracy of suspended sediment concentration in inland lakes. And, in the process of assimilation simulation, the existing initial perturbation generation methods consider that the perturbation error is a random Gaussian distribution. But in the process of dynamic prediction, the spatial distribution of the simulation error often contains structural information. Therefore, in this paper, the spatial structure characteristics of the simulated error of suspended sediment concentration is considered in the generation of perturbation. The proposed approach was validated through its application to assimilation simulation of suspended

sediment of Taihu Lake in China.

## 2. Study area and data

### 2.1. Study area

The study area, Taihu Lake, is typical inland shallow lake in China's East Asia Monsoon region, with a water surface area of approximately 2,338 km<sup>2</sup> and a mean depth of 1.9 m (Figure 2.1). The lake bed is covered by the mud composed of silt and clay with grain sizes ranging between 2 and 20 micrometers, which has accumulated high concentrations of nutrients for a long time. For the flat lake bed, the release of nutrients will occur due to sediment resuspension on windy days, resulting in the frequent occurrence of harmful algal blooms.

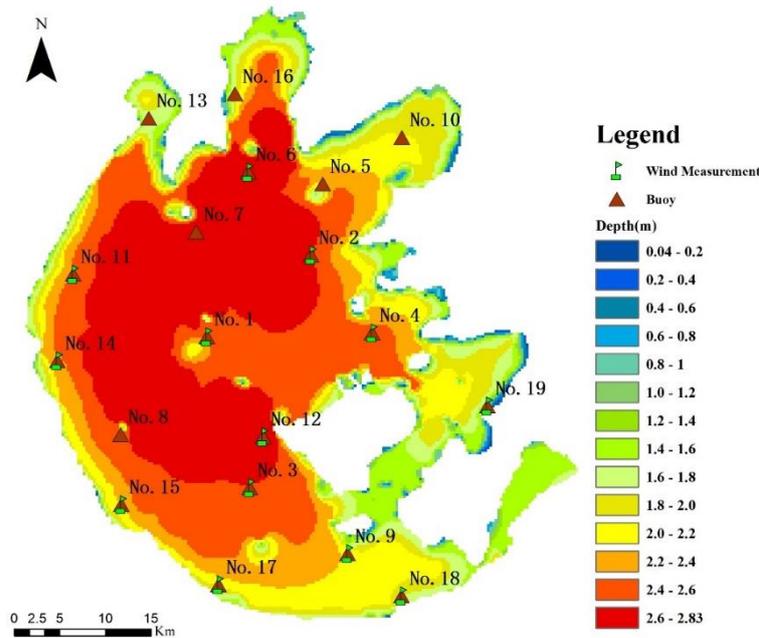


Figure 2.1 Underwater topography and buoys location of Taihu Lake

### 2.2. Data sources

Buoys located in Taihu Lake (Figure 2.1) measured real-time series data during the simulation period at one hour intervals. Each buoy has turbidity and pollutant data. Among the 19 buoys, only 14 buoys have meteorological data (wind speed, wind

direction). The concentration of total suspended sediments ( $C_{TSM}$ ) can be obtained according to the empirical relationship between turbidity (T) and  $C_{TSM}$  (following Equation 1) (Huang et al., 2015):

$$C_{TSM} = 0.9364 \times T + 12.046 \quad (N = 126; R^2 = 0.91; p < 0.0001) \quad (1)$$

For Taihu Lake, water enters the lake from the west and flows out to the east generally. According to the runoff model and measured hydrological data, eighteen main rivers and tributaries (including inflows and outflows) were selected for the simulation.

### 3. Methods

#### 3.1. Basic idea

We believe that there is spatial variability in hydrodynamic simulation of suspended sediment. the resulting error is composed of structural error and random error. It is then possible to make use of structural error in generating perturbation field for assimilation simulation than to treat all error as random error. Based on this idea, we propose a new method to generate the perturbation field and realize the assimilation simulation. The design of this method is mainly divided into three steps. The first is the construction of hydrodynamic model of suspended sediment. The second step is the generation of structural perturbation, which is combined with correlation analysis of time series data and dynamic model simulation. The third step is the construction and simulation experiment of assimilation coupling simulation model based on hybrid perturbation.

#### 3.2. Description of the models

##### 3.2.1. Hydrodynamic model

Three-dimensional shallow water equations are composed of the continuity equation, the momentum equation, and the free surface equation(Chen et al., 2003; Zhang et al., 2014, 2016):

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} - fv = -\frac{1}{\rho_0} \frac{\partial P}{\partial x} + \frac{\partial}{\partial x} (k_m \frac{\partial u}{\partial z}) + F_u \quad (2)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} + fu = -\frac{1}{\rho_0} \frac{\partial P}{\partial y} + \frac{\partial}{\partial x} (k_m \frac{\partial v}{\partial z}) + F_v \quad (3)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \quad (4)$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} \left[ \int_{-h}^{\eta} u dz \right] + \frac{\partial}{\partial y} \left[ \int_{-h}^{\eta} v dz \right] = 0 \quad (5)$$

where  $x$ ,  $y$ , and  $z$  are the east, north, and vertical axes in the Cartesian coordinate system;  $u$ ,  $v$ , and  $w$  are the velocity components in  $x$ ,  $y$ ,  $z$  directions;  $t$  is the time;  $\eta$  is water level;  $f$  is the Coriolis parameter;  $\rho$  is the density;  $P$  is the pressure;  $K_m$  is the vertical eddy viscosity coefficient;  $F_u$  and  $F_v$  represent the horizontal momentum; and  $h$  is the static water depth denoting the vertical length between the static water level and bottom of the lake with the positive direction downwards.

### 3.2.2. Suspended sediment model

The suspended sediment model (TSM model) uses a concentration-based approach subject to the following evolution equation (Chen et al., 2003; Zhang et al., 2014, 2016:

$$\frac{\partial C}{\partial t} + \frac{\partial uC}{\partial x} + \frac{\partial vC}{\partial y} + \frac{\partial (w - \omega_c)C}{\partial z} = \frac{\partial}{\partial x} \left( A_h \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left( A_h \frac{\partial C}{\partial y} \right) + \frac{\partial}{\partial z} \left( K_h \frac{\partial C}{\partial z} \right) \quad (6)$$

where  $C$  is the suspended sediment concentration,  $A_h$  is the horizontal eddy viscosity and  $K_h$  is the vertical eddy viscosity,  $\omega_c$  is the settling velocity of the sediment. The model includes transport of both suspended load and bedload, ignoring the process of condensation and dissolution.

At the surface, a no-flux boundary condition is used for the sediment concentration:

$$K_h \frac{\partial C}{\partial z} = 0, \quad z = \eta \quad (7)$$

The bottom boundary condition for sediment are:

$$K_h \frac{\partial C}{\partial z} = E - D, \quad z = \eta \quad (8)$$

where  $E$  ( $\text{kg m}^{-2} \text{h}^{-1}$ ) denotes the erosion flux, also called resuspension and  $D$  ( $\text{kg m}^{-2} \text{h}^{-1}$ ) denotes the deposition flux.

$$E = \Delta t Q (1 - P_b) F_b \left( \frac{\tau_b}{\tau_c} - 1 \right) \quad (9)$$

where  $Q$  is the erosive flux,  $P_b$  is the bottom porosity,  $F_b$  is the bottom fraction,  $\tau_b$  is the bottom shear stress, and  $\tau_c$  is the critical shear stress for erosion.

### 3.2.3 Description of EnKF method

Let  $X_t$  be an array of the true values,  $X_b$  be an array of the forecast values,  $X_a$  an array of analysis values, indicating the assimilation results, and  $Y$  an array of observational values. In a forecast model, the value of  $X_t$  at time step  $i$  can be predicted by

$$X_t^i = MX_t^{i-1} + \varepsilon_m^{i-1} \quad (10)$$

Where  $M$  presents the nonlinear model operator,  $\varepsilon$  is model error. In the Kalman filter forecast model system, the analysis values  $X_a$  is calculated by

$$X_a^i = X_b^i + K(Y^i - HX_b^i) \quad (11)$$

where  $K$  is the Kalman gain matrix, and  $H$  is an observation operator.

The EnKF is constructed by running a forecast model driven by a set of initial conditions and then estimate the error covariance relative to the ensemble mean to determine the ensemble analysis values for the next time step forecast. Let  $k$  denote the  $k$ th ensemble, and  $m$  the total number of the ensemble members, then

$$X_t^i = \frac{1}{m} \sum_{k=1}^m X_b^i \quad (12)$$

the analysis values at time step  $i$  for the  $k$ th ensemble model run are calculated by

$$X_{ak}^i = X_{bk}^i + K(Y_k^i - HX_{bk}^i) \quad (13)$$

The expression of gain matrix  $K$  is as follows:

$$\bar{K} = \bar{B}^i H^T (H \bar{B}^i H^T + \bar{R}^i)^{-1} \quad (14)$$

Where  $B$  is forecast error covariance matrix,  $R$  is the observational error covariance matrix. In the situation with a sufficiently large number of ensembles, the ensemble analysis error covariance matrix can be updated with a relationship as

$$\bar{A}^i = (I - KH) \bar{B}^i \quad (15)$$

To conduct the EnKF, we need to create an ensemble of the observational data constructed with the perturbation relative to the real value,

$$y_i = y + \delta_i, \delta_i = N(0, \sqrt{R}) \quad (16)$$

### 3.3. Modified perturbation field generation method

We believe that the structural error is difficult to obtain accurately, but we can think that it is closely related to the observation error at the adjacent time, that is, the spatial structure of the error at different time has correlation and hysteresis. At the time step  $i$ , the suspended sediment concentration of different buoys at the same time is selected to construct the data set  $H_i$ . The correlation analysis is carried out on the data sets at different time steps. According to the correlation coefficient, the time lag of time series data is determined.

Let  $m$  be the time lag,  $R_o^i$  is the observation error of  $i$  time step, then

$$R_o^i = Y^i - X_b^i \quad (17)$$

The structural error of this time step can be calculated by the average observation error within the time lag.

$$\overline{R_o^{i-1}} = \frac{1}{m} \sum_{i=i-m+1}^i R_o^i \quad (18)$$

In order to maintain the spatial structure characteristics of errors, the error matrix is standardized.

$$R_{std}^i = (\overline{R_o^i} - R_{mean}^i) / S_R^i \quad (19)$$

Where  $R_{mean}^i$  is the mean of error at time step  $i$ ,  $S_R^i$  is the standard deviation of the error at time step  $i$ .

Let  $S_p$  be the variance of a given perturbation field or the specified value. Then, the structural perturbation  $R_m^i$  can be calculated by

$$R_m^i = R_{std}^i \cdot \sqrt{S_p} \quad (20)$$

At time step  $i$ , the structural and random perturbation are superimposed to generate perturbation fields and then be substituted into the model for the next step calculation.

#### 4. Data assimilation experiments

A data assimilation experiment was performed on the basis of ensemble Kalman filter (EnKF) assimilation coupled with dynamic model. This section describes the setup of the experiments for evaluating the proposed method.

##### 4.1. Description of assimilation simulation process

The process of assimilation coupled simulation is shown in Figure 4.1, which includes five parts: data organization, parameter calibration, TSM model calculation, assimilation calculation and result analysis.

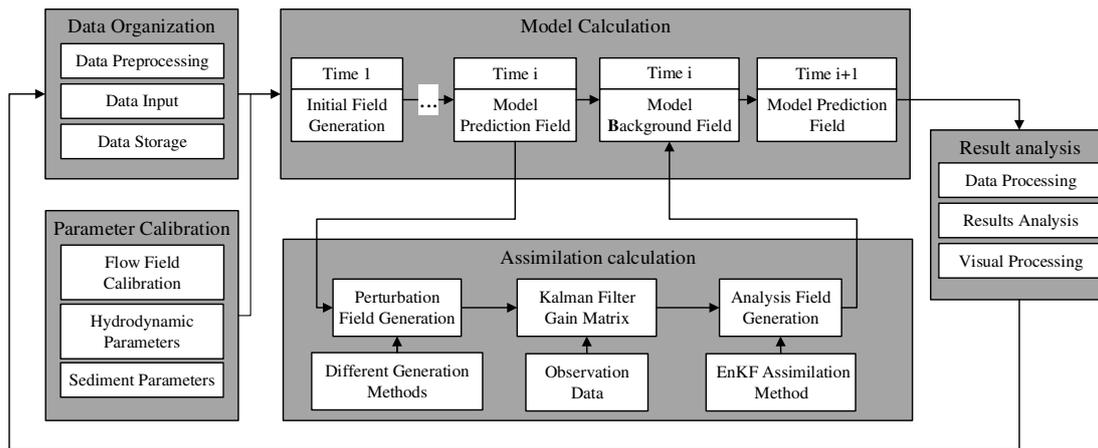


Figure 4.1 TSM model assimilation coupled simulation process

Data preparation is the data processing required by coupling simulation, and model parameter calibration is to find the optimal model parameters. In TSM model calculation, the initial background field is generated first, and the simulation results at  $i$  time step are taken as the background field of  $i+1$  time step. In assimilation simulation, at  $i-1$  time step, the modified perturbation method is used to generate the perturbation field. Then the  $i$ -step model prediction field is used as the background field. And the assimilation analysis field is generated by combining the observation data, which is imported into the model calculation module as the background field of the  $i+1$  time step for TSM model calculation. In all parts, model calibration is the precondition of assimilation coupled simulation and is further described below.

#### 4.2. TSM model calibration

The parameters of TSM model were calibrated by using the buoy turbidity and meteorological data from November 9 to 11 in 2015. Combined with the existing research results (Pang et al., 2008; Zhang et al., 2014, 2016), and through the verification of suspended sediment concentration and flow field, the final parameter calibration results are shown in Table 1. During the model calibration, the simulated flow field is consistent with the existing results. The concentration of suspended sediments simulated by TSM model was compared with the data of 19 buoys, The average RMSE is 22.36mg/l, and the average relative error is about 19.38%. The accuracy of the model is better than the existing research results of inland water. The

calibration results show that the TSM MODEL can be used for assimilation simulation.

Table 1. Calibration results of TSM model parameters

Model parameter	The ranges of parameters	Preferred value
critical shear stress	0.05-0.2(N/m <sup>2</sup> )	0.15
porosity	0.3-0.6	0.40
erosion rate	10 <sup>-5</sup> -5×10 <sup>-4</sup> (kg·m <sup>-2</sup> s <sup>-1</sup> )	6.0×10 <sup>-5</sup>
Settling velocity	0.01-0.2 (mm/s)	0.07185
Sediment density	1200-3800(g/cm <sup>3</sup> )	2650.0
Sediment mean diameter	0.002-0.1(mm)	0.032
Bottom friction coefficient	--	0.0025
Lake bottom roughness	--	0.001

### 4.3. Experiments design

In the assimilation simulation experiments, the turbidity data measured at the buoy stations in Taihu Lake and meteorological data are used. And the concentration of suspended sediment( $C_{TSM}$ ) is calculated by Eq.1. Since the data release interval of buoy data is 1 hour, this study simulated 72 hours of data from November 12 to 14, 2015, with a total of 72 time-steps. Taking 1 hour as the time interval, we analyzed the correlation coefficient of  $C_{TSM}$  on 19 buoys at different time steps. In this paper, We select time-step lag of 3 to participate in the perturbation of disturbance field. In this case, the maximum correlation coefficient is 0.99 and the minimum correlation coefficient is 0.76. Then, we designed the following five experiments.

*Experiment 1:* TSM model simulation experiment. In this experiment, combined with the parameters calibrated in Table 1, the  $C_{TSM}$  was simulated and predicted.

*Experiment 2:* Standard EnKF assimilation simulation experiment (S-EnKF). In this experiment, Monte Carlo method is used to generate the initial perturbation, and coupled with TSM model for data assimilation.

*Experiment 3:* EnKF assimilation simulation experiment with ETKF initial perturbation (ET-EnKF). In this experiment, Monte Carlo method is used to generate the random perturbation field at the initial time, and then the prediction perturbation is updated to analysis perturbation by ETKF, and the next

step ensemble prediction is carried out. Based on the transformation matrix and expansion factor, the stable initial perturbation field of ETKF is obtained.

*Experiment 4:* EnKF simulation experiment with modified perturbation (M-EnKF). In this experiment, the perturbation field is modified by the method proposed in this paper. It represents structural disturbance and random errors, and will be added in each time step for assimilation simulation.

*Experiment 5:* EnKF simulation experiment with modified ETKF perturbation (M-ET-EnKF). The perturbation generation method is the same as experiment 4, which is used to replace the Monte Carlo method, and add in each time step.

Among the above experiments, Experiment 1 provides a dynamic model for assimilation simulation. In Experiments 2 and 3, only random error was considered to generate perturbation field; in Experiments 4 and 5, the improved method was used to generate perturbation field. In this way, two different assimilation methods can be used to verify the effectiveness of the proposed method.

## **5. Results**

To examine the applicability of the assimilation simulation method based on mixed perturbation, the match-up points were used to test the modified method. Table 2 lists the RMSE (root mean square error) of errors computed of five assimilation simulation experiments and figure 5.1 shows the RMSE histogram. The experimental results show that the assimilation simulation based on random perturbation is better than TSM-model simulation on each buoy, and the mean values of root mean square errors for TSM-model, S-EnKF and ET-EnKF model are 22.14, 9.56, 9.55, respectively. For the assimilation simulation based on hybrid perturbation, the accuracy is further improved, the mean values of root mean square errors for M-EnKF model and M-ET-EnKF model are 8.70 and 8.68.

The concentration of total suspended sediments calculated from turbidity values measured by buoys are used to validate and compare the models results shown in Figure

5.2, which means that the results produced using assimilation simulation based on hybrid perturbation are more consistent with the measurements than those based on random perturbation. The reason is that the influence of structural error is considered in the generation of hybrid perturbation field, and the modified perturbation generation method is more suitable for assimilation simulation than the random perturbation generation method.

Table 2 RMSE of 19 buoys with 5 experimental results

Serial number	Buoys	TSM-Model	S-EnKF	M-EnKF	ET-EnKF	M-ET-EnKF
No.1	Pingtai Mountain	21.91	10.67	10.13	10.66	10.10
No.2	South of Wugui Mountain	13.49	7.74	7.26	7.74	7.23
No.3	No.14 Beacon	23.18	8.90	7.78	8.89	7.75
No.4	Man Mountain	28.91	10.38	9.29	10.37	9.26
No.5	South of Shazhu	12.73	7.31	6.92	7.28	6.90
No.6	South of Tuo Mountain	15.56	8.57	8.09	8.58	8.09
No.7	Jiao Mountain	14.51	7.30	7.11	7.29	7.11
No.8	Dalei Mountain	22.22	8.30	7.36	8.30	7.35
No.9	Ze Mountain	34.24	11.81	9.82	11.80	9.78
No.10	South of Xidongxi	39.56	12.82	10.79	12.80	10.80
No.11	Dapu Port	21.27	10.51	9.00	10.52	9.00
No.12	West of Xishan	18.65	7.55	6.99	7.55	6.96
No.13	Middle of Zhushan Lake	18.36	8.79	8.39	8.78	8.41
No.14	Spur of Lan Mountain	18.22	9.66	9.14	9.65	9.13
No.15	Xintang Harbor	21.97	9.47	8.33	9.47	8.33
No.16	Middle of Meiliang Lake	26.36	10.38	9.76	10.38	9.76
No.17	Xiaomeikou	20.30	11.93	11.39	11.94	11.37
No.18	Yangxi Harbor	27.71	11.09	10.00	11.08	9.96
No.19	Middle of Xu Lake	21.54	8.39	7.71	8.38	7.68
Average		22.14	9.56	8.70	9.55	8.68

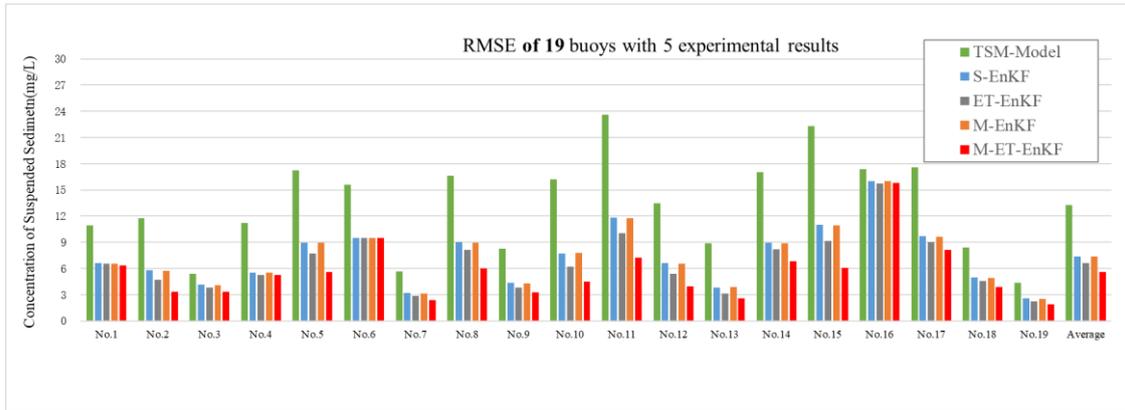
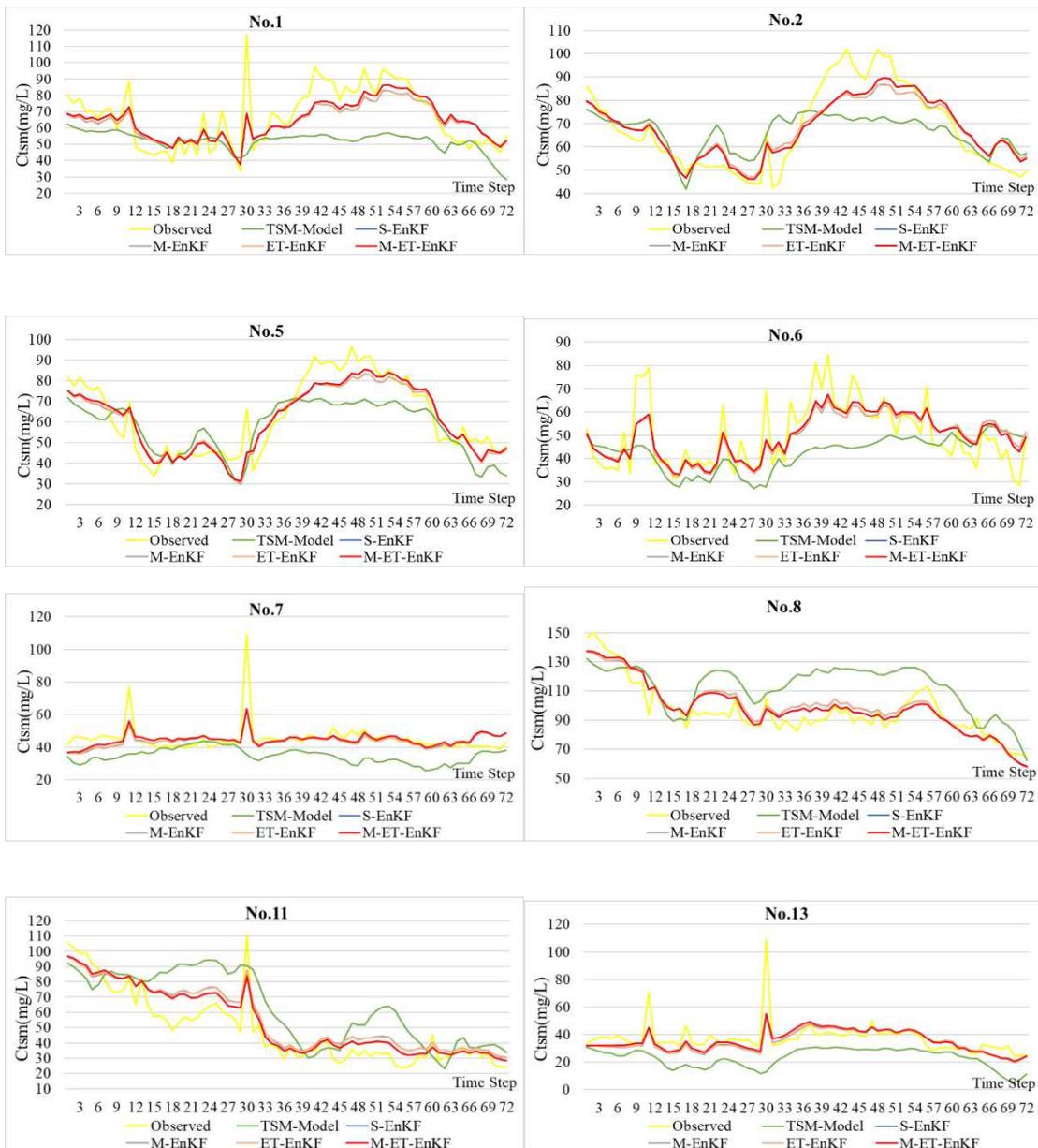


Figure 5.1. RMSE of 19 buoys with 5 experimental results



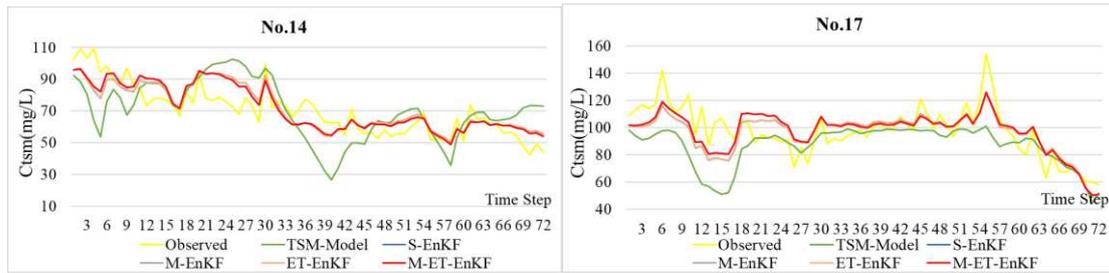


Figure 5.2. Comparison of model simulation results of some buoys in five experiments

## 6. Discussion

### 6.1. Determination of the variance of perturbation field

The basic purpose of ensemble perturbation generation is to generate uncorrelated initial values of perturbation, which can more accurately represent the analysis error covariance. Therefore, error structure and amplitude of perturbation are the key problems in the study of initial perturbation in ensemble forecasting. The error structure reflects the spatial distribution characteristics of the analysis error, while the perturbation amplitude reflects that the analysis error should be equivalent to the corresponding prediction error, and directly affects the dispersion of the prediction set. Due to the existence of uncertain factors, this paper uses the control variable method. Keeping the other factors unchanged, each factor was assimilated several times for 24 hours, and finally the optimal value was obtained based on the average RMSE of the results. Finally, the observation error variance is 4.5, and the assimilation model error variance is 4.0, which are selected (Figure 6.1). In this case, the RMSE of assimilation simulation results is the minimum.

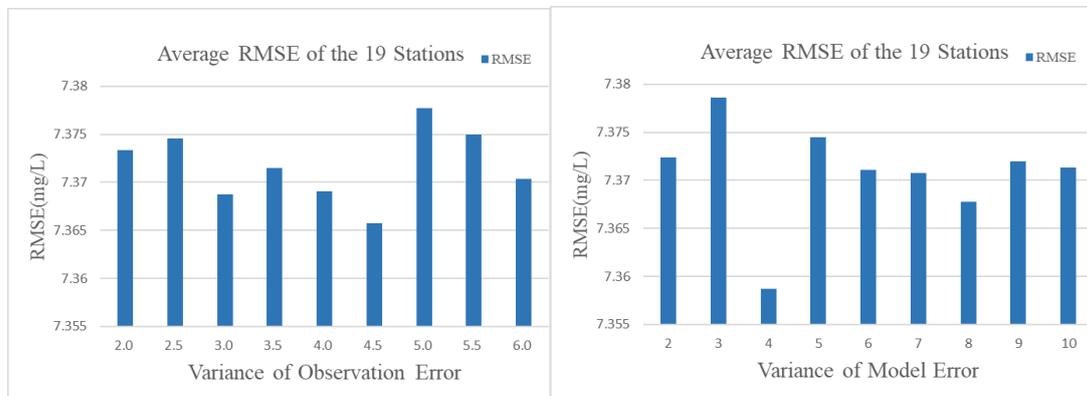


Figure 6.1. Calibration results of error field based on control variable method

We consider that the assimilation model error variance represents the ratio of structural error to perturbation field, and finally generates the perturbation field of each time step. However, due to the complexity of error disturbance, whether there is a better structural error variance generation method remains to be studied.

## 6.2. The impact of ensemble size on assimilation simulation results

The ensemble size in EnKF method will affect the assimilation simulation results, so we must first determine the total number of the ensemble members selected in the forecast model run. In this paper, Monte Carlo method is used to generate initial perturbation, and EnKF method is used to compare and analyze the influence of ensemble size on the forecasting. We set the number of ensemble members from 10 to 100 with an interval of 10. In the same way, the control variable method is used to calibrate the ensemble size. With other factors unchanged, when the number of ensemble members is equal to 40 the assimilation simulation results is better than others (Figure 6.2). Obviously, this method needs to be tried many times and needs to be improved in the future.

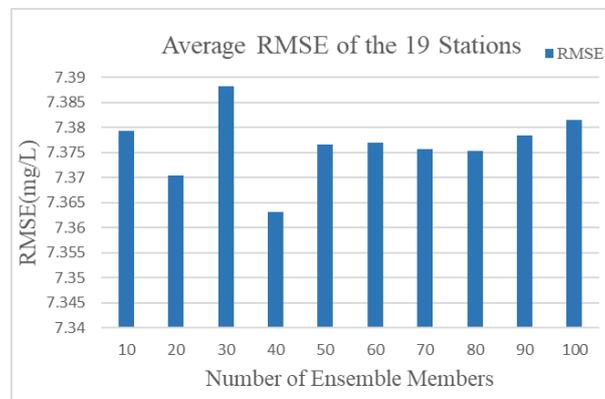


Figure 6.2. Calibration results of ensemble size based on control variable method

## 7. Conclusion

In this paper, a new method to generate the perturbation field for assimilation simulation was proposed. The method mainly verifies an idea that the combined effect of structural error should be considered in the generation of perturbation field. Based on this idea, we designed five assimilation experiments to simulate the concentration of suspended sediments in Taihu Lake of China. After three days and 72 time steps of assimilation

simulation based on hybrid perturbation generation, we found that the proposed assimilation method provided results that were more consistent with buoy-measured data. The RMSE results show that the assimilation method based on hybrid perturbation field generation method has higher simulation accuracy than other methods. This study shows that the method is effective and provides a new idea for the assimilation simulation of suspended sediment concentration in inland lakes.

### **Acknowledgements**

This work was supported by the National Natural Science Foundation of China (No. 41571386 and No. U1811464) and the Priority Academic Program Development of the Jiangsu Higher Education Institutions (PAPD). We thank the anonymous reviewers for their constructive comments on an earlier version of this paper.

### **Declarations**

#### **Funding**

This work was supported by the National Natural Science Foundation of China (No. 41571386 and No. U1811464) and the Priority Academic Program Development of the Jiangsu Higher Education Institutions (PAPD).

#### **Conflicts of interest/Competing interests**

No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

#### **Availability of data and material**

The high-resolution underwater terrain data were provided by the Chinese government. The authors do not have the right to make them publicly available on the Internet. Meteorological data and other model simulation data sets of this study are available from the corresponding author on reasonable request.

#### **Code availability**

The MATLAB code for assimilation and FORTRAN code for dynamic model of this study are available from the corresponding author on reasonable request.

### **Authors' contributions**

The method was conceived by Fei Guo and A-xing Zhu, the experiments were designed by Fei Guo and A-xing Zhu, and performed by Jingjia Zhang. The algorithm was conceived, implemented and optimized by Fei Guo, Jingjia Zhang, Zhuo Zhang and Hong Zhang. Fei Guo and Jingjia Zhang took part in writing the paper. A-xing Zhu provided critical review and substantially revised the manuscript. All authors read and approved the final manuscript.

### **References**

- Barasa B, Wanyama J (2020) Freshwater lake inundation monitoring using Sentinel-1 SAR imagery in Eastern Uganda. *Annals of GIS* 26: 191-200.
- Bishop CH, Etherton B J, Majumdar SJ (2001) Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review* 129(3):420-436.
- Burgers G, Jan Van Leeuwen P, Evensen G (1998) Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review* 126(6):1719-1724.
- Chen C, Liu H, Beardsley RC (2003) An Unstructured Grid, Finite-Volume, Three-Dimensional, Primitive Equations Ocean Model: Application to Coastal Ocean and Estuaries. *Journal of Atmospheric and Oceanic Technology* 20(1):159-186.
- Chen C, Xiang LI, Hongrang HE, et al (2018) Algorithm based on local breeding of growing modes for convection-allowing ensemble forecasting. *Science China Earth Sciences* 61(4):1-11.
- Chen J, Gao G, et al (2006) Hydrostatic settling of suspended matter of Lake Taihu. *Journal of Lake Sciences* 18(5): 528-534.
- Clark MP, Rupp DE, Woods RA, et al (2008) Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources* 31(10):1309-1324.
- Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99(C5):10143.

Evensen G (2003) The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* 53(4):343-367.

Friese K, Schmidt G, Lena JCD, et al (2010) Anthropogenic influence on the degradation of an urban lake – The Pampulha reservoir in Belo Horizonte, Minas Gerais, Brazil. *Limnologica* 40( 2):114-125.

Ghale YAG, Altunkaynak A, Unal A (2018) Investigation Anthropogenic Impacts and Climate Factors on Drying up of Urmia Lake using Water Budget and Drought Analysis. *Water Resources Management* 32(1):1-13.

Hoffman RN, Kalnay E (2010) Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A* 35a(2):100-118.

Hollingsworth CG, Heydrick FP (1980) Investigator initiated multi-center clinical trials and the national institutes of health. *Controlled Clinical Trials* 1(2):179-180.

Huang C, Yang H, Zhu AX, et al (2015) Evaluation of the Geostationary Ocean Color Imager (GOCI) to monitor the dynamic characteristics of suspension sediment in Taihu Lake. *International journal of remote sensing* 36(15-16):3859-3874.

Kalman RE, Bucy RS (1961) *New Results in Linear Filtering and Prediction Theory* 1. Transaction of the ASME, Journal of Basic Engineering 1961.

Kouakou KL, Kouame KI, Konan KS, et al. Two-Dimensional Numerical Simulation of the Hydro-Sedimentary Phenomena in Lake Taabo, Côte d'Ivoire. *Water Resources Management* 27: 4379–4394.

Leith CE (1974) Theoretical Skill of Monte Carlo Forecasts. *Monthly Weather Review* 102(6):409.

Majumdar SJ, Bishop, et al (2010) A comparison of ensemble-transform Kalman-filter targeting guidance with ECMWF and NRL total-energy singular-vector guidance. *Quarterly Journal of the Royal Meteorological Society* 128(585):2527-2549.

Majumdar S J, Chen S G, Wu C C (2015) Characteristics of Ensemble Transform Kalman Filter adaptive sampling guidance for tropical cyclones. *Quarterly Journal of the Royal Meteorological Society* 137(655):503-520.

Kazuyuki, Miyazaki (2009) Performance of a local ensemble transform Kalman filter for the analysis of atmospheric circulation and distribution of long-lived tracers under

idealized conditions. *Journal of Geophysical Research: Atmospheres*, 2009.

Pang Y, Zhuang W, et al (2008) Experiment and model Simulation of suspended Solids in Taihu Lake under wind-wave disturbance. *Environmental Science (China)*10:2743-2748.

Song C, Shi X, Wang J (2020) Spatiotemporally Varying Coefficients (STVC) model: a Bayesian local regression to detect spatial and temporal nonstationarity in variables relationships. *Annals of GIS* 26:277-29.

Szunyogh I, Kostelich E J, Gyarmati G, et al (2010) A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus Series A-dynamic Meteorology & Oceanography* 60(1):113-130.

Toth Z, Kalnay E (1997) Ensemble Forecasting at NCEP and the Breeding Method. *Monthly Weather Review* 125(12):3297-3319.

Vrugt JA, Gupta HV, Nualláin BÓ, et al (2006) Real-Time Data Assimilation for Operational Ensemble Streamflow Forecasting. *Journal of Hydrometeorology* 7(3):548.

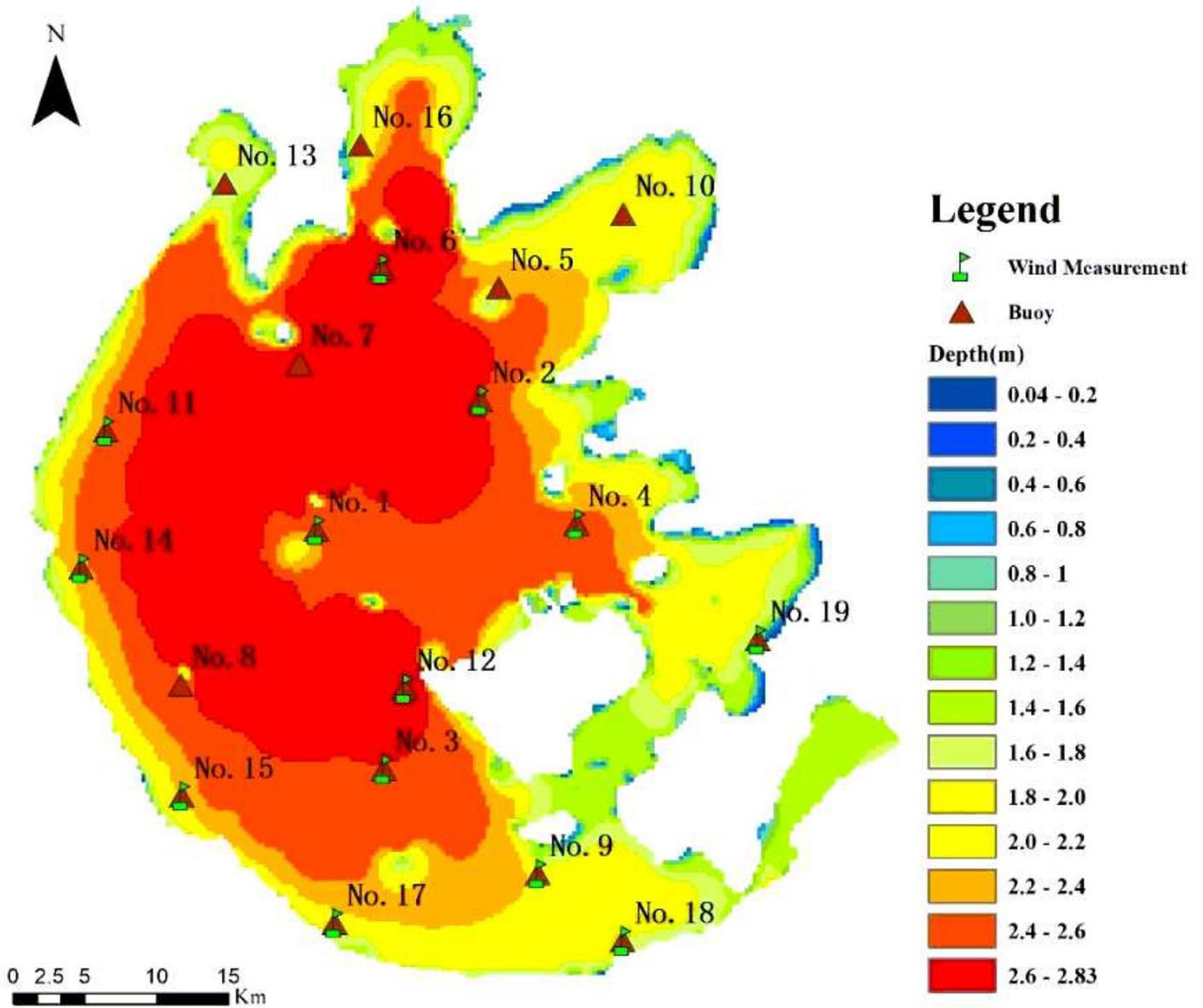
Wei L, Huang C, Zhong Y, et al (2019) Inland Waters Suspended Solids Concentration Retrieval Based on PSO-LSSVM for UAV-Borne Hyperspectral Remote Sensing Imagery. *Remote Sensing* 11(12):1455.

Wu J, Zeng H, Yu H, et al (2012) Water and Sediment Quality in Lakes along the Middle and Lower Reaches of the Yangtze River, China. *Water Resources Management* 26(12):3601-3618.

Zhang Z, Huang C, Guo F, et al (2019) A new technology for suspended sediment simulation in Lake Taihu, China: Combination of hydrodynamic modeling and remote sensing. *Journal of limnology* 121-134

Zhang Z, Peng G, Guo F, et al (2016) The key technologies for eutrophication simulation and algal bloom prediction in Lake Taihu, China. *Environmental Earth Sciences* 75(18):1295.

# Figures



**Figure 1**

Underwater topography and buoys location of Taihu Lake Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its

authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

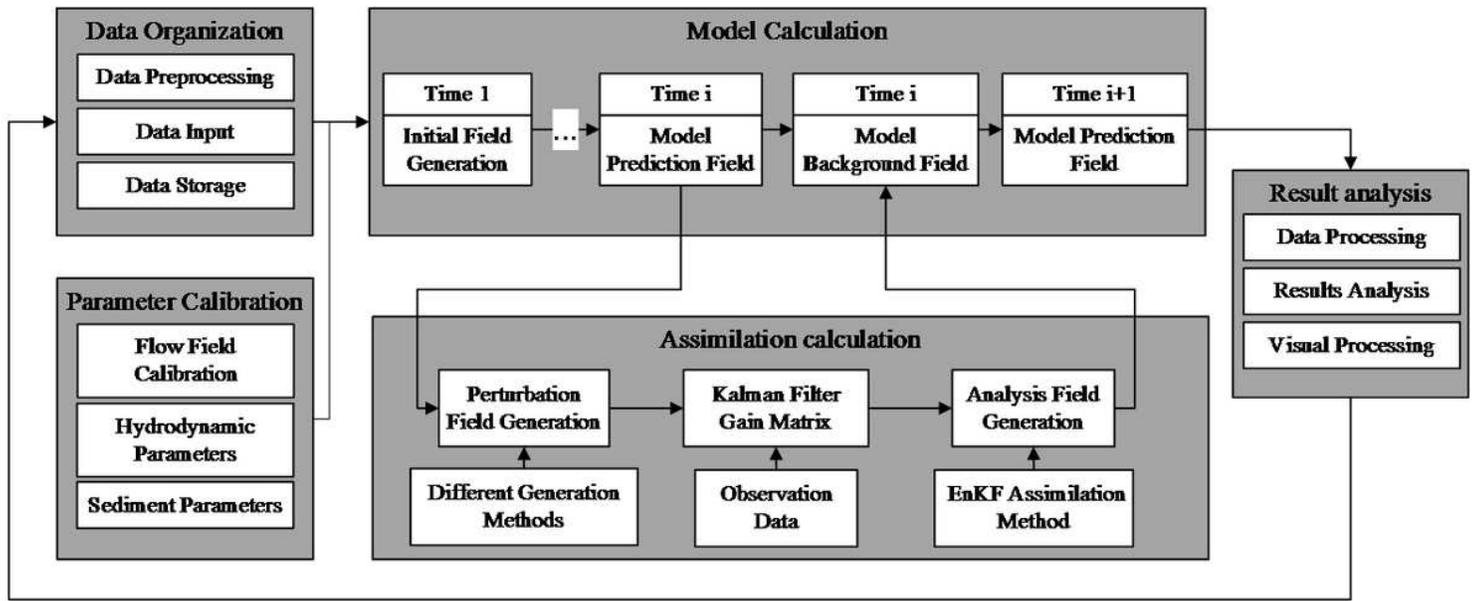


Figure 2

TSM model assimilation coupled simulation process

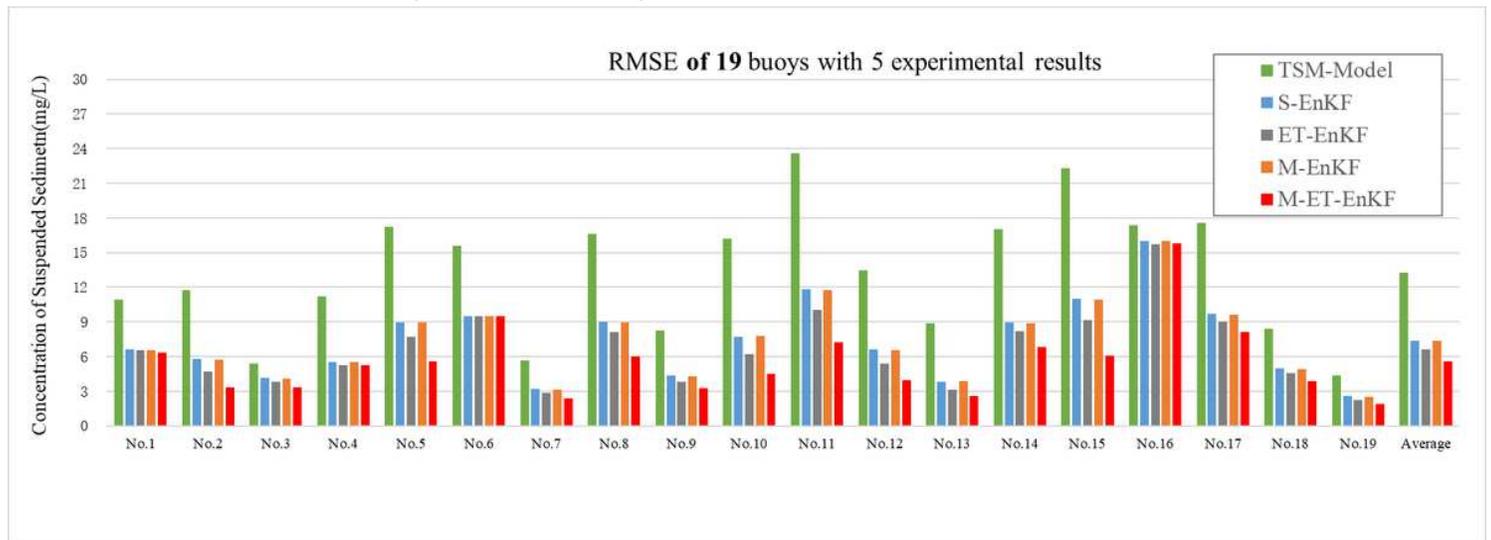
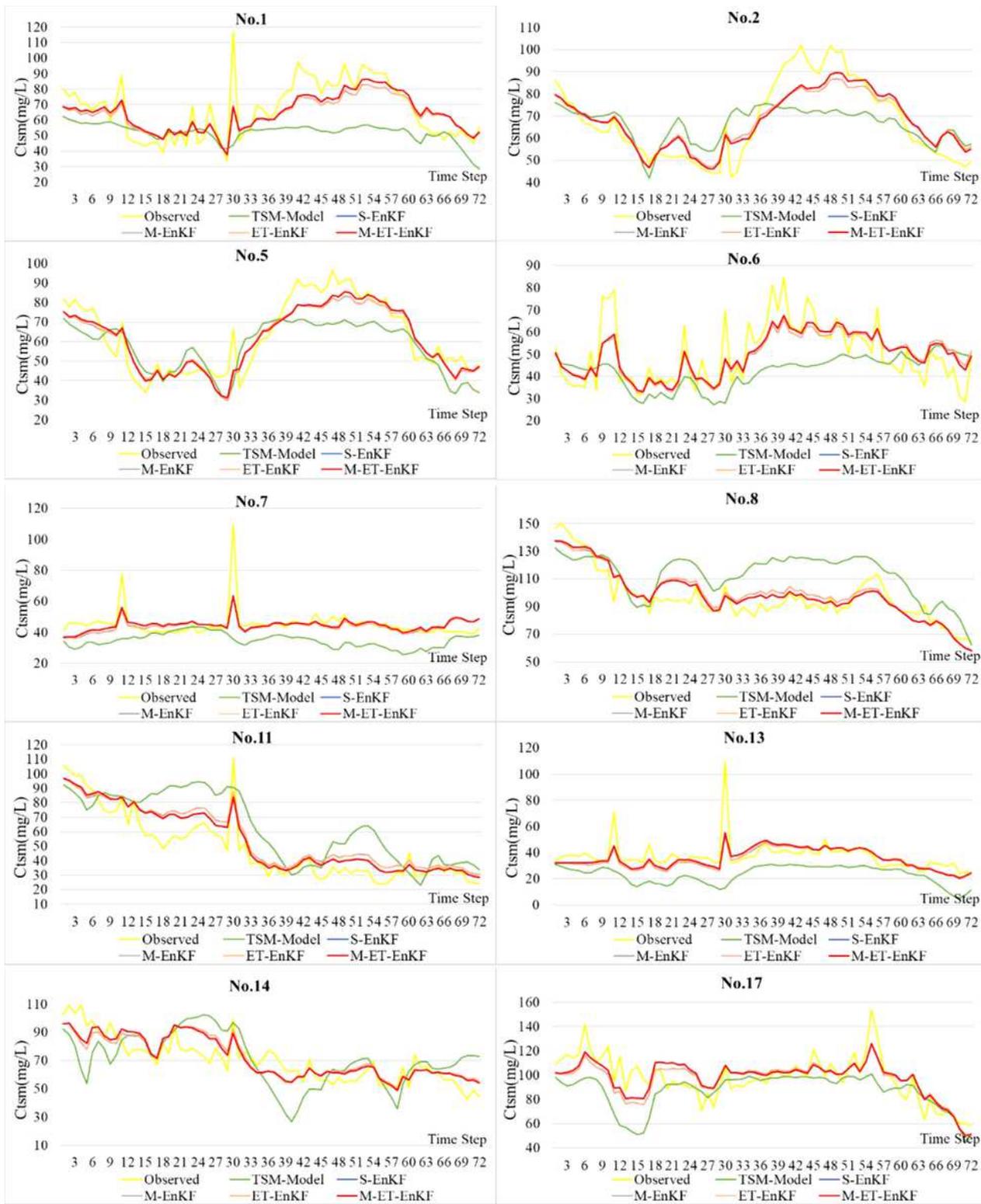


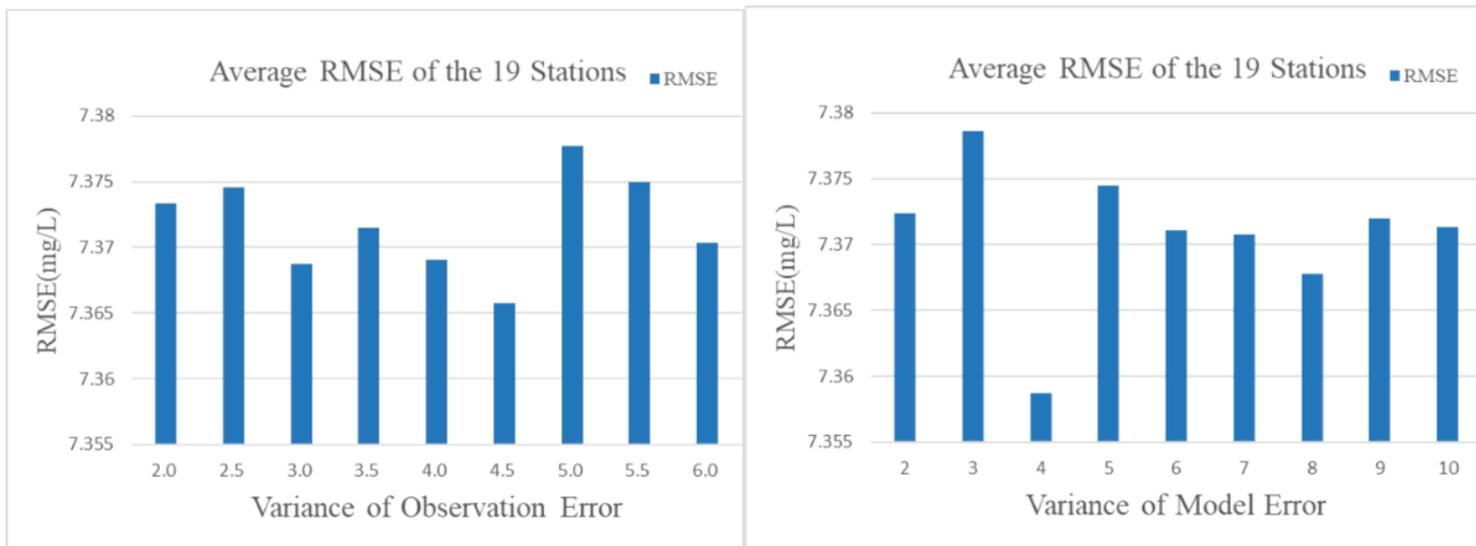
Figure 3

RMSE of 19 buoys with 5 experimental results



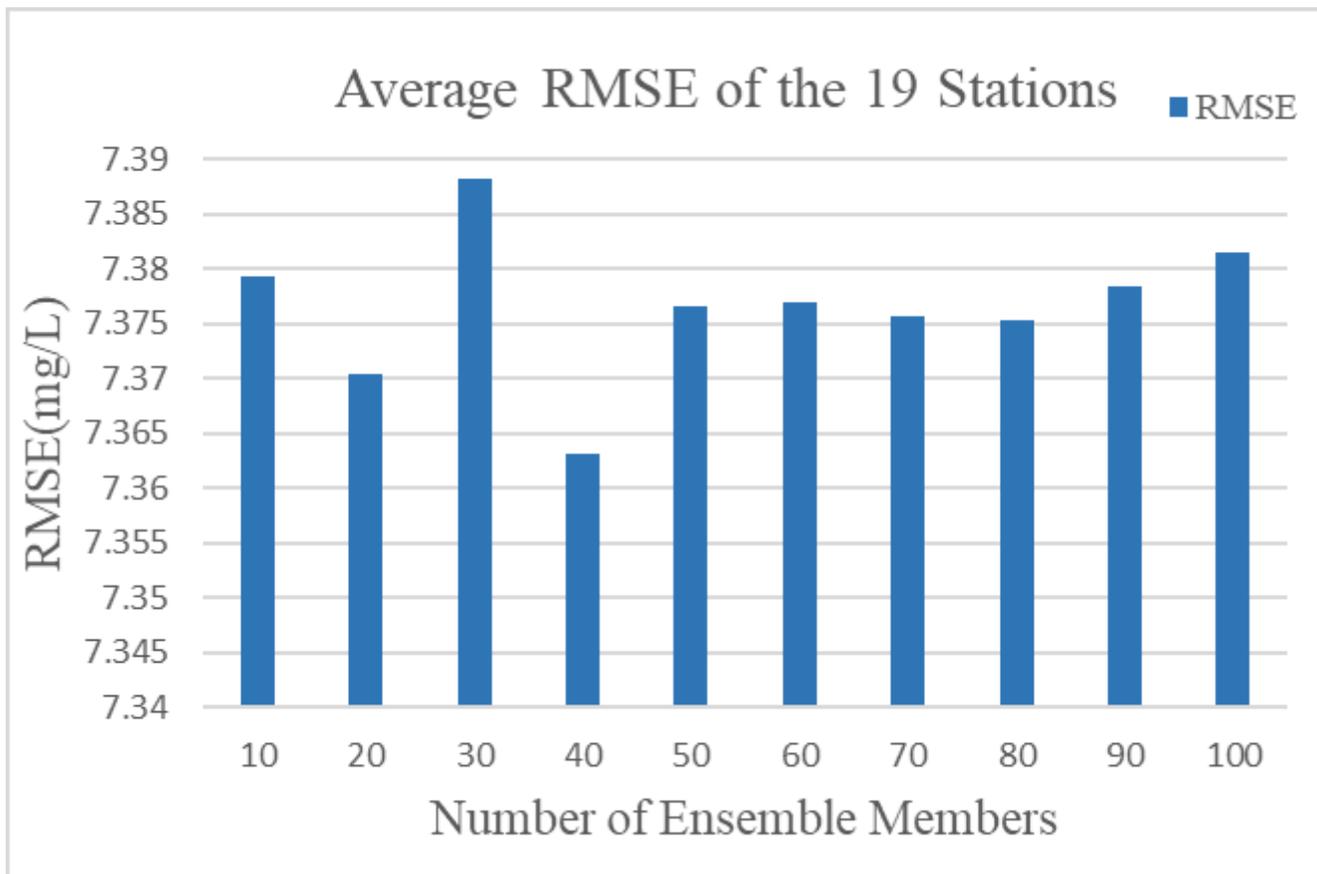
**Figure 4**

Comparison of model simulation results of some buoys in five experiments



**Figure 5**

Calibration results of error field based on control variable method



**Figure 6**

Calibration results of ensemble size based on control variable method