JGR Oceans

RESEARCH ARTICLE
10.1029/2021JC017911

Key Points:

- This is the first study to apply the ensemble Kalman filter to an actual coastal estuary using real 1-year observation data
- It is possible to maintain the ensemble spread by perturbing atmospheric forcing, lateral boundary conditions, and river discharge forcing
- This method achieves robust annual data assimilation and reflects seasonal fluctuations

Correspondence to:
Y. Matsuzaki,
matsuzaki-y@p.mpat.go.jp

Citation:
Matsuzaki, Y., & Inoue, T. (2022). Perturbation of boundary conditions to create appropriate ensembles for regional data assimilation in coastal estuary modeling. *Journal of Geophysical Research: Oceans*, *127*, e2021JC017911. https://doi.org/10.1029/2021JC017911

Received 18 AUG 2021
Accepted 21 MAR 2022

Perturbation of Boundary Conditions to Create Appropriate Ensembles for Regional Data Assimilation in Coastal Estuary Modeling

Yoshitaka Matsuzaki[1] and Tetsunori Inoue[1,2]

[1]Marine Environment Control System Department, Port and Airport Research Institute, Yokosuka, Japan, [2]Estuary Research Center, Shimane University, Matsue, Japan

Abstract Regional data assimilation is conducted for a coastal estuary using the ensemble Kalman filter, real observation data from Ise Bay, Japan, and a simulation model called the Ise Bay Simulator. The applicability and robustness of the method are then examined. We also analyze the relationship between the boundary conditions, which add perturbations and the data assimilation results of water temperature and salinity. A method of creating an ensemble by perturbing three boundary conditions (atmospheric forcing, lateral boundary conditions [open boundary conditions], and river discharge forcing) is then proposed. In situ water temperature and salinity profiles observed at fixed points are assimilated daily. The proposed assimilation method provides stable data assimilation without unnatural values for water temperature and salinity throughout the year. Further, applying a perturbation to the three boundary conditions does not lead to filter divergence, thus indicating good applicability and robustness, that is, applying a perturbation to the three boundary conditions does not degenerate the ensemble spread. According to a sensitivity experiment, perturbing the atmospheric boundary conditions of both air temperature and wind speed increases the ensemble spread of water temperature, especially near the surface layer. Wind speed has the greatest influence on the magnitude of the salinity ensemble spread, and its dominance depends on location. Perturbation of lateral boundary conditions increases the ensemble spread of water temperature and salinity at all water depths near the bay mouth, and the observations are effectively assimilated. Perturbation of river discharge forcing successfully assimilates water temperature and salinity near the estuary.

Plain Language Summary The accuracy of numerical simulations of physical features such as water temperature and salinity in coastal estuaries may be limited by uncertainties in weather, river, and ocean conditions. Therefore, we attempt to improve the accuracy of numerical simulation by inputting observed values into a simulation through data assimilation: a method of extracting information from both observational data and numerical models and combining them statistically to obtain the optimum solution. In this study, we developed a regional data assimilation system for the coastal estuary: Ise Bay, Japan, by considering model errors stemming from uncertainties in atmospheric forcing, lateral boundary conditions, and river discharge forcings. We evaluate the data assimilation results using objective evaluation indexes. The data assimilation results for water temperature and salinity for 1-year exhibit high accuracy and verify the applicability and robustness of the proposed data assimilation method. This is the first step in applying data assimilation for coastal estuaries, and this study will help many researchers set up data assimilation systems for other coastal estuaries.

1. Introduction

There are certain difficulties in conducting precise numerical simulation of physical phenomena at coastal estuaries. Data assimilation methods can improve the reproduction accuracy and advance our understanding of physical processes. However, applying data assimilation to coastal numerical simulations is still challenging because of the complexity of the physical process (Stanev et al., 2016). One of the most important conditions of data assimilation is the background error covariance (forecast error covariance; Edwards et al., 2015; Hoteit et al., 2018; Moore et al., 2011; Sakov et al., 2012). Although there are several methods for calculating the background error covariance (Fisher & Courtier, 1995; Fu et al., 1993; Weaver & Courtier, 2001), an appropriate method for regional data assimilation in coastal estuaries has not yet been determined. From this viewpoint, the ensemble Kalman filter (EnKF), which can express and update the background error covariance using ensemble members

that indicate the system error, that is, the numerical simulation error (Evensen, 1994), is a potential procedure for coastal calculation. Compared to the ocean, coastal estuaries have a complex flow field with a short timescale that includes density-driven currents, tidal currents, wind-induced currents, and ocean currents; thus, EnKF is a suitable method for analyzing the physical processes of coastal estuaries.

Ensemble members are created by perturbing the error factors of numerical simulations to represent the ensemble spread or variability. There are approximately three types of error factors that contribute to the error of a numerical simulation: (a) initial conditions, (b) forcing data, and (c) model equations and parameters (Turner et al., 2008). For numerical models of open oceans, which are relatively advanced in data assimilation, several studies have suggested calculating ensembles to represent the initial conditions (Sakov et al., 2012), the atmospheric forcing errors (Lima et al., 2019; Mirouze & Storto, 2019; Penny et al., 2015; Sakov et al., 2012), parameter errors (Brankart et al., 2015), and combinations of the atmospheric forcing errors and parameter errors (Baduru et al., 2019; Kwon et al., 2016; Sanikommu et al., 2020; Vandenbulcke & Barth, 2015). This reflects the assumption that initial conditions, models, and atmospheric boundary conditions are important for the precise simulation of physical processes in the open ocean, which has a relatively large calculation area and long-term fluctuations.

However, the successful perturbation of error factors to generate ensembles has not yet been achieved for regional data assimilation in coastal estuaries. We suggest that perturbation of three boundary conditions is required to generate ensembles for regional data assimilation of a coastal estuary specifically, atmospheric forcing, lateral boundary conditions, and river discharge forcing. This is because coastal areas are more affected by boundary conditions because of the small calculation area. Moreover, it is very difficult to set accurate boundary conditions because of limitations of available data set, despite their substantial influence on the results of regional coastal numerical simulations. Previous studies have reported that error variability caused by the initial conditions decreases with time in coastal numerical models (Turner et al., 2008) because such models are dominated by relatively short-term fluctuations. Moreover, the error caused by the initial conditions can be maintained by multiplicative inflation (Anderson & Anderson, 1999; Whitaker & Hamill, 2012); however, this technique does not generate a consistent physical model (Sanikommu et al., 2020).

Some previous studies have conducted regional data assimilation for coastal estuaries using EnKF. For example, Turner et al. (2008) generated ensemble members for EnKF by perturbing atmospheric forcing, lateral boundary conditions, and river discharge forcing. They also proposed adding random noise with a normal distribution to the boundary conditions of the ensemble members as a method of perturbation. They applied this method to observing system simulation experiments (OSSEs) in Port Phillip Bay, Australia, using synthetic sea surface temperature (SST) data, and reported good prediction capability. Hoffman et al. (2012) also conducted OSSEs in Chesapeake Bay, USA. The assimilated synthetic data included fixed point water temperature, salinity, and SST data. They created ensembles by perturbing the initial conditions and wind via atmospheric forcing. Although they did not add perturbations to lateral boundary conditions and river discharge forcing, they noted it may be necessary to add perturbations to lateral boundary conditions and river discharge forcing for generating ensembles when data assimilation is conducted using real observation data. Furthermore, Khanarmuei et al. (2021) conducted twin experiments for the shallow estuary of Currimundi Lake, Australia. They perturbed the lateral boundary condition of water level and river discharge, and the synthetic observed values of water level and current velocity were assimilated. They revealed the importance of perturbing the lateral boundary condition of the water level when assimilating the observed water level value and that of perturbing river discharge forcing when assimilating the observed current velocity value. However, real observation data were not included in their previous experiments and synthetic observation data were simulated numerically. In addition, the error factors were already known because the experiments were virtual.

Thus, we conducted the EnKF in the Ise Bay, Japan using actual observed data (Matsuzaki & Inoue, 2020). Ise Bay is a coastal area with an estuary. Ensembles were made to perturb lateral boundary condition of water temperature and river water temperature. The assimilation results were compared with the observed values, and it was confirmed that the water temperature improved. However, Matsuzaki and Inoue (2020) was conducted only in the summer, and the data assimilation performance and the robustness of the data assimilation method throughout the year have not been evaluated. Therefore, it is imperative to conduct assessment throughout the year to respond to seasonal fluctuations and confirm applicability and robustness of the methods (Turner et al., 2008).

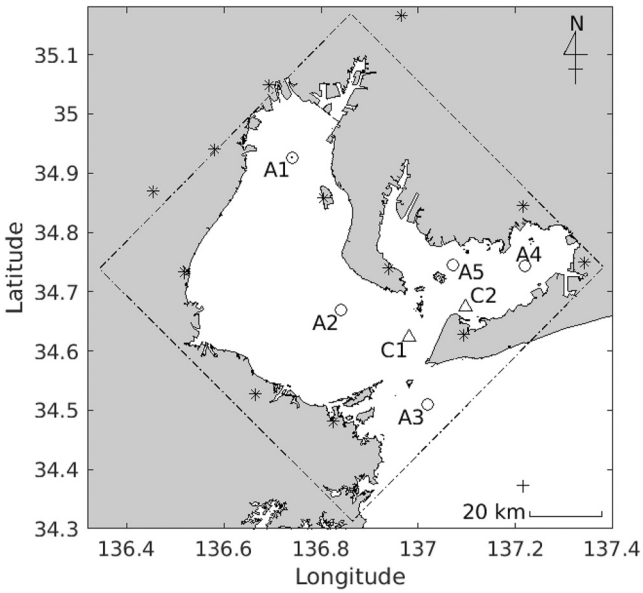

Figure 1. Location of Ise Bay, Japan. Dashed line indicates the experimental area for data assimilation. Circles and triangles represent observation stations used for data assimilation and accuracy validation, respectively. A1 to A5 circles represent the back of Ise Bay, the center of Ise Bay, the mouth of Ise Bay, No. 1 buoy, and No. 2 buoy, respectively. C1 and C2 triangles indicate the Nakayama Channel and No. 3 buoy, respectively. Asterisks represent the observation stations used to generate atmospheric forcing data. Cross represents the observation point used to generate the lateral boundary conditions.

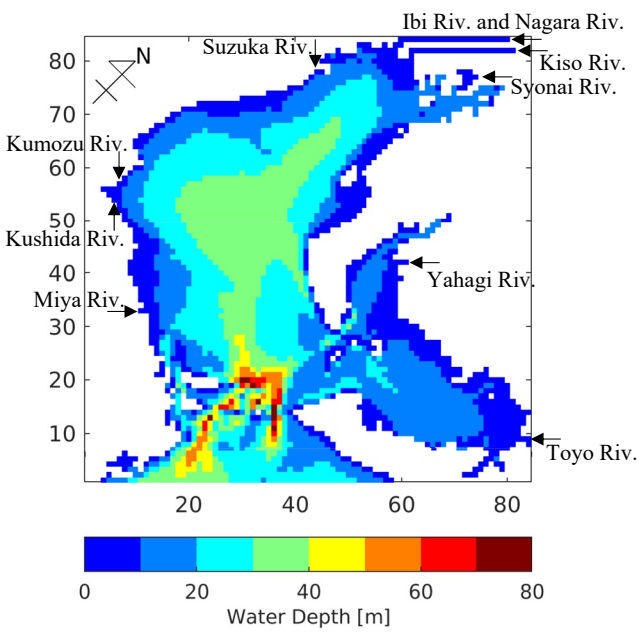

Figure 2. Calculation grid and its water depth. Arrows indicate the inflow position and directions of 10 class A rivers. The horizontal and vertical axes indicate a calculation grid of x and y-direction and the grid number is 85 × 85. In this study, the entire sea area is defined as Ise Bay, the eastern side is defined as Mikawa Bay, and the Ise Bay excluding Mikawa Bay is defined as the western side of Ise Bay.

In this study, we conducted EnKF in the Ise Bay, Japan, and evaluate the applicability of the data assimilation method. Specifically, we analyze the optimal method for adding perturbations to create ensemble members for regional data assimilation of a coastal estuary. This study also analyzes the relationship between the boundary conditions, which add perturbations and the assimilated water temperature and salinity data results as well as their ensemble spread. To the best of our knowledge, this is the first study to employ EnKF with actual water temperature and salinity data for a coastal estuary over 1 year. Additionally, no previous studies have generated ensembles by perturbing lateral boundary conditions and river discharge forcing under practical conditions; thus, this study reveals the effect of perturbing boundary conditions. In addition, we confirm the robustness of the regional coastal data assimilation method by performing long-term integral data assimilation and quantitative evaluation using the data assimilation results. The proposed data assimilation method is characterized by high applicability to coastal estuaries and responds to both short-term and long-term fluctuations, including seasonal changes.

2. Materials and Methods

2.1. Simulation Model and Setup

Simulations were conducted using the Ise Bay Simulator (Tanaka & Suzuki, 2010), which is a nonhydrostatic numerical simulation model. The model was configured to cover the entire area of Ise Bay (Figure 1; surface area: 2,342 km^2, volume: 3.94×10^{10} m^3, mean depth: 17 m), which is located in the south-central part of Honshu Island, Japan. The bay is approximately 70 km long in both longitudinal and latitudinal directions and is divided into two. The western side has a surface area of 1,738 km^2, a volume of 3.39×10^{10} m^3, and a mean depth of 20 m (Figure 2). The water depth was about 35 m in the center of the bay, and the maximum water depth is 100 m at the bay's mouth. The eastern side is called Mikawa Bay, which has a surface area, volume, and mean depth of 604 km^2, 5.5×10^9 m^3, and 9 m, respectively. The maximum water depth was about 35 m, but most water was under 20 m deep (Figure 2). The lateral boundary borders the Pacific Ocean. The annual river discharge (about 2.0×10^{10} m^3) is about half of the volume of Ise Bay. Ten class A rivers, which are defined as nationally controlled rivers that are generally large in scale, and 91 medium and small-sized rivers flow into Ise Bay. The Kiso, Nagara, and Ibi rivers have the highest discharges, and these three rivers flow in from the back of the bay (Figure 2). Eight class A rivers flow into the west side of Ise Bay and two into Mikawa Bay. Thus, freshwater inflow is biased to the west side of Ise Bay. There are two types of characteristic wind patterns in Ise Bay; from fall to spring, seasonal wind flow from the northwest, and in the summer, land and sea breeze flow from the southeast during the day and from the northwest at night (Sekine et al., 2002). Since the water is relatively shallow, seawater exchange occurs due to the influence of wind.

The Ise Bay model uses the Cartesian coordinate system, which simulates the water current structure of coastal estuaries with horizontal resolution of 800 m for both x and y-direction (Figure 2). The coordinate system is set by rotating it clockwise by 45°. The number of vertical layers is 32, with 0.5-m spacing near the water surface and 30-m spacing near the seabed. Input water depth data were created by reading the water depth from a chart made by the Japan Coast Guard. A subgrid-scale model was used for the horizontal

turbulence model; the model of Nakamura and Hayakawa (1991), which has been modified from the model of Henderson-Sellers (1985), was used for the vertical turbulence model. The Sommerfeld radiation condition was applied for the transmission condition of the lateral boundary (Orlanski, 1976).

2.2. Boundary Condition Settings

Three boundary conditions were set: atmospheric forcing, lateral boundary conditions, and river discharge forcings. This simulation system, which includes data assimilation, is designed from the perspective of short-term forecasts. Therefore, the data used for the three boundary conditions were created using only data available in real time. Thus, more accurate data were not used for boundary conditions unless they could be obtained in real time. Thus, although a system that uses the output of an atmospheric simulation model as a boundary condition has since been developed for this numerical simulation model (Hafeez et al., 2021; Matsuzaki et al., 2021), this study adopted a system that creates boundary conditions based on observed values. Atmospheric forcings were created through spatial interpolation from data obtained from 12 observation stations on the land surrounding Ise Bay. The lateral boundary conditions were based on the average values from 10 years of monthly data. Owing to the use of climate values rather than real-time data, the water temperature and salinity values are considered to be significantly erroneous. The river discharge was calculated using a storage function method from precipitation in the basin. The precipitation was created using observation data. River water temperature was estimated from the air temperature near the mouth of the river using a regression line. Details of boundary condition settings are shown in the Appendix A.

2.3. Assimilation Model

The EnKF model for the Ise Bay Simulator was coded (Matsuzaki & Inoue, 2020) based on the work of Evensen (2003). The basic analysis steps of Kalman filter and EnKF are as follows. The analysis value, $\boldsymbol{x}^{\mathrm{a}}$ (data assimilation results), at a certain time, t, is obtained by the optimal weighted average of the background value $\boldsymbol{x}^{\mathrm{f}}$ (forecast/simulation results) and the observed value, $\boldsymbol{y}$, as shown by Equation 1.

$$\boldsymbol{x}_t^a = \boldsymbol{x}_t^f + \boldsymbol{K}_t \left(\boldsymbol{y}_t - \boldsymbol{H}_t \boldsymbol{x}_t^f \right) \tag{1}$$

where $\boldsymbol{x}$ represents a state vector whose elements are physical quantities such as water temperature, salinity, and current velocity in each mesh of the numerical model. $\boldsymbol{y}$ is an observation vector with observed values such as water temperature, salinity, and current velocity as elements. $\boldsymbol{H}$ is an observation matrix and an operator that extracts the physical quantity of the mesh corresponding to the observation point from the result (state vector) of the numerical model. $\boldsymbol{K}$ is a weight matrix (Kalman gain) that determines the extent that the observed values are assimilated and is calculated as follows.

$$\boldsymbol{K}_t = \boldsymbol{P}_t^f \boldsymbol{H}_t^T \left(\boldsymbol{R}_t + \boldsymbol{H}_t \boldsymbol{P}_t^f \boldsymbol{H}_t^T \right)^{-1} \tag{2}$$

where $\boldsymbol{P}$ is the background error covariance matrix, and $\boldsymbol{R}$ is the observation error covariance matrix. Superscript T means transpose. Equation 2 shows that the assimilation rate of the observed values is determined by the relationship between $\boldsymbol{P}$ and $\boldsymbol{R}$ and the spatial correction by the observed values is determined by $\boldsymbol{P}$. As mentioned in the introduction, the main purpose of this study is to calculate the appropriate ensembles for regional coastal estuary modeling, and this is dependent on the background error covariance matrix $\boldsymbol{P}$. The data assimilation flow in the EnKF is as follows:

1. Ensemble members are calculated using a numerical model; the Ise Bay Simulator. In this study, ensemble members are calculated under different boundary conditions, as explained in Section 2.4. The number of ensembles is expressed as L.
2. The background error covariance matrix is estimated from ensembles.

$$\bar{\boldsymbol{P}}_t^f = \frac{1}{L-1} \sum_{l=1}^{L} \left(\boldsymbol{x}_t^{f(l)} - \bar{\boldsymbol{x}}_t^f \right) \left(\boldsymbol{x}_t^{f(l)} - \bar{\boldsymbol{x}}_t^f \right)^T \tag{3}$$

$$\bar{\boldsymbol{x}}_t^f = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{x}_t^{f(l)} \tag{4}$$

Here, the mean value of the ensemble and the covariance matrix are represented by an overbar.

3. The observation error covariance matrix is estimated from ensembles.

$$\bar{\boldsymbol{R}}_t = \frac{1}{L-1}\sum_{l=1}^{L} \left(\boldsymbol{r}_t^{(l)} - \bar{\boldsymbol{r}}_t\right)\left(\boldsymbol{r}_t^{(l)} - \bar{\boldsymbol{r}}_t\right)^T \tag{5}$$

$$\bar{\boldsymbol{r}}_t = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{r}_t^{(l)} \tag{6}$$

where $\boldsymbol{r}$ is the observation error, which is the value of the measurement error that occurs depending on the characteristics of the observation equipment and the measurement environment, and the expression error that is caused by the phenomenon that the numerical model cannot express.

4. The Kalman gain is estimated as follows

$$\overline{\boldsymbol{K}}_t = \overline{\boldsymbol{P}}_t^f \boldsymbol{H}_t^T \left(\overline{\boldsymbol{R}}_t + \boldsymbol{H}_t \overline{\boldsymbol{P}}_t^f \boldsymbol{H}_t^T\right)^{-1} \tag{7}$$

5. The analysis values of ensemble members are calculated as follows.

$$\boldsymbol{x}_t^{a(l)} = \boldsymbol{x}_t^{f(l)} + \overline{\boldsymbol{K}}_t \left(\boldsymbol{y}_t + \boldsymbol{r}_t^{(l)} - \boldsymbol{H}_t \boldsymbol{x}_t^{f(l)}\right) \ (l = 1, \ldots, L) \tag{8}$$

The averaged value, $\bar{\boldsymbol{x}}_t^a$, of the analysis values of each ensemble member is the data assimilation result.

$$\bar{\boldsymbol{x}}_t^a = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{x}_t^{a(l)} \tag{9}$$

The settings for the ensemble simulation were the same as those described in Section 2.1, and a novel data assimilation method with a high-resolution horizontal grid size (800 m) was employed. EnKF was implemented with 32 members ($L = 32$). The ensemble number was selected as shown in a previous study (Matsuzaki & Inoue, 2020). The observation data described below were assimilated once per day at 00:00, and localization technique (Evensen, 2009; Gaspari & Cohn, 1999; Hamill et al., 2001) was not applied; thus, it was possible to correct the entire Ise Bay based on the background error covariance using a physical model, instead of nonphysical techniques such as the distance function. The multiplicative inflation technique was not applied because multiplicative inflation generates nonphysical and spurious error covariance (Sanikommu et al., 2020). Moreover, correlation of the observation error was ignored, that is, the observation error covariance matrix was set to diagonal. As explained in Section 2.4, perturbations were added to the boundary conditions to represent the system error. When assimilation was performed near the lateral boundary, the assimilation system became unstable. To stabilize the data assimilation, the variables at two meshes adjacent to the lateral boundary were excluded from the Kalman gain calculation and assimilation, which ensured stable data assimilation performance.

2.4. Method of Adding Perturbations to Boundary Conditions

Generating a perturbation and determining its magnitude is a challenging task. Previous research has employed various methods to determine the boundary conditions for expressing an ensemble containing system noise, for example: (a) a method of adding noise according to a normal distribution (Turner et al., 2008), (b) a method of adding red noise (Evensen, 2003; Sakov et al., 2012), (c) a method of using ensemble simulation results (Bougeault et al., 2010) as the boundary condition (Sanikommu et al., 2020), and (d) a method that considers the difference in state quantities at different times as a perturbation (Kunii & Miyoshi, 2012). This study employed method (a), as shown in Equation 10, because it was previously used to conduct successful data assimilation for a coastal estuary; however, the study of Turner et al. (2008) employed OSSE instead of real data.

$$\boldsymbol{F}_{\text{mem}} = \boldsymbol{F}_{\text{base}} + \boldsymbol{v} \tag{10}$$

Table 1
Magnitude of Perturbations to Boundary Conditions

Boundary condition		Method	ξ
Atmospheric forcing	Air temperature	Equation 10	3.04°C
	Wind speed	Equation 10	3.45 m s^{-1}
Lateral boundary conditions	Water temperature	Equation 10	0.73°C
	Salinity	Equation 10	0.20
River discharge forcing	River discharge	Equation 11	0.35
	River water temperature	Equation 10	1.21°C

Note. Calculation of ξ values is shown in Appendix B.

Here, $\boldsymbol{F}_{\text{mem}}$ indicates the boundary conditions for data assimilation with perturbation, $\boldsymbol{F}_{\text{base}}$ indicates the boundary conditions for numerical simulation, and $\boldsymbol{v}$ indicates the perturbations that have a normal distribution with a mean of zero and variance of ξ^2. For some boundary conditions, such as that shown in Equation 10, the additive method is not valid. For example, the boundary condition of river discharge may have a negative value when the river discharge is close to zero and noise with a normal distribution is added. In addition, when river discharge is larger, the error of river discharge forcing appears to decrease relatively. Thus, the following multiplication method was introduced:

$$\boldsymbol{F}_{\text{mem}} = (1 + \boldsymbol{v})\boldsymbol{F}_{\text{base}} \tag{11}$$

The model outputs evaluated in this study, which are explained in Section 2.7, are water temperature and salinity. The boundary conditions considered having a large effect on the simulation error of water temperature and salinity were selected as follows. For the numerical simulation model, the atmospheric forcing includes air temperature, shortwave radiation, longwave radiation, atmospheric pressure, wind direction, wind speed, water vapor pressure, and precipitation. The lateral boundary conditions include water temperature, salinity, and water level. The river discharge forcing boundary conditions include river discharge and river water temperature. Of these, the errors in the boundary conditions of air temperature, shortwave radiation, longwave radiation, lateral boundary water temperature, and river water temperature were considered directly linked to the numerical simulation error of water temperature. Similarly, the errors in the boundary conditions of precipitation, lateral boundary salinity, and river discharge were considered directly linked to the numerical simulation error of salinity. In addition, as water temperature and salinity are advected and diffused by the flow of water mass, the errors in the boundary conditions of wind speed, atmospheric pressure, and tide level of the lateral boundary were also considered having an effect. For these boundary conditions, these three assumptions were set. First, the shortwave radiation and longwave radiation errors were included in the air temperature error. Second, the precipitation error was included in the river discharge error. Third, the influence of the error between the atmospheric pressure and the tide level of the lateral boundary was relatively small, so it was ignored. Therefore, the perturbations for atmospheric forcing were air temperature and wind speed (wind direction was not altered), the perturbations for lateral boundary conditions were water temperature and salinity, and the perturbations for river discharge forcing were river discharge and river water temperature.

As the magnitude of the error is considered to correlate with the accuracy of the boundary conditions, the magnitude of the perturbation, ξ, for creating the ensemble must be determined by the same method used to generate the boundary conditions. In this study, ξ values were estimated according to the assumption that all calculated error distributions follow a normal distribution; ξ values calculated on a trial basis are shown in Appendix B and summarized in Table 1. When the normal distribution is expressed by a normal random number with a few members, there is the potential for a large deviation from the normal distribution due to sampling error. In this study, we did not use normal random numbers, but set the value of each member to match the cumulative value of the normal distribution so the normal distribution can be expressed even with a few members. To avoid unintended correlation of each boundary condition, the Fisher-Yates shuffle (Fisher & Yates, 1948) was used to perform 10,000 replacement attempts, and the boundary conditions were set for each ensemble member using the combination with the lowest correlation.

Table 2
Assimilated Observation Data

No.	Station name	Latitude (ºN)	Longitude (ºE)	Observation type	Observation depth [m]
A1	Back of Ise Bay	34.926	136.741	Automatic elevating	Every 1.0 m
A2	Center of Ise Bay	34.669	136.841	Automatic elevating	Every 1.0 m
A3	Mouth of Ise Bay	34.509	137.018	Fixed	1.0 m, 11.8 m, and 23.2 m from low water level
A4	No. 1 buoy	34.743	137.220	Automatic elevating	Every 1.0 m
A5	No. 2 buoy	34.745	137.072	Automatic elevating	Every 1.0 m

2.5. Assimilated Observations

In situ water temperature and salinity profiles observed at fixed points were used for the data assimilation. Seven in situ observation stations are in operation in Ise Bay. Data from the five observation stations in Table 2 were assimilated. The observation time of the assimilated value is 00:00, and is not the 24-hr average. Observation error variance values were set to $(1.0°C)^2$ for water temperature and $(1.0)^2$ for salinity. These values were set referring to a previous study (Matsuzaki & Inoue, 2020). Gross error check was performed as background quality control. The difference between the observed value and the first guess value was calculated, and if the difference was over 3°C for water temperature, and 6 for salinity; the observed value was rejected.

2.6. Experimental Setup

Experiments were conducted for six cases (Table 3). In the standard experiment (control run [CR]), data assimilation was not applied, that is, CR was a normal numerical simulation. DAAll included the optimal settings determined before the experiment. Perturbations were applied to three boundary conditions: atmospheric forcing, lateral boundary conditions, and river discharge forcing. The other four experiments included the assimilation results but used different methods of generating the ensembles. These experiments were conducted to confirm the effect of adding perturbations to the boundary conditions by comparing the results with those of DAAll. DAwoAtm had the same conditions as DAAll but did not perturb the atmospheric forcing of air temperature and wind speed, it analyzed the effect of considering the uncertainty of atmospheric forcing on the data assimilation results. As DAwoWind applied perturbations to air temperature but not to wind speed, it isolated the effects of air temperature and wind speed among the atmospheric forcing boundary conditions. Finally, as DAwoLBC and DAwoRiv had the same conditions as DAAll but did not perturb the lateral boundary conditions or river discharge forcing, these experiments examined the effect of considering the uncertainty of lateral boundary conditions and river discharge forcing on the data assimilation results. The assimilation experiments were conducted for 1 year from 1 January 2016, to evaluate the applicability of the proposed method to long-term fluctuations, including seasonal changes, and to verify the robustness of the data assimilation method. Initial ensembles for the assimilation experiments on 1 January 2016 were generated using an 8-month spin-up period from 1 April 2015. In the spin-up period, the ensemble members were calculated under the boundary conditions including the

Table 3
Experimental Conditions

Experiment	Assimilation	Atmospheric forcing		Lateral boundary condition	River discharge forcing
		Air temperature	Wind speed		
CR	Control run without DA	NA	NA	NA	NA
DAAll	Assimilated	Perturbed	Perturbed	Perturbed	Perturbed
DAwoAtm	Assimilated	Not perturbed	Not perturbed	Perturbed	Perturbed
DAwoWind	Assimilated	Perturbed	Not perturbed	Perturbed	Perturbed
DAwoLBC	Assimilated	Perturbed	Perturbed	Not perturbed	Perturbed
DAwoRiv	Assimilated	Perturbed	Perturbed	Perturbed	Not perturbed

Table 4
Comparison Observation Data Not Assimilated

No.	Station name	Latitude (ºN)	Longitude (ºE)	Observation type	Observation depth
C1	Nakayama Channel	34.623	136.982	Fixed	1.4 m, 8.2 m, 12.4 m from low water level
C2	No. 3 buoy	34.675	137.097	Automatic elevating	Every 1.0 m

perturbations, and exhibited an ensemble spread according to the position and magnitude of the perturbation of the initial conditions.

2.7. Accuracy Validation

Water temperature and salinity data of the model output were compared with the in situ observation data of water temperature and salinity profiles observed at fixed points (Table 4). Unfortunately, a suitable map data set for salinity similar to SST does not exist, making the comparison more difficult. As the next best measure, the model outputs and data used for the assimilation (Table 2) were compared to evaluate and discuss the effects of perturbing boundary conditions to generate ensembles. Observation data were collected every hour, but the assimilations were conducted every day; thus, comparisons were conducted every day. The in situ observation time of the value was 00:00, and it is not the 24-hr average. Water temperature data of the model output were also compared with the SST data observed by Terra and Aqua (Moderate Resolution Imaging Spectroradiometer: MODIS) to evaluate the correction of water temperature in the spatial direction. The MODIS SST data were obtained by assuming that all data observed between 22:00 and 02:00 were observed at midnight; the SST data were then compared with model output data for assimilation. The reproducibility of the planar distribution of water temperature was then discussed.

The accuracy of the model output was evaluated using the following skills: the bias (Equation 12) and the root-mean-square error (RMSE, Equation 13):

$$\text{bias} = \frac{1}{N}\sum_{i=1}^{N}(m_i - o_i) \tag{12}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2} \tag{13}$$

where m_i and o_i are model output and observation, respectively, and N is the number of model outputs and observations. SST was also evaluated using centered (bias removed) RMSE (CRMSE, Equation 14) and correlation coefficient (CC, Equation 15):

$$\text{CRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(m_i - \overline{m}\right) - \left(o_i - \overline{o}\right)\right]^2} \tag{14}$$

$$\text{CC} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(m_i - \overline{m}\right)\left(o_i - \overline{o}\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(m_i - \overline{m}\right)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(o_i - \overline{o}\right)^2}} \tag{15}$$

where $\bar{m}$ and $\bar{o}$ are the average value of the model output and observation, respectively.

The assimilation result of EnKF was obtained by the weighted average of the background value and the observed value. The weighted average was determined by the ensemble spread, which was the amount of error in the background value. Therefore, degeneration of the ensemble spread becomes a problem when executing EnKF (termed as filter divergence). Then, the magnitude of the ensemble spread was evaluated (Equation 16):

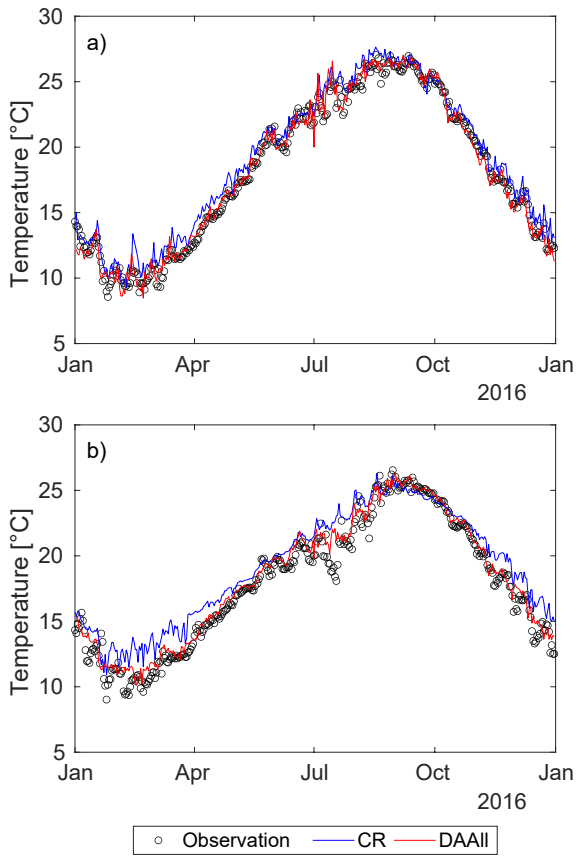

Figure 3. Time series of water temperature data at C1 for observations, CR, and DAAll. (a) Water depth of 1.0 m; (b) water depth of 12.0 m.

$$\text{Ensemble spread} = \sqrt{\frac{1}{L-1}\sum_{i=1}^{L}\left(m_i - \overline{m}_L\right)^2} \tag{16}$$

where $\bar{m}_L$ is the ensemble average value of m_i.

3. Results

3.1. Performance and Robustness of Data Assimilation

In this section, the results of the CR and data assimilation (DAAll) are compared to show the validity and effectiveness of the data assimilation method. Figures 3 and 4 compare the time series of observed water temperature data in C1 and C2 (Table 4) and the model output of CR and DAAll. CR exhibits the same water temperature fluctuation trend as the observed values; however, the water temperature is higher than the observed values. This difference is particularly large in the lower layer. DAAll shows the water temperature corrected to match the observations. Moreover, DAAll was able to perform the data assimilation for 1 year without the divergence of numerical operation. Figures 5 and 6 show the biases and RMSEs between the observed and simulated water temperatures for C1 and C2. Bias and RMSE values are lower for DAAll than CR at all depths. The bias improvement is approximately the same near the water surface and near the bottom, with an average difference between CR and DAAll of 0.78°C for C1 and 1.09°C for C2. Conversely, the RMSE improvement is greater near the bottom than near the sea surface. The average difference between CR and DAAll is 0.57°C for C1 and 0.86°C for C2. These results indicate that the proposed regional data assimilation method for a coastal estuary is effective for correcting water temperature and highly robust, that is, it can be applied throughout the year and reflects seasonal variations.

Figures 7–10 show the spatial distributions of bias, RMSE, CRMSE, and CC values between SST data observed by MODIS and the model outputs, respectively. The bias, RMSE, and CRMSE scores of DAAll are better than those of CR throughout Ise Bay, particularly at the west side of the bay, although the observed values used for data assimilation extend from the center of the bay to the east side, and there are no observation points on the west side. Data assimilation corrects the water temperature for the entire bay, despite sparse observations in the horizontal direction, because the error covariance is properly expressed by the proposed perturbation. The CC values of DAAll have almost the same distribution as that of CR. The bias, RMSE, and CRMSE values of SST were 0.67°C, 0.54°C, and 0.34°C lower, respectively, in DAAll (Figure 11). Nevertheless, the bias, RMSE, and CRMSE values do not exhibit substantial improvement on the east side of the bay mouth and in parts of the river mouth. Thus, there is still room for further improvement. The RMSE and CRMSE of the simulation and data assimilation results were high in the river mouth where the three largest class A rivers flow (around 55 mesh in the x-direction and 75 mesh in the y-direction). One reason may be that in the case of the simulation, the bias of the river water temperature in the boundary conditions was small, but the variation against actual river water temperature was large; hence, the RMSE and CRMSE were high. In the case of data assimilation, the observation station was not near the river mouth and the correlation between the observation station and the river mouth was low; therefore, it was probable that the correction was not made.

Figures 12 and 13 show time series of the observed salinity in C1 and C2 and the model outputs of CR and DAAll. Although the effect of assimilation on salinity is not as clear as that for water temperature, the assimilation performance is stable throughout the year. Figures 14 and 15 show the bias and RMSE values of salinity in C1 and C2. Bias and RMSE values decrease from CR to DAAll at all depths in C1. The average difference in bias and RMSE values between the two experiments are 0.17 and 0.06, respectively. At C2, the bias values are lower at all depths in DAAll; however, the RMSE values do not show this trend; the average difference in bias and

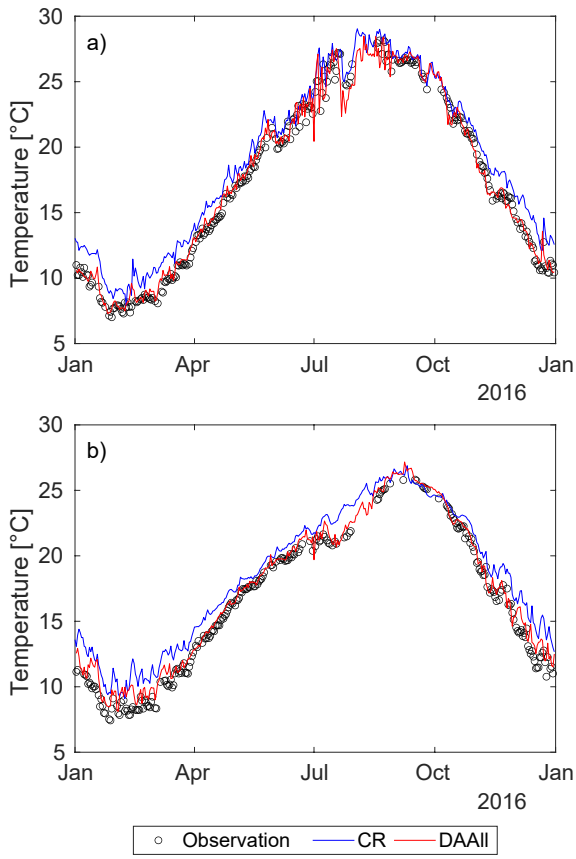

Figure 4. Time series of water temperature data at C2 for observations, CR, and DAAll. (a) Water depth of 1.0 m; (b) water depth of 12.0 m.

RMSE values between the two experiments are 0.07 and −0.09, respectively. One reason for this finding could be that the magnitude of perturbations (ξ) for assimilation of salinity data was not appropriate in the boundary conditions. When data assimilation is performed by only changing the magnitude of the perturbation of the boundary conditions from DAAll (the results of the sensitivity experiments are not shown, but ξ was set to 1.00°C for air temperature, 2.00 m s^{-1} for wind speed, 0.50°C and 0.25 for water temperature and salinity of the lateral boundary, 0.36 for river discharge, and 0.50°C for river water temperature), the average RMSE of salinity at C2 is 0.01 smaller for DAAll than CR. Therefore, the optimal magnitude of perturbation should be carefully considered. Nevertheless, the results indicate that the proposed regional data assimilation method for coastal estuaries is an effective and robust method for both water temperature and salinity data.

3.2. Effect of Perturbations on Boundary Conditions

3.2.1. Atmospheric Forcing

This subsection examines the effect of perturbation on atmospheric forcing on the data assimilation results. Compared to DAAll, DAwoAtm, which does not perturb the air temperature and wind speed, does not improve the bias and RMSE values of water temperature in C1 and C2 (Figures 5 and 6). This finding is particularly noticeable at C2. DAwoAtm was the least improved among the data assimilation results for the bias and RMSE scores of water temperature (Figures 5 and 6). Additionally, DAwoAtm was the least improved among the data assimilation results for the bias, RMSE, and CRMSE scores of SST (Figure 11). However, DAwoWind, which perturbs the atmospheric forcing condition of air temperature, improves the water temperature from that of DAwoAtm (Figures 5, 6, and 11). DAwoWind also exhibits better bias and RMSE scores than DAAll at a depth of −4 m or more at C2, and better bias scores at a depth of −10 m or more in C1. On the other hand, DAwoWind does not exhibit improvements from DAAll at the other depths, in the SST in Mikawa Bay on the east side of Ise Bay (Figures 7 and 8), or in the SST bias and RMSE scores (Figure 11). Moreover, DAwoWind has a small improvement from CR in the SST CRMSE. Also, DAwoWind was worse than CR in SST CC. Therefore, the scores of DAAll are generally better than those of DAwoWind. The ensemble spread of water temperature in the C1 (Figure 16) is smaller in DAwoAtm than in DAAll, especially in the surface layer. Moreover, the ensemble spread of DAwoWind is larger than of DAwoAtm, but smaller than of DAAll. At C2 (Figure 17), the ensemble spread of DAwoAtm is even smaller than in C1; thus, it is considered that the boundary conditions of air temperature and wind speed is a large error factor. Thus, perturbation of the atmospheric boundary conditions increases the ensemble spread of water temperature, especially near the surface layer, enabling effective assimilation of observed water temperature values.

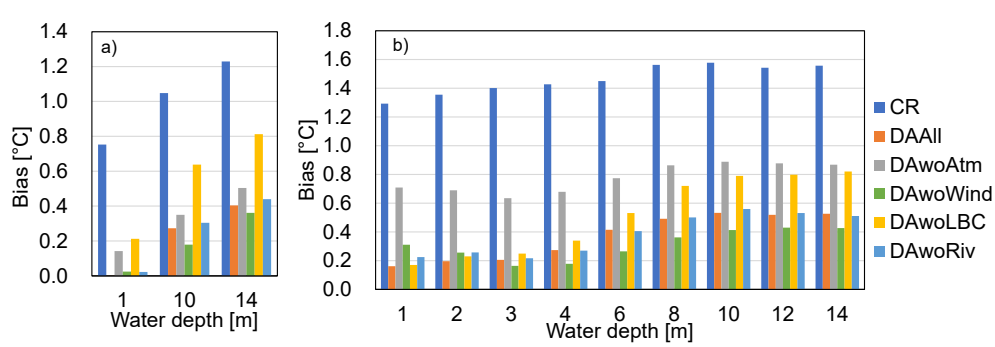

Figure 5. Bias between observed and modeled water temperature for all experiments. (a) C1; (b) C2.

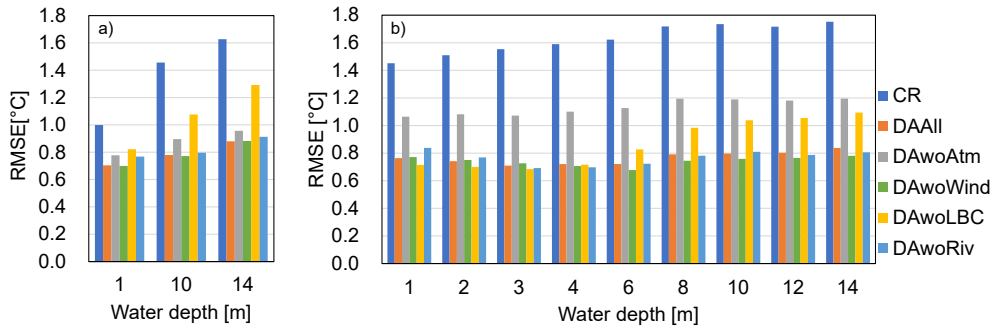

Figure 6. RMSE between observed and modeled water temperature for all experiments. (a) C1; (b) C2.

For salinity (Figures 14 and 15), DAwoAtm and DAwoWind exhibit similar bias and RMSE scores to DAAll. However, at C2 (at a water depth of −4 m or more in Figure 14) and in the center of the bay (at a water depth of −10 m or less in Figure 18), the bias score is significantly worse. Therefore, it is considered preferable to perturb atmospheric forcing to avoid local salinity errors in the data assimilation. The difference in the ensemble spread of salinity is small between DAwoAtm and DAwoWind (Figures 20 and 21). In addition, the ensemble spread of DAwoAtm and DAwoWind is smaller than of DAwoLBC and DAwoRiv (Figures 20 and 21), particularly at C2. These results indicate that, among the boundary conditions, wind speed has the greatest influence on the magnitude of the salinity ensemble spread and can be dominant depending on the location.

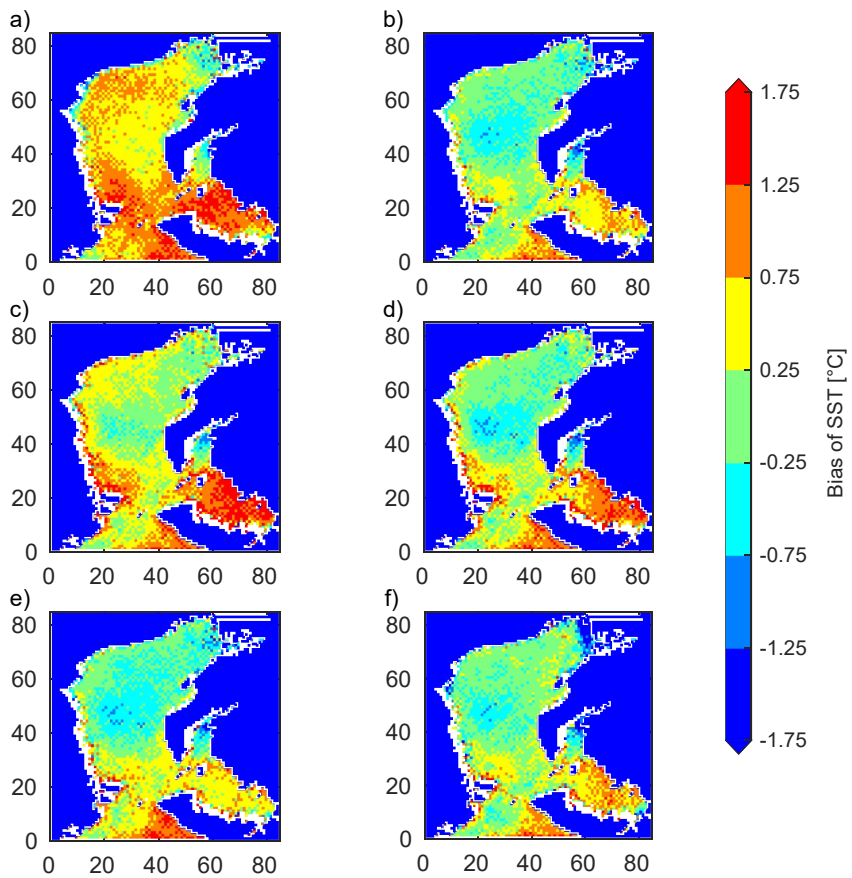

Figure 7. Planar images of the sea surface temperature bias for (a) CR, (b) DAAll, (c) DAwoAtm, (d) DAwoWind, (e) DAwoLBC, and (f) DAwoRiv. The horizontal and vertical axes indicate a calculation grid of x and y-direction and the grid number is 85 × 85.

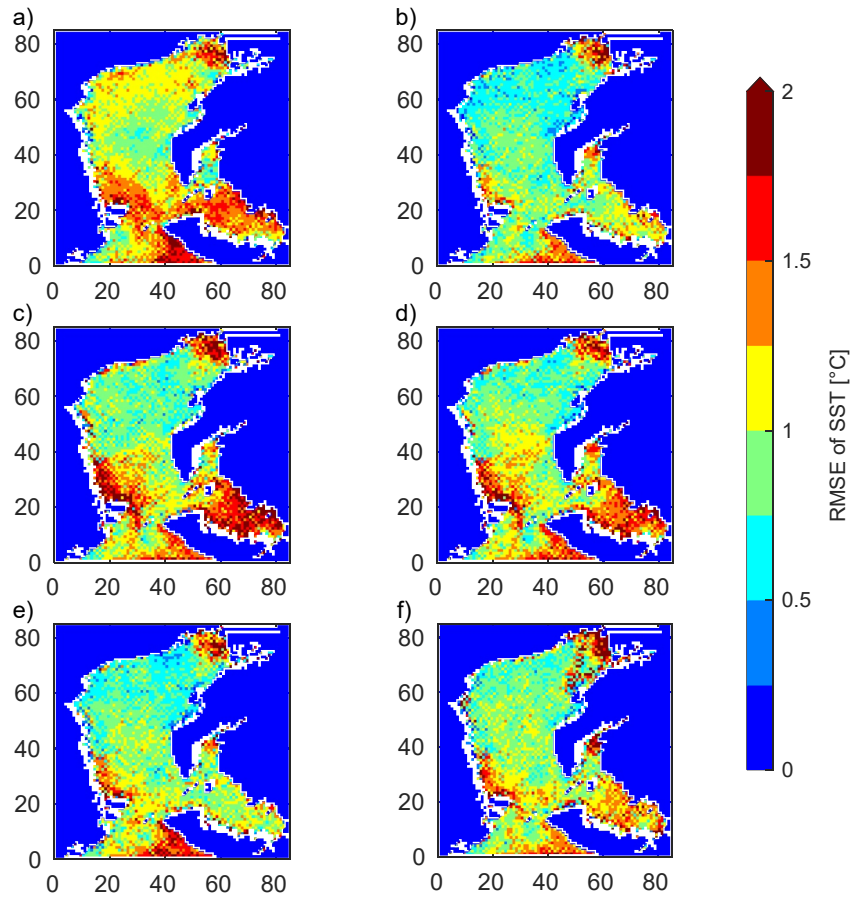

Figure 8. Planar images of the RMSE values of sea surface temperature for (a) CR, (b) DAAll, (c) DAwoAtm, (d) DAwoWind, (e) DAwoLBC, and (f) DAwoRiv. The horizontal and vertical axes indicate a calculation grid of x and y-direction and the grid number is 85 × 85.

3.2.2. Lateral Boundary Conditions

This subsection examines the effect of perturbation to the lateral boundary conditions on the data assimilation results. In DAwoLBC, which does not perturb the lateral boundary conditions, the bias and RMSE scores of water temperature in C1 and at C2 are not improved by data assimilation compared to those of DAAll (Figures 5 and 6). This finding is particularly noticeable in the C1. DAwoLBC also shows that the bottom water temperature errors increase whereas surface water temperatures improve. This improvement is due to perturbations of atmospheric forcing (Figure 16). DAwoLBC exhibits the least improvement in bias and RMSE scores among all data assimilation results in C1 and a slight improvement in SST scores around the bay mouth (Figures 7 and 8). However, when looking at the entire Ise bay, the effect of perturbing the lateral boundary conditions was small for the magnitude of the SST error (CRMSE: Figure 9) and variation pattern (CC: Figure 10). The ensemble spread of water temperature was smaller for DAwoLBC than for DAAll for all water depths in C1 (Figure 16). The large ensemble spread for DAwoLBC from January to March and in December is thought to be because of perturbing the atmospheric boundary conditions because the ensemble spread for DAwoAtm during the same period is small. However, at C2, there is minimal difference in the ensemble spread between DAAll and DAwoLBC (Figure 17). Therefore, perturbation of the lateral boundary conditions increases the ensemble spread of water temperature at all water depths, especially near the bay mouth, and enables the effective assimilation of observed values.

DAwoLBC exhibits lower bias and RMSE scores for salinity than DAAll in C1 (Figures 14 and 15) and at the mouth of the bay (Figures 18 and 19). The ensemble spread of salinity is smaller at C1 (Figure 20), which is similar to the results of water temperature. Again, there is almost no difference in the ensemble spread between DAAll and DAwoLBC at C2 (Figure 21). Therefore, as with water temperature, perturbation of the lateral boundary

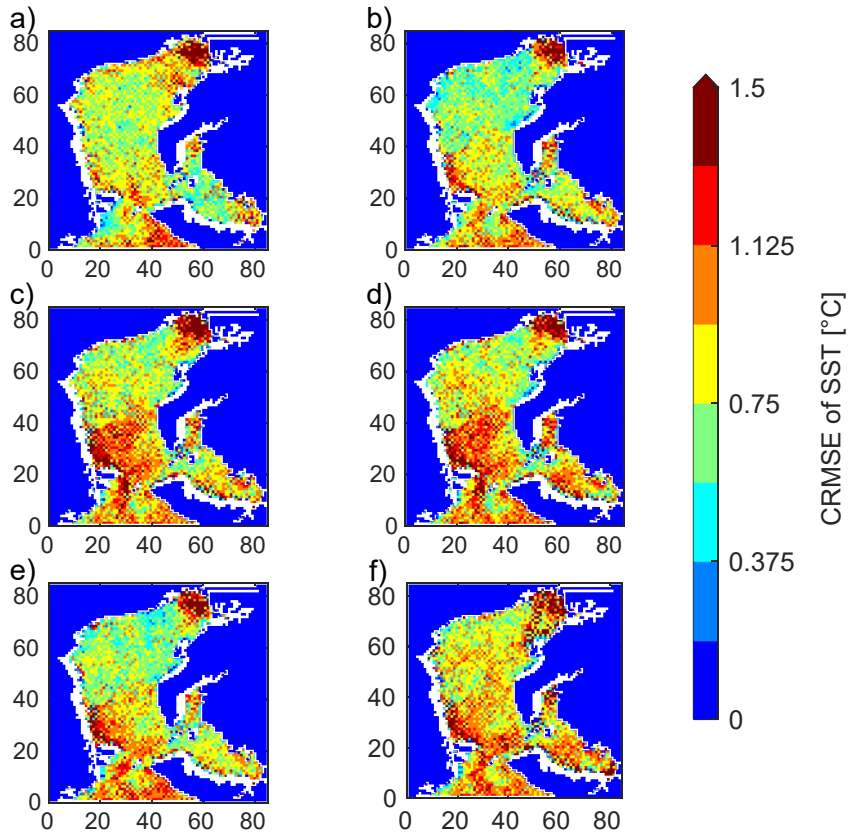

Figure 9. Planar images of the CRMSE values of sea surface temperature for (a) CR, (b) DAAll, (c) DAwoAtm, (d) DAwoWind, (e) DAwoLBC, and (f) DAwoRiv. The horizontal and vertical axes indicate a calculation grid of x and y-direction and the grid number is 85 × 85.

conditions increases the ensemble spread at all water depths, especially near the bay mouth, and enables the effective assimilation of observed values.

3.2.3. River Discharge Forcing

DAwoRiv, which does not perturb the river discharge forcing, shows a similar improvement in the bias and RMSE scores of water temperature from those of DAAll in C1 and at C2 (Figures 5 and 6). However, the bias, RMSE, CRMSE, and CC scores of SST are worse than those of DAAll in the inner part (near river mouth: Figure 2) of the bay (Figures 7–10). The ensemble spread of water temperature for DAwoRiv and DAAll show similar trends in C1 and at C2 (Figures 16 and 17). This result indicates that the effect of perturbing river discharge forcing is particularly large near the river mouth and decreases with distance from the river mouth. Therefore, perturbation of river discharge forcing ensures appropriate assimilation of water temperature data in the coastal estuary.

For salinity, the RMSE score of DAwoRiv is worse than DAAll at the back of bay (Figure 19). Like water temperature, the error of the river boundary conditions has an increasing influence on salinity with proximity to the river mouth. Moreover, it is necessary to perturb river discharge forcing to improve the data assimilation results, especially near the river mouth.

4. Discussion

4.1. Performance and Robustness of Data Assimilation

Previous studies have not examined the long-term applicability of regional data assimilation methods for coastal estuaries, nor their ability to reflect seasonal fluctuations. Moreover, although EnKF has been applied to OSSEs, before this study, it had not been applied to actual observation data from coastal areas. In this study, the proposed

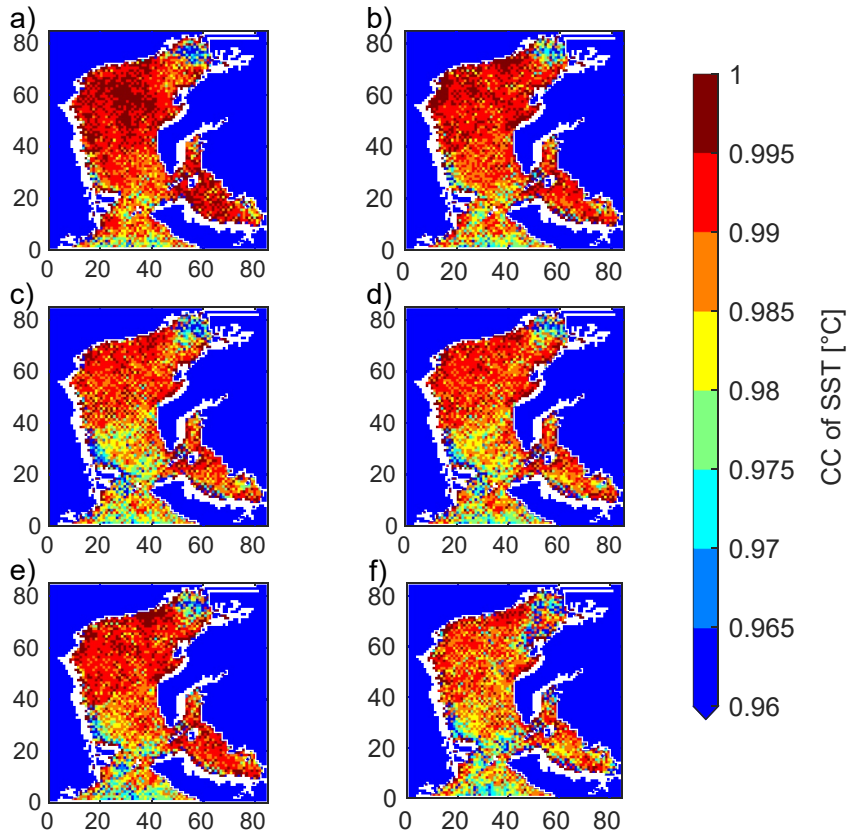

Figure 10. Planar images of the CC values of sea surface temperature for (a) CR, (b) DAAll, (c) DAwoAtm, (d) DAwoWind, (e) DAwoLBC, and (f) DAwoRiv. The horizontal and vertical axes indicate a calculation grid of x and y-direction and the grid number is 85 × 85.

EnKF method achieved stable assimilation results for both water temperature (Figures 3 and 4) and salinity (Figures 12 and 13) throughout the year, and reflected seasonal fluctuations. Thus, the proposed regional data assimilation method for coastal estuaries exhibits good applicability and robustness. The assimilation of water temperature (Figures 5 and 6) and salinity (Figures 14 and 15) data contributed to error correction in the vertical direction (i.e., with water depth). Water temperature was also corrected in the horizontal direction (Figures 7 and 8). This is because the error covariance was appropriately expressed by generating ensembles using the proposed method of perturbing boundary conditions.

4.2. Effect of Perturbations to Boundary Conditions

In comparison to the open ocean, accuracy of lateral boundary conditions and river discharge forcing are relatively more important in a coastal estuary. However, due to inadequate observation data, it is difficult to provide accurate boundary conditions, causing substantial errors in coastal numerical simulations. Therefore, in this study, a perturbation was applied to the three boundary conditions. Although the ensemble spread generally tends to degenerate in coastal estuary modeling, this was avoided by applying perturbations to lateral boundary conditions and river discharge forcing (Figures 16, 17, 20, and 21). Although perturbations are often applied to atmospheric forcing in ocean data assimilation methods, this is the first study to indicate the importance of applying perturbations to lateral boundary conditions and river discharge forcing in regional data assimilation for a coastal estuary. The SST bias and RMSE scores of DAwoAtm, DAwoWind, DAwoLBC, and DAwoRiv were better than CR (Figures 7 and 8). However, SST CRMSE score of DAwoAtm, DAwoWind, DAwoLBC, and DAwoRiv has only a small improvement. Furthermore, the CC score of all data assimilation results excluding DAAll was worse than CR. These results show that data assimilation without the appropriate perturbations does not contribute to the improvement of error without bias (CRMSE: Figure 9), but rather worsens the pattern of

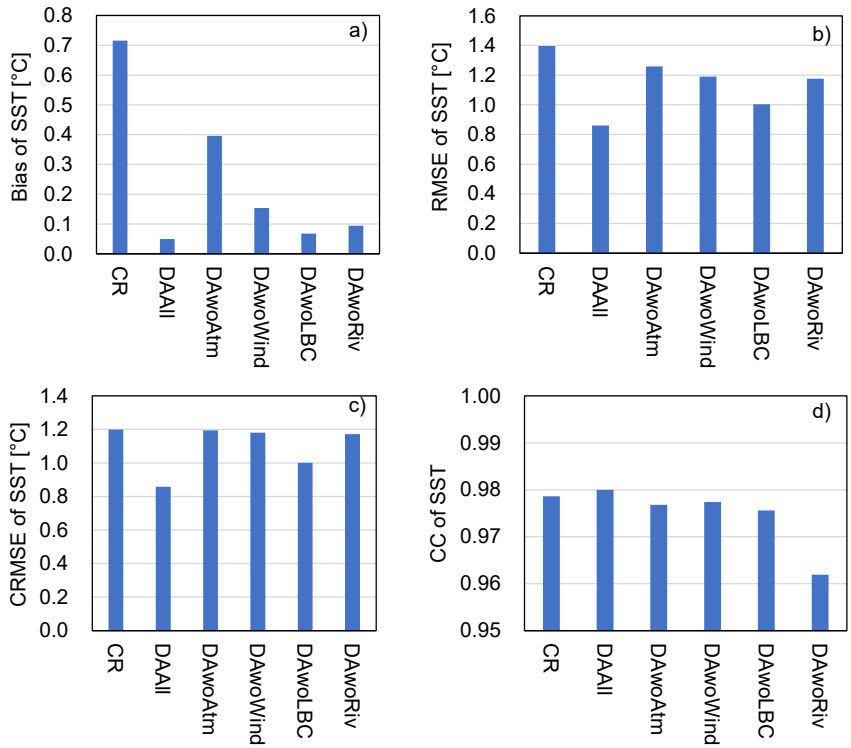

Figure 11. (a) Bias, (b) RMSE, (c) CRMSE, and (d) CC values of sea surface temperature for all experiments.

variation (CC: Figure 10). Thus, these results suggest that it is necessary to create a background error covariance matrix that recognizes the main error factors of numerical simulations when performing regional data assimilation in a coastal estuary.

Water temperature reproducibility improved when both air temperature and wind speed were perturbated. In other words, perturbing only one of them does not improve the water temperature. This could be because the air temperature affects the heating/cooling of the water temperature directly under the calculation grid, which inputs as atmospheric forcing. Considering only the temperature error, the water mass whose water temperature has been corrected does not advect and diffuse accurately; hence, a sufficient assimilation effect cannot be obtained. Wind also affects the advection and diffusion of water masses. Advection/diffusion of water mass is corrected by considering mainly the wind speed error. However, a good assimilation effect cannot be obtained because heating/cooling from the atmosphere is not calculated accurately. This is supported by the fact that salinity data assimilation is greatly affected by wind speed perturbations and not by temperature perturbations.

Within the possible magnitude of error to boundary conditions (Appendix B), wind speed and air temperature were shown to be important for water temperature and salinity correction. That is, the error of the boundary condition between wind speed and temperature is more sensitive to simulation accuracy than other boundary conditions. Therefore, it can be said that it is necessary to carefully set the boundary conditions of wind speed and temperature to accurately simulate the entire Ise Bay. Mikawa Bay is shallower than the west side of Ise Bay (Figure 2); relatively shallow water is considered to be more susceptible to wind accuracy, and results show that this is partly attributed to wind effect.

Vervatis et al. (2021) noted that in the open ocean, perturbing the wind speed had the greatest effect on the ensemble spread of water temperature during data assimilation by EnKF, and that perturbation of other atmospheric forcing conditions (air temperature and sea level pressure) was less dominant. They also reported that wind uncertainty had a significant impact on upper ocean uncertainty for both the geostrophic and Ekman components defined by Sverdrup dynamics. In contrast, in our regional data assimilation for coastal estuaries, perturbation of both the wind speed and air temperature was important for the ensemble spread of water temperature (Figures 16 and 17). These results show the difference between open ocean and coastal modeling. Figure 4 (b) in Vervatis

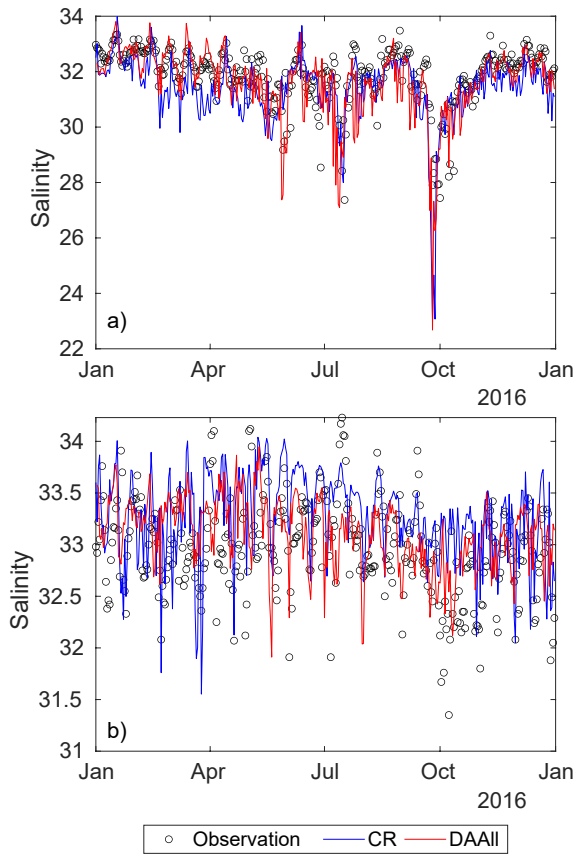

Figure 12. Time series of salinity at C1 for observations, CR, and DAAll. (a) Water depth at 1.0 m; (b) water depth at 12.0 m.

et al. (2021) shows that the ensemble spread caused by perturbation of the air temperature was large near the coastline (coastal area). A possible reason is that the perturbation of air temperature becomes relatively significant in the shallow sea area because the amount of water mass in the shallow sea area with respect to the heat flux of the atmosphere and seawater is smaller than that in the deep sea area. Therefore, the effect of air temperature perturbations cannot be neglected during data assimilation in coastal areas.

In this study, we selected six boundary conditions (air temperature, wind speed, water temperature, salinity of the lateral boundary, river discharge, and river water temperature), which are considered the main error factors, instead of all the boundary conditions as variables that cause perturbations. As a result, data assimilation was carried out stably and effectively. This indicates the importance of adding perturbations to these six conditions. Moreover, it is necessary to examine the boundary conditions to which perturbations should be added to further improve the assimilation results.

In this study, the location where the perturbation was applied was examined, and the magnitude was obtained by error analyses through comparisons with observation data. According to the data assimilation results, the magnitude of perturbation was qualitatively appropriate. Therefore, the method of estimating the magnitude of the perturbation (Appendix B) is considered appropriate, and the error estimation method implemented in this study can be used for general purposes. However, this study did not evaluate the optimal magnitude of the perturbation; therefore, this should be considered in future work.

4.3. Future Work

The results here are a crucial first step in regional coastal data assimilation; however, many issues remain unresolved. Specifically, the correlation of different boundary conditions was set to be small to avoid unintended accidental correlations. However, we could not confirm that there were no problems with this setting. For example, the lateral boundary conditions of water temperature and salinity exhibit a certain correlation. Thus, it is necessary to verify the assimilation when the perturbation is applied according to the correlation obtained from observed values. Furthermore, the correlation coefficient between the discharge forcing of each river was set to 1, which is not the true value. Although the correlation for rivers with short distances between them is close to 1, rivers with long distances between them may require comparison of the observed river discharge and water temperatures to estimate the correlation coefficient.

Abundant observation data are obtained from satellite and in situ observations in coastal areas. However, the data assimilation method used in this study cannot simultaneously assimilate more observation data than ensemble members. Therefore, experiments with a greater amount of ensemble members are required to assimilate large amounts of observational data. Moreover, system error in this study was assumed to be constant, regardless of the time or season, and the perturbations (standard deviation ξ) of boundary conditions were set to constant values. Therefore, future research should examine whether the proposed data assimilation method is suitable for detailed event analysis (e.g., strong winds, large-scale floods, and water mass intrusion from the open ocean to the inner bay) where the model error, not the boundary conditions, has a significant effect.

Furthermore, confirmation of the reproducibility of salinity data was limited to a comparison of bias and RMSE scores using in situ observations, and the reproducibility of salinity distributions was not discussed. However, a method for calculating the highly accurate planar distribution of coastal areas using satellite observations has recently been developed (Nakada et al., 2018), which will be used to conduct salinity reproducibility analyses in future works.

Finally, instead of relying on data assimilation, it is necessary to improve the accuracy of boundary conditions. As described in Section 2.2, the accuracy of the boundary conditions is low because only real-time data were used for the three boundary conditions. The reason underlying the bias of the water temperature in the CR may

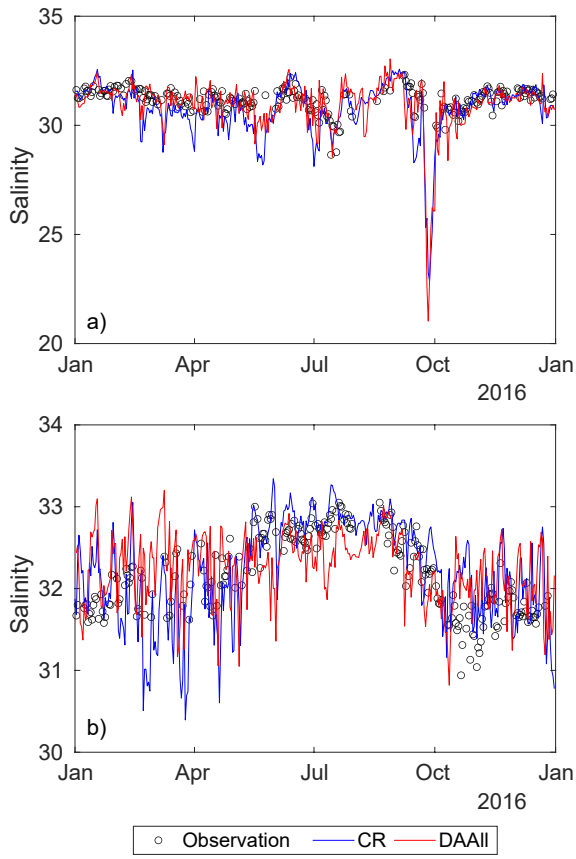

Figure 13. Time series of salinity at C2 for observations, CR, and DAAll. (a) Water depth at 1.0 m; (b) water depth at 12.0 m.

be due to the low accuracy of the boundary conditions, especially the bias of two boundary conditions: wind velocity of atmospheric forcing and water temperature of lateral boundary condition. The wind velocity of atmospheric forcing is set from observation data from the observation stations on land. Wind speed is corrected to fit on the water surface while considering the roughness of land/sea surface using the method reported by Kuwagata and Kondo (1990) (see Appendix A for details). However, the annual difference (bias) between the wind speed for the boundary condition and observed data at sea surface was −1.55 m/s on average. Therefore, it is highly possible that the wind speed of atmospheric forcing was set lower in CR than it actually was. A test simulation was carried out under the same conditions as those of CR, but only the wind velocity of the atmospheric forcing was replaced with the results of the mesoscale meteorological simulation by the Japan Meteorological Agency. As a result, from October to March, the bias of the water temperature of the bottom layer was improved. The bias of water temperature was presumed to be improved by changing the wind velocity of atmospheric forcing due to the vertical mixing of the cooling seawater near the water surface. However, the bias of water temperature from April to September was not improved by changes in atmospheric forcing. This bias from April to September could have presumably be caused by the water temperature under the lateral boundary condition. The observed average value for 10 years is input for the water temperature of the lateral boundary condition. The error of the water temperature of the lateral boundary condition is considered to be large. In fact, the observed water temperature in 2016, which is the calculation period, is lower than the water temperature under the lateral boundary condition. The results have not been organized because the test simulation with changing water temperature of the lateral boundary condition has not been performed. Therefore, we intend to study the importance of boundary conditions in future research.

And also, it is necessary to improve the simulation model. For example, the salinity bias is reversed between the surface and bottom layers in this study, which may be because the salinity of the model output is less diffused in the vertical direction than in reality. As the positive and negative biases are the same in the data assimilation results (Figures 14 and 18), it is necessary to modify the simulation model to consider diffusion in the vertical direction.

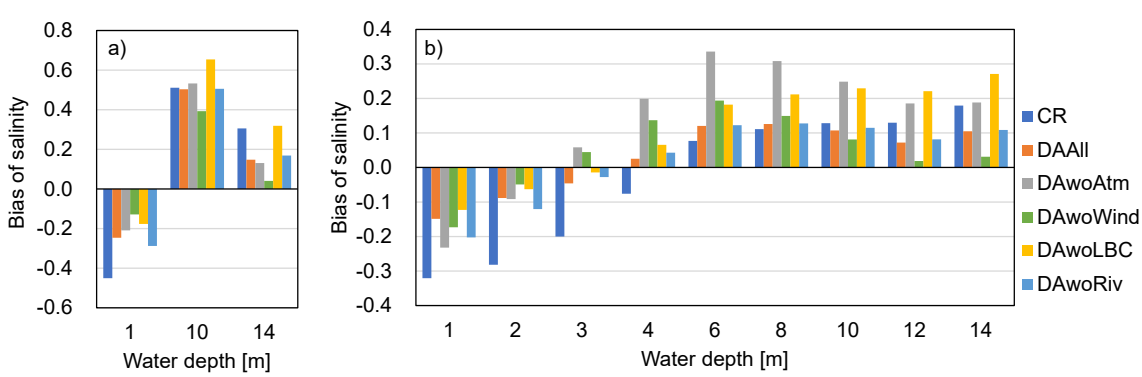

Figure 14. Bias between observed and modeled salinity for all experiments. (a) C1; (b) C2.

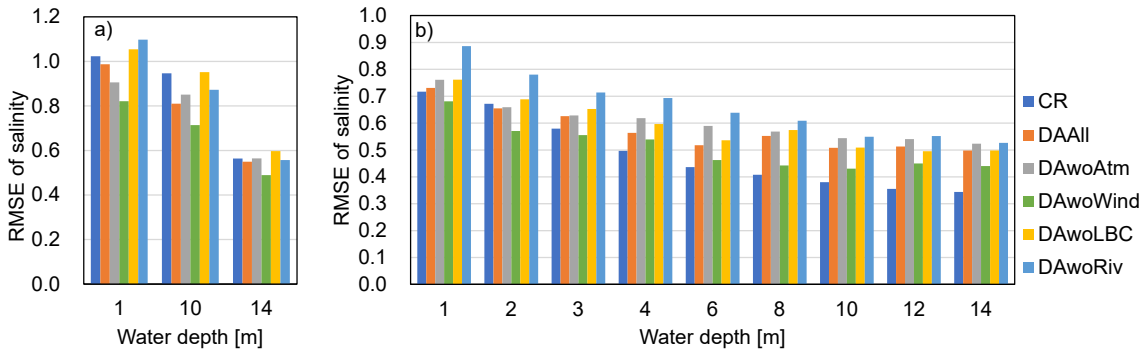

Figure 15. RMSE between observed and modeled salinity for all experiments. (a) C1; (b) C2.

5. Conclusions

There have been previous reported numerical experiments on data assimilation (OSSEs); however, this is the first study to apply the EnKF to regional data assimilation of coastal estuaries using actual long-term observation data. Specifically, data assimilation was performed for water temperature and salinity. According to comparisons with observation data not used in the assimilation, the simulated water temperature and salinity data were corrected in the horizontal and vertical directions (i.e., with water depth). In addition, the proposed method achieved stable long-term data assimilation over 1 year and responded to seasonal fluctuations. Besides perturbations to atmospheric forcing adopted in previous open ocean data assimilation, model accuracy scores, and the ensemble spread of water temperature and salinity revealed that perturbations of the lateral boundary conditions and river discharge forcing are important for regional data assimilation in coastal estuaries. To correct the entire Ise Bay,

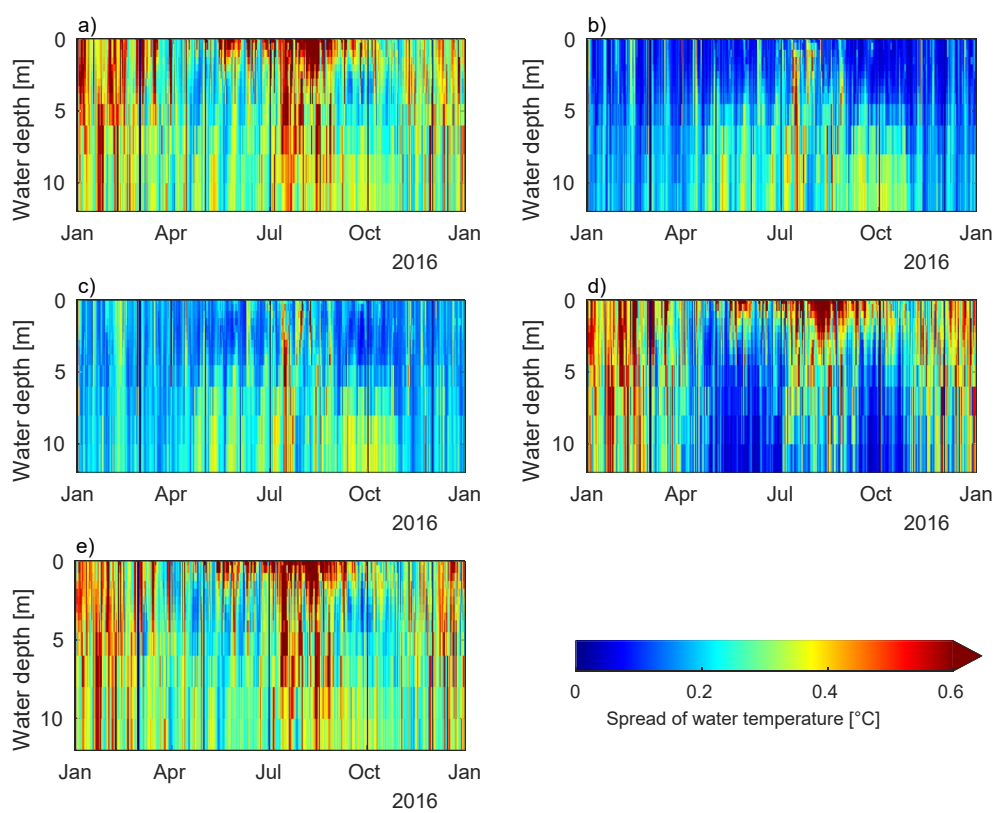

Figure 16. Temporal evolution of the ensemble spread of water temperature at C1 with water depth. (a) DAAll, (b) DAwoAtm, (c) DAwoWind, (d) DAwoLBC, and (e) DAwoRiv.

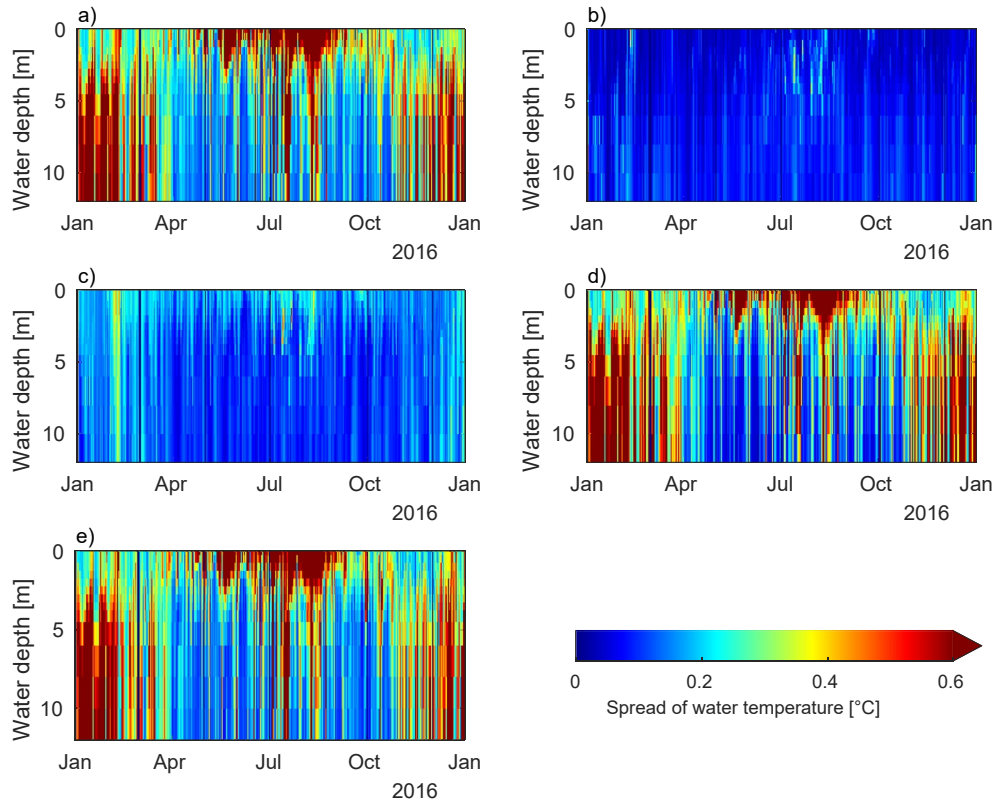

Figure 17. Temporal evolution of the ensemble spread of water temperature at C2 with water depth. (a) DAAll, (b) DAwoAtm, (c) DAwoWind, (d) DAwoLBC, and (e) DAwoRiv.

the influence of perturbation of atmospheric forcing should be greater than that of lateral boundary conditions and river discharge forcings within the possible magnitude of error of boundary conditions. It is important to add perturbations to both wind speed and air temperature to correct the water temperature.

Appendix A: Details of Boundary Condition Settings

A1. Atmospheric Forcing

Atmospheric forcing data were generated from observation data from 12 terrestrial observation stations of the Automated Meteorological Data Acquisition System (AMeDAS) near Ise Bay (Nagoya, Centrair, Gamagori, Minamichita, Toyohashi, Irago, Kuwana, Yokkaichi, Kameyama, Tsu, Omata, and Toba). All atmospheric forcing data at each calculation grid were interpolated using weighting interpolation with a normal distribution (the variance was 100 km^2) according to the distance from the observation stations. Shortwave radiation was calculated from daylight hours following the method of Nimiya et al. (1997). Longwave radiation was calculated according to the method of Nimiya et al. (1996). Wind velocity was set as follows. The observed wind speed was converted to wind speed at an altitude of 100 m using the logarithmic law in Equations A1 and A2:

$$W = \frac{U^*}{\kappa} \ln \frac{Z}{Z_0} \tag{A1}$$

$$U^* = \frac{W_0 \cdot \kappa}{\ln \frac{h_m}{Z_0}} \tag{A2}$$

where W is the converted wind speed, U^* is the friction speed, κ is the Kalman constant ($\kappa = 0.4$), Z is the height from the bottom, Z_0 is the roughness length, W_0 is the wind speed at the observation station, and h_m is the altitude of the wind anemometer. The roughness length at the sea surface was set to 0.001 m, and the roughness length

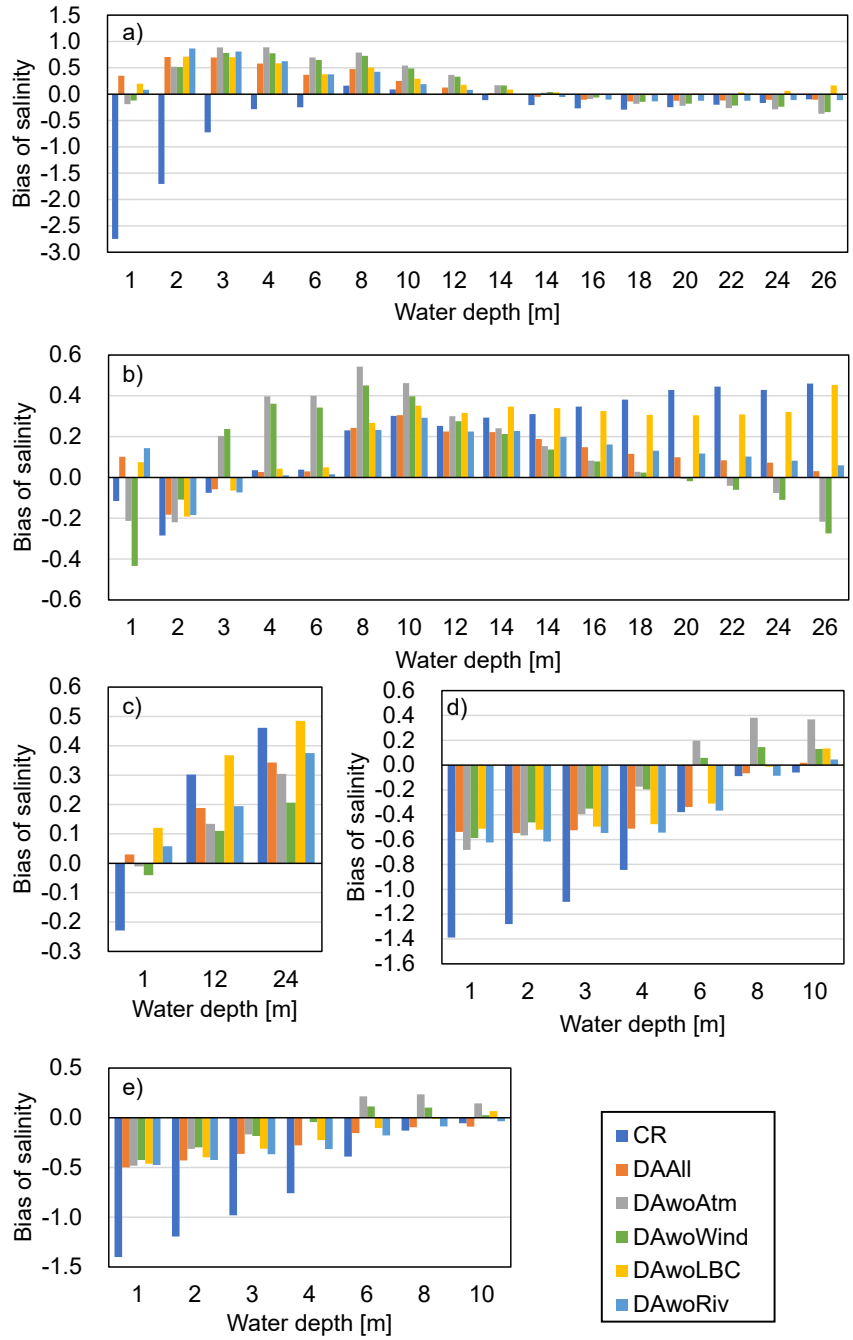

Figure 18. Bias of salinity between observations and model output using assimilated data from Table 2. (a) Back of bay, (b) center of bay, (c) mouth of bay, (d) No. 1 buoy, and (e) No. 2 buoy.

at each observation station was set according to the work of Kuwagata and Kondo (1990). Wind velocity at each calculation grid was interpolated using the same method as that for other weather data. Then, the wind speed at an altitude of 10 m was obtained by Equation A1. Vapor pressure e [hPa] was calculated using Equations A3 and A4:

$$e = es \times RH/100 \tag{A3}$$

$$es = 6.112 \times exp\left(\frac{17.62T_a}{243.12 + T_a}\right) \tag{A4}$$

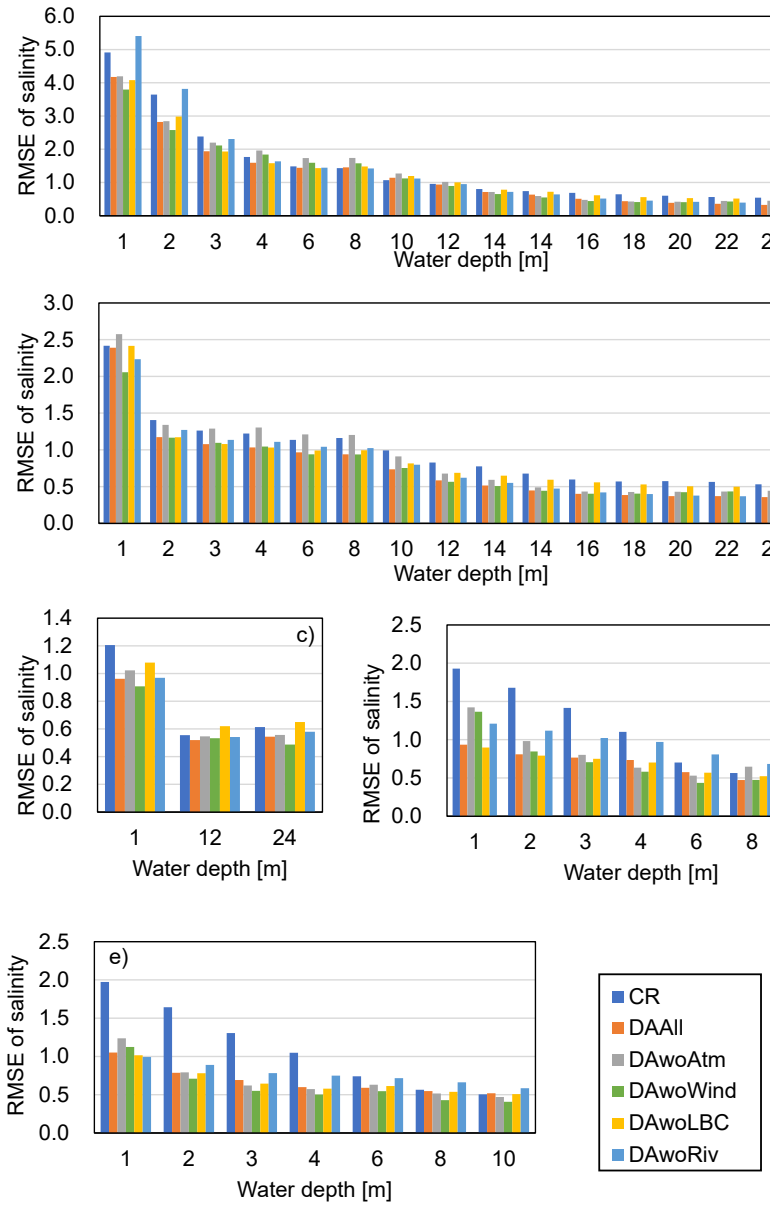

Figure 19. RMSE of salinity between observations and model output using assimilated data from Table 2. (a) Back of bay, (b) center of bay, (c) mouth of bay, (d) No. 1 buoy, and (e) No. 2 buoy.

where *es* is the saturation vapor pressure [hPa], *RH* is the relative humidity [%], and T_a [°C] is the air temperature. The parameter *es* was calculated using the method of the World Meteorological Organization (2008).

A2. Lateral Boundary Condition

The average water temperature and salinity each day of the year are calculated from monthly observation data (observation point number A10; latitude, 34.373; longitude, 137.216; measurement depth: 0, 10, 20, 30, 50, 75, 100, and 150 m below sea level) for 10 years (2004–2013) obtained by the Aichi Fisheries Research Institute. Their data were used to generate the lateral boundary conditions of water temperature and salinity. The observation data were uniformly interpolated in the horizontal direction since only one point was observed in the horizontal direction, linearly interpolated in the vertical direction, and linearly interpolated in the time direction. The tide level for the lateral boundary conditions was estimated using the amplitude and phase of 14 major tide components (Sa, Ssa, Mm, MSf, Mf, Q1, O1, P1, S1, K1, N2, M2, S2, K2) obtained from observation data of the

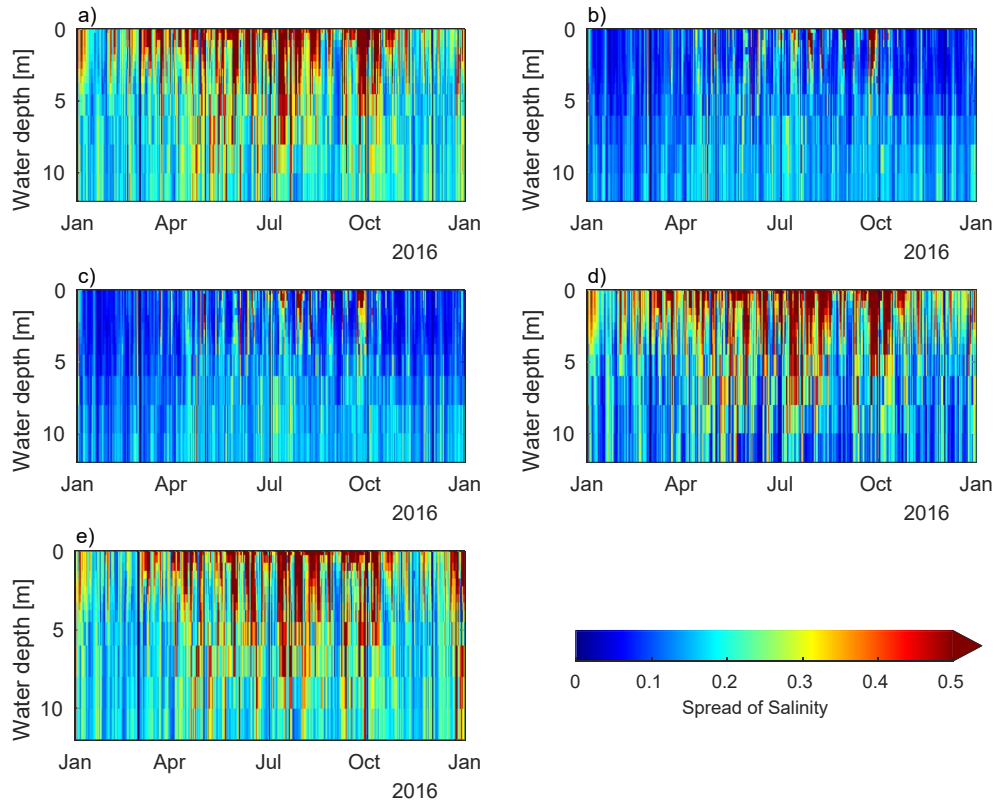

Figure 20. Temporal evolution of the ensemble spread of salinity at C1 with water depth. (a) DAAll, (b) DAwoAtm, (c) DAwoWind, (d) DAwoLBC, and (e) DAwoRiv.

Akabane tide station (latitude 34.6, longitude 137.18333) located near the lateral boundary. The estimated tide level was corrected using the atmospheric pressure.

A3. River Discharge Forcing

The river discharge was calculated by a storage function method, as follows.

$$\frac{ds}{dt} = q_{\mathrm{up}}(t) + r(t) - q(t) - q_{\mathrm{base}} \tag{A5}$$

$$s = k_1 q^p + k_2 \frac{dq}{dt} \tag{A6}$$

$$Q(t) = \frac{q(t)}{3.6} A \tag{A7}$$

where s is the apparent storage height of the basin [mm], t is time [h], r is the average precipitation in the basin [mm h^{-1}], q is the runoff over time t [mm h^{-1}], q_{up} is the runoff from the upper area [mm h^{-1}], q_{base} is the base runoff [mm h^{-1}], k_1, k_2, and p are constant values, Q is the river discharge [m^3 s^{-1}], and A is the basin area [km^2]. Equation A6 is based on Prasad (1967). For the class A river in the basin, k_1, k_2, and p were obtained to compare the observed river discharge values. For other smaller rivers, few river discharge observations are made during precipitation events; therefore, the parameters were estimated using the average precipitation value in the basin multiplied by the basin area to obtain the river discharge. The average precipitation (r) in each basin was calculated as follows. Each river basin was divided into a grid. The distance between each grid point and the AMeDAS observation point was calculated, and any AMeDAS data point less than 30 km from a grid point was extracted. Here, the maximum number of AMeDAS observation points used at each grid point was 10. Precipitation at each

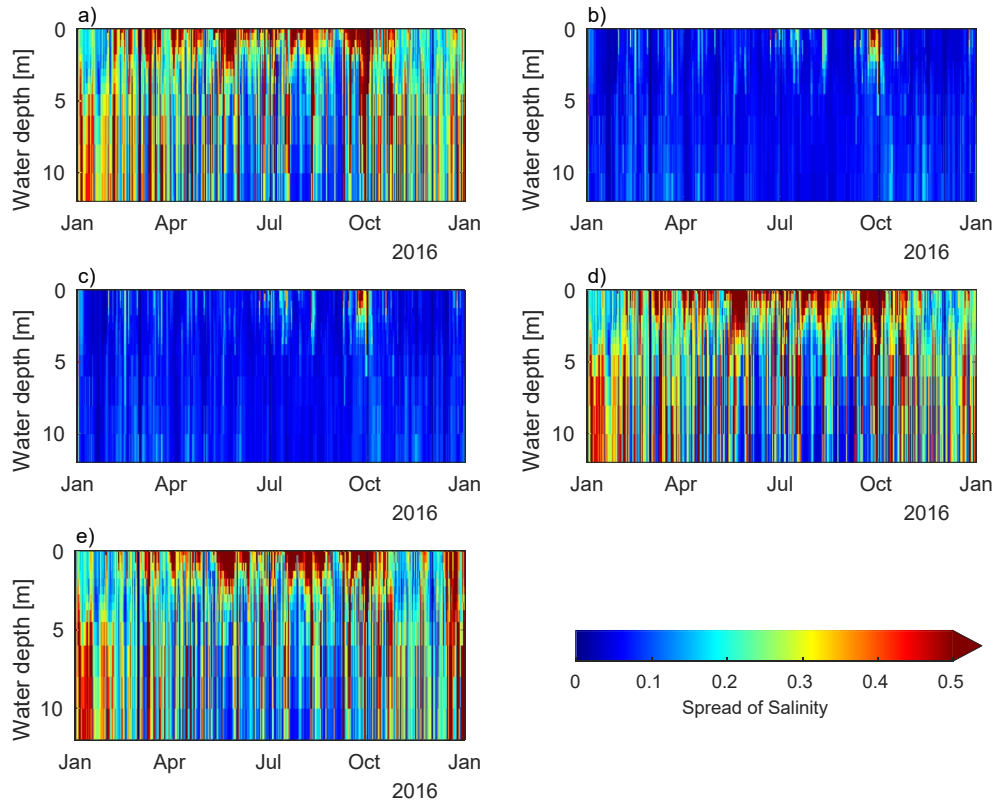

Figure 21. Temporal evolution of the ensemble spread of salinity at C2 with water depth. (a) DAAll, (b) DAwoAtm, (c) DAwoWind, (d) DAwoLBC, and (e) DAwoRiv.

grid was calculated by weighting according to the same method used for other weather data. The sum of precipitation for each grid was taken as the average precipitation of the basin.

River water temperature was calculated from the air temperature near the mouth of the river using Equation A8:

$$T_w = aT_a + b \tag{A8}$$

where T_w [°C] is the river water temperature, and a and b are parameters calculated from the relationship between the observed air temperature near the river mouth and the observed river water temperature. The parameters a and b are calculated in each river.

Appendix B: Estimation of the Magnitude of Perturbation to Boundary Conditions

B1. Air Temperature

The dominant error factors of the atmospheric forcing condition of air temperature were the differences between observation points (sea and ground) and the influence of spatial interpolation. Therefore, it is assumed that the air temperatures are accurate at five locations in Ise Bay (center of the bay, mouth of the bay, and buoys 1–3), where the observed air temperature is shown in Tables 2 and 4, and from April 2015 to December 2019. The boundary condition between the air temperature observed at the monitoring locations in Ise Bay and the air temperature calculated at the same position was extracted every hour. The cumulative frequency distribution of the absolute difference between the observed value and the calculated value was obtained after subtracting the average error, and the temperature at which the cumulative frequency was 68.2% was calculated as 3.05°C. Therefore, we added system noise with a normal distribution and a standard deviation of the ξ value of 3.05°C to the boundary conditions of air temperature for each ensemble member.

B2. Wind Speed

The error factor and ξ of the atmospheric forcing condition of wind speed was estimated using the same method as that for air temperature. The cumulative frequency distribution of the absolute difference between the observed value and the boundary condition was obtained, and the value at which the cumulative frequency was 68.2% was calculated as 3.45 m s^{-1}. Therefore, we added system noise with a normal distribution and a standard deviation of the ξ value of 3.45 m s^{-1} to the boundary conditions of wind speed for each ensemble member.

B3. Water Temperature of the Lateral Boundary

The error factor of the lateral boundary condition of water temperature was mainly caused because the original data used to create the boundary conditions was not observed during the simulation period, but was the average value over 10 years, as explained in Section 2.2. Then, ξ was estimated as follows. First, it was assumed that the observed water temperature is accurate. Second, the error was estimated by comparing the observed values with the open boundary conditions. The comparison period was for 1 year (2015). The cumulative frequency distribution of the absolute difference between the observed value and the boundary condition was calculated after subtracting the average error, and the value at which the cumulative frequency was 68.2% was calculated as 0.73°C. Therefore, we added system noise with a normal distribution and a standard deviation of the ξ value of 0.73°C to the open boundary condition of water temperature for each ensemble member.

B4. Salinity of the Lateral Boundary

The error factor and ξ of the lateral boundary condition of salinity was estimated using the same method as that for water temperature. The cumulative frequency distribution of the absolute difference between the observed value and the boundary condition was obtained, and the value at which the cumulative frequency was 68.2% was calculated as 0.20. Therefore, we added system noise with a normal distribution and a standard deviation of the ξ value of 0.20 to the boundary conditions of salinity for each ensemble member.

B5. River Discharge

The error factors of river discharge were predominantly the estimation error of the storage function method and the spatiotemporal error of input precipitation. Thus, the ξ value of river discharge was estimated as follows. It was assumed that the rate of fluctuation inherent in river discharge is the same for each river simultaneously. When the rate of discharge fluctuation varies for each river, the variation is regarded as the error of the river discharge. The analysis period was set from April 2015 to December 2019, and the average discharge was calculated for the 10 major rivers flowing into Ise Bay. The river discharge change rate was calculated by dividing the discharge of each river at each time by the average discharge for each river, and the standard deviation for each time was obtained. When the cumulative frequency of the standard deviation was 68.2%, the value was calculated as 0.35. Therefore, the boundary condition was multiplied by the system noise with a normal distribution and a standard deviation of 0.35.

B6. River Water Temperature

The spatial correlation error and estimation error were considered the dominant error factors of river water temperature. Therefore, when there was a difference in water temperature between rivers, system noise was added by assuming that it was an error. The standard deviation regarding the variation in water temperature at each time for each river was calculated for the 10 major rivers that flow into Ise Bay. The analysis period was from April 2015 to December 2019. Then, if the distribution of the magnitude of the error for the entire period follows a normal distribution, the cumulative frequency distribution was created, and the value at which the cumulative frequency was 68.2% was calculated. Therefore, we added system noise with a normal distribution and a standard deviation of the ξ value of 1.21°C to the boundary condition of temperature for each ensemble member.

Data Availability Statement

AMeDAS is operated by the Japan Meteorological Agency. Data are available at https://www.data.jma.go.jp/obd/stats/etrn. The observations of water temperature, salinity, wind speed, and air temperature are operated by Chubu Regional Bureau, Ministry of Land, Infrastructure, Transport, and Tourism of Japan, and Aichi Fisheries Research Institute. Data are available at http://www.isewan-db.go.jp. SST observations by MODIS is operated by NASA. Data are available at https://oceancolor.gsfc.nasa.gov.

Acknowledgments
This study results from commissioned research from Chubu Regional Bureau, Ministry of Land, Infrastructure, Transport, and Tourism, Japan. This work was supported by JSPS KAKENHI Grant No. 21K14255. Observed data were obtained from Chubu Regional Bureau, Ministry of Land, Infrastructure, Transport, and Tourism of Japan, the Japan Meteorological Agency, Aichi Fisheries Research Institute, and the National Aeronautics and Space Administration (NASA). We extend thanks to Shun Ohishi and Teruhisa Okada for their advice on our study. We would like to thank Editage (www.editage.com) for the English language editing. No potential conflict of interest was reported by the authors.

References

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, *127*(12), 2741–2758. https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2

Baduru, B., Paul, B., Banerjee, D. S., Sanikommu, S., & Paul, A. (2019). Ensemble based regional ocean data assimilation system for the Indian Ocean: Implementation and evaluation. *Ocean Modelling*, *143*, 101470. https://doi.org/10.1016/j.ocemod.2019.101470

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., et al. (2010). The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, *91*(8), 1059–1072. https://doi.org/10.1175/2010BAMS2853.1

Brankart, J.-M., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P.-A., et al. (2015). A generic approach to explicit simulation of uncertainty in the NEMO ocean model. *Geoscientific Model Development*, *8*(5), 1285–1297. https://doi.org/10.5194/gmd-8-1285-2015

Edwards, C. A., Moore, A. M., Hoteit, I., & Cornuelle, B. D. (2015). Regional ocean data assimilation. *Annual Review of Marine Science*, *7*(1), 21–42. https://doi.org/10.1146/annurev-marine-010814-015821

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. https://doi.org/10.1029/94JC00572

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2009). *Data assimilation: The ensemble Kalman filter* (2). Springer.

Fisher, M., & Courtier, P. (1995). *Estimating the covariance matrices of analysis and forecast error in variational data assimilation* (p. 220). ECMWF Technical Memoranda. https://doi.org/10.21957/1DXRASJIT

Fisher, R. A., & Yates, F. (1948). *Statistical tables for biological, agricultural and medical research*. Oliver and Boyd.

Fu, L.-L., Fukumori, I., & Miller, R. N. (1993). Fitting dynamic models to the geosat sea level observations in the Tropical Pacific Ocean. Part II: A linear, wind-driven model. *Journal of Physical Oceanography*, *23*(10), 2162–2181. https://doi.org/10.1175/1520-0485(1993)023<2162:FDMTTG>2.0.CO;2

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Hafeez, M. A., Nakamura, Y., Suzuki, T., Inoue, T., Matsuzaki, Y., Wang, K., & Moiz, A. (2021). Integration of Weather Research and Forecasting (WRF) model with regional coastal ecosystem model to simulate the hypoxic conditions. *The Science of the Total Environment*, *771*, 145290. https://doi.org/10.1016/j.scitotenv.2021.145290

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Henderson-Sellers, B. (1985). New formulation of eddy diffusion thermocline models. *Applied Mathematical Modelling*, *9*(6), 441–446. https://doi.org/10.1016/0307-904X(85)90110-6

Hoffman, M. J., Miyoshi, T., Haine, T. W. N., Ide, K., Brown, C. W., & Murtugudde, R. (2012). An advanced data assimilation system for the Chesapeake Bay: Performance evaluation. *Journal of Atmospheric and Oceanic Technology*, *29*(10), 1542–1557. https://doi.org/10.1175/JTECH-D-11-00126.1

Hoteit, I., Luo, X., Bocquet, M., Köhl, A., & Ait-El-Fquih, B. (2018). Data assimilation in oceanography: Current status and new directions. In E. P. Chassignet, A. Pascual, J. Tintoré, & J. Verron (Eds.), *New frontiers in operational oceanography*. GODAE OceanView. https://doi.org/10.17125/gov2018.ch17

Khanarmuei, M., Mardani, N., Suara, K., Sumihar, J., Sidle, R. C., McCallum, A., & Brown, R. J. (2021). Assessment of an ensemble-based data assimilation system for a shallow estuary. *Estuarine, Coastal and Shelf Science*, *257*, 107389. https://doi.org/10.1016/j.ecss.2021.107389

Kunii, M., & Miyoshi, T. (2012). Including uncertainties of sea surface temperature in an ensemble Kalman filter: A case study of Typhoon Sinlaku (2008). *Weather and Forecasting*, *27*(6), 1586–1597. https://doi.org/10.1175/WAF-D-11-00136.1

Kuwagata, T., & Kondo, J. (1990). Estimation of aerodynamic roughness at the regional meteorological stations (AMeDAS) in the central part of Japan. *TENKI*, *37*(3), 55–59.

Kwon, K. M., Choi, B.-J., Lee, S.-H., Kim, Y. H., Seo, G.-H., & Cho, Y.-K. (2016). Effect of model error representation in the Yellow and East China Sea modeling system based on the ensemble Kalman filter. *Ocean Dynamics*, *66*(2), 263–283. https://doi.org/10.1007/s10236-015-0909-8

Lima, L. N., Pezzi, L. P., Penny, S. G., & Tanajura, C. A. S. (2019). An investigation of ocean model uncertainties through ensemble forecast experiments in the Southwest Atlantic Ocean. *Journal of Geophysical Research: Oceans*, *124*(1), 432–452. https://doi.org/10.1029/2018JC013919

Matsuzaki, Y., Fujiki, T., Kawaguchi, K., Inoue, T., & Iwamoto, T. (2021). Application of the WRF model to the coastal area at Ise Bay, Japan: Evaluation of model output sensitivity to input data. *Coastal Engineering Journal*, *63*(1), 17–31. https://doi.org/10.1080/21664250.2020.1830485

Matsuzaki, Y., & Inoue, T. (2020). *A study of calculation conditions of data assimilation for flow analysis of closed waters by ensemble Kalman filter* (Vol. 1367, p. 35). Technical Note of the Port and Airport Research Institute.

Mirouze, I., & Storto, A. (2019). Generating atmospheric forcing perturbations for an ocean data assimilation ensemble. *Tellus A: Dynamic Meteorology and Oceanography*, *71*(1), 1624459. https://doi.org/10.1080/16000870.2019.1624459

Moore, A. M., Arango, H. G., Broquet, G., Powell, B. S., Weaver, A. T., & Zavala-Garay, J. (2011). The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I – System overview and formulation. *Progress in Oceanography*, *91*(1), 34–49. https://doi.org/10.1016/j.pocean.2011.05.004

Nakada, S., Kobayashi, S., Hayashi, M., Ishizaka, J., Akiyama, S., Fuchi, M., & Nakajima, M. (2018). High-resolution surface salinity maps in coastal oceans based on geostationary ocean color images: Quantitative analysis of river plume dynamics. *Journal of Oceanography*, *74*(3), 287–304. https://doi.org/10.1007/s10872-017-0459-4

Nakamura, Y., & Hayakawa, N. (1991). Modelling of thermal stratification in lakes and coastal seas. In *Presented at the IAHS publication (international association of hydrological sciences)* (pp. 227–236). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0025843997&partnerID=40&md5=7952862f01733aee89cce203e30718ca

Nimiya, H., Akasaka, H., & Matsuo, Y. (1996). A method to estimate the hourly downward atmospheric radiation using AMeDAS data. *The Society of Heating, Air-Conditioning Sanitary Engineers of Japan*, (60), 133–144.

Nimiya, H., Akasaka, H., Matsuo, Y., & Soga, K. (1997). A method to estimate the hourly solar radiation using AMeDAS data Part 2-Application to the improved sunshine recorder of AMeDAS. *The Society of Heating, Air-Conditioning Sanitary Engineers of Japan*, (65), 53–65.

Orlanski, I. (1976). A simple boundary condition for unbounded hyperbolic flows. *Journal of Computational Physics*, *21*(3), 251–269. https://doi.org/10.1016/0021-9991(76)90023-1

Penny, S. G., Behringer, D. W., Carton, J. A., & Kalnay, E. (2015). A hybrid global ocean data assimilation system at NCEP. *Monthly Weather Review*, *143*(11), 4660–4677. https://doi.org/10.1175/MWR-D-14-00376.1

Prasad, R. (1967). A nonlinear hydrologic system response model. *Journal of the Hydraulics Division*, *93*(4), 201–222. https://doi.org/10.1061/jyceaj.0001646

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., & Korablev, A. (2012). TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, *8*(4), 633–656. https://doi.org/10.5194/os-8-633-2012

Sanikommu, S., Toye, H., Zhan, P., Langodan, S., Krokos, G., Knio, O., & Hoteit, I. (2020). Impact of atmospheric and model physics perturbations on a high-resolution ensemble data assimilation system of the Red Sea. *Journal of Geophysical Research: Oceans*, *125*(8), e2019JC015611. https://doi.org/10.1029/2019JC015611

Sekine, Y., Nakamura, S., & Wang, Y. W. (2002). Variation in wind in the region around Ise Bay. Mie University. *The Bulletin of the Faculty of Bioresources*, *28*, 1–9.

Stanev, E. V., Schulz-Stellenfleth, J., Staneva, J., Grayek, S., Grashorn, S., Behrens, A., et al. (2016). Ocean forecasting for the German Bight: From regional to coastal scales. *Ocean Science*, *12*(5), 1105–1136. https://doi.org/10.5194/os-12-1105-2016

Tanaka, Y., & Suzuki, K. (2010). Development of non-hydrostatic numerical model for stratified flow and upwelling in estuary and coastal areas. *Report of the Port and Airport Research Institute*, *49*(1), 3–26.

Turner, M., Walker, J., & Oke, P. (2008). Ensemble member generation for sequential data assimilation. *Remote Sensing of Environment*, *112*(4), 1421–1433. https://doi.org/10.1016/j.rse.2007.02.042

Vandenbulcke, L., & Barth, A. (2015). A stochastic operational forecasting system of the Black Sea: Technique and validation. *Ocean Modelling*, *93*, 7–21. https://doi.org/10.1016/j.ocemod.2015.07.010

Vervatis, V. D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ghantous, M., Kailas, M., et al. (2021). Assessment of a regional physical–biogeochemical stochastic ocean model. Part 1: Ensemble generation. *Ocean Modelling*, *160*, 101781. https://doi.org/10.1016/j.ocemod.2021.101781

Weaver, A., & Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, *127*(575), 1815–1846. https://doi.org/10.1002/qj.49712757518

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089. https://doi.org/10.1175/MWR-D-11-00276.1

World Meteorological Organization. (2018). *Guide to instruments and methods of observation—Volume 1 measurement of meteorological variables*. World Meteorological Organization Report WMO-No. 8. Retrieved from https://library.wmo.int/index.php?id=12407&lvl=notice_display