

## Assimilating In Situ and Remote Sensing Observations in a Highly Variable Estuary–Shelf Model

WENFENG LAI,<sup>a,b,c</sup> JIANPING GAN,<sup>a,b,c</sup> YE LIU,<sup>d</sup> ZHIQIANG LIU,<sup>e</sup> JIPING XIE,<sup>f</sup> AND JIANG ZHU<sup>g</sup>

<sup>a</sup> Center for Ocean Research in Hong Kong and Macau, Hong Kong University of Science and Technology, Hong Kong

<sup>b</sup> Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

<sup>c</sup> Department of Ocean Science, Hong Kong University of Science and Technology, Hong Kong

<sup>d</sup> Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

<sup>e</sup> Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China

<sup>f</sup> Nansen Environmental and Remote Sensing Center, Bergen, Norway

<sup>g</sup> Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

(Manuscript received 11 June 2020, in final form 14 November 2020)

**ABSTRACT:** To improve the forecasting performance in dynamically active coastal waters forced by winds, tides, and river discharges in a coupled estuary–shelf model off Hong Kong, a multivariable data assimilation (DA) system using the ensemble optimal interpolation method has been developed and implemented. The system assimilates the conductivity–temperature–depth (CTD) profilers, time series buoy measurement, and remote sensing sea surface temperature (SST) data into a high-resolution estuary–shelf ocean model around Hong Kong. We found that the time window selection associated with the local dynamics and the number of observation samples are two key factors in improving assimilation in the unique estuary–shelf system. DA with a varied assimilation time window that is based on the intratidal variation in the local dynamics can reduce the errors in the estimation of the innovation vector caused by the model–observation mismatch at the analysis time and improve simulation greatly in both the estuary and coastal regions. Statistically, the overall root-mean-square error (RMSE) between the DA forecasts and not-yet-assimilated observations for temperature and salinity has been reduced by 33.0% and 31.9% in the experiment period, respectively. By assimilating higher-resolution remote sensing SST data instead of lower-resolution satellite SST, the RMSE of SST is improved by ~18%. Besides, by assimilating real-time buoy mooring data, the model bias can be continuously corrected both around the buoy location and beyond. The assimilation of the combined buoy, CTD, and SST data can provide an overall improvement of the simulated three-dimensional solution. A dynamics-oriented assimilation scheme is essential for the improvement of model forecasting in the estuary–shelf system under multiple forcings.

**KEYWORDS:** Remote sensing; Forecasting techniques; Data assimilation; Interpolation schemes; Model evaluation/performance; Ocean models

### 1. Introduction

In ocean model simulation, data assimilation (DA) provides an efficient way to address the uncertainties and improve the forecasts of the ocean model by using observations. Due to the uncertainties over model parameterization, driving forces, and initial and boundary conditions, ocean models are imperfect representations of the actual environments. Assimilation techniques for ocean forecasts are commonly classified into sequential methods (e.g., Kalman filter) and variational methods (e.g., 3D/4D-Var). The variational DA algorithms are based on optimal control theory and consider the initial state as the control vector. They need to compute both the linearized state and the adjoint state and use a minimization algorithm to minimize the cost-function. However, complex algorithms make their implementations difficult. The sequential methods are generally adopted, which, based on filter theory, estimate the system state sequentially forward in time by adding a statistically based correction term to the solution of the balance equations. This algorithm is relatively easy to implement. One popular

sequential approach is the ensemble Kalman filter (EnKF) that was introduced by Evensen (1994, 2003) and Burgers et al. (1998). The EnKF is a Monte Carlo approximation to the Kalman filter. However, a practical limitation of the EnKF is the ensemble size (Oke et al. 2007). Thus, the ensemble optimal interpolation (EnOI), an approximation of the EnKF, was born and provides a cost-effective solution (Oke et al. 2002; Evensen 2003). In theory, the EnOI performs the analysis by a time-invariant ensemble of model states sampled from a long-term model integration. The EnOI requires only one deterministic model run, and only one background state needs to be updated. Thus, the computational cost is inexpensive compared to the EnKF (Oke et al. 2007).

Owing to its attractive characteristics like quasi-dynamically consistent, multivariate, inhomogeneous, anisotropic covariances, and high efficiency, the EnOI methods have been applied by many operational ocean forecasting centers, for example, in China, Brazil, South Africa, and Australia (Lyu et al. 2014; Tanajura et al. 2014; Backeberg et al. 2014; Oke et al. 2008). Oke et al. (2007) have demonstrated the power of the multivariate EnOI in solving the eddy-resolving problem in a region of energetic mesoscale variability where the

Corresponding author: Jianping Gan, [magan@ust.hk](mailto:magan@ust.hk)

DOI: 10.1175/JTECH-D-20-0084.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

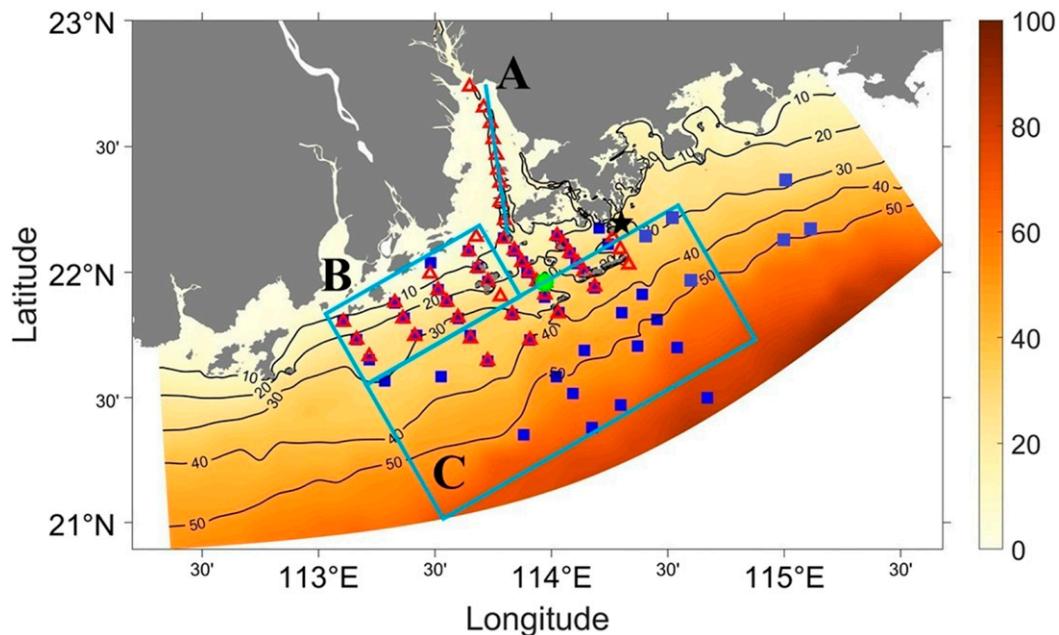


FIG. 1. CTD cast stations during the field survey period, and the PRE model domain and bathymetry. The red triangles indicate the stations in leg 1, and the blue squares indicate the stations in leg 2. The green dot indicates the buoy location. The black star indicates the location of Waglan Island.

Leeuwin Current frequently sheds eddies to the west of Australia. [Counillon and Bertino \(2009\)](#) also suggested that EnOI can effectively improve the forecasts of the mesoscale variability in the Gulf of Mexico. To date, real-time buoy-mounted, shipboard in situ, and remote sensing SST data are the three main types of data applied in the DA scheme. For instance, [Liu et al. \(2013\)](#) assimilated temperature and salinity profiles into an ocean model of the Baltic Sea by using the EnOI approach and proposed that the static ensemble can well resolve the high singularity in the coastal and offshore waters. The remote sensing SST data have been widely applied to the operational forecast, reanalysis, or validation of the model due to their excellent spatiotemporal coverage. [Losa et al. \(2012\)](#) reduced the model uncertainties by assimilating the Advanced Very High Resolution Radiometer (AVHRR) SST data into the North Sea and Baltic Sea forecasting model. [O'Dea et al. \(2012\)](#) improved the SST prediction by assimilating both in situ and remote sensing SST data into the European northwest shelf operational model. Based on EnOI, [D. Liu et al. \(2018\)](#) carried on observing system simulation experiments (OSSE) to study the impact of assimilating moored velocity on the improvement of the simulation in the intraseasonal variability. [Crosby et al. \(2017\)](#) reported that assimilating buoy observations can improve model predictions and wave hindcasts and suggested that dense observational networks lead to a significant improvement in model performance. Despite successful applications of DA in the mesoscale ocean simulation in many previous studies, its application in estuary–shelf waters is still a challenge ([De Mey et al. 2017](#)). The water properties (e.g., temperature, salinity, and velocity) in coastal waters are highly variable due to the amplified forcing effects of winds, tides, and

freshwater flux, and topographic control. Indeed, a well-performed DA scheme in the open ocean may not be reliable in the dynamic coastal waters ([Barth et al. 2011](#); [Sperrevik et al. 2015](#); [Edwards et al. 2015](#); [Botto et al. 2018](#)).

The Pearl River Estuary (PRE) is a semienclosed coastal embayment in the northern South China Sea (NSCS; [Fig. 1](#)). The upper estuary is linked to the Pearl River and the lower estuary to the adjacent continental shelf. Together, they form an estuary–shelf system off Hong Kong. The Pearl River discharges on average  $10\,000\text{ m}^3\text{ s}^{-1}$  of freshwater annually into the NSCS. The PRE is shaped like a trumpet with a width of 5 km at the northern end and 35 km at the southern end. Over the continental shelf outside the estuary, the water depth increases to more than 20 m and the isobaths run approximately parallel to the coastline with a strong cross-shelf pressure gradient. The coastal waters of the estuary–shelf system have mixed tides predominated by the diurnal frequency with an amplitude of  $\sim 1$  m and a strong spring–neap tidal cycle ([Mao et al. 2004](#)). The tidal currents change stratification structures in the PRE and play an important role in vertical mixing, especially in the shallow nearshore regions ([Lai et al. 2018](#); [Z. Liu et al. 2018](#)). In summer/winter, the monsoon-driven currents form a freshwater plume that extends eastward/westward over the shelf ([Pan et al. 2020](#)). The circulation and associated river plume exhibit a distinct seasonal variation, as a result of the interaction between the wind-driven current and estuarine circulation ([Gan et al. 2009](#); [Zu et al. 2014](#)). The overall circulation in the estuary–shelf system off Hong Kong is highly variable in space and time subject to the controls of an irregular coastline, variable topography, and the multiple forcings of winds, river discharges, and tides. Inevitably, the

model forecast contains uncertainties. Shu et al. (2011a) analyzed the different sources of error in the circulation simulation in NSCS and found that the major forecast errors come from river discharges, atmospheric forcing, and initial and open boundary conditions. Further, Shu et al. (2011b) reproduced the structure of Pearl River Plume and coastal upwelling by assimilating Group for High Resolution (GHR) SST and in situ data using an ensemble data assimilation method.

To improve the accuracy of initial conditions and reduce forecast uncertainties in the estuary–shelf model off Hong Kong, a DA system can be applied to constrain the model with coastal observations. To date, in situ and remote sensing data have been collected in this region. Since the in situ measurements are sparsely and asynchronously observed in a highly variable coastal environment, that can potentially introduce error in the estimation of the innovation vectors (observation minus background) by comparing observations with the background at the analysis time. In this study, we propose a novel DA scheme with a variable assimilation time window to correct the mismatch between the observations and the model forecast in the analysis process due to asynchronous observation in a highly variable estuary–shelf system off Hong Kong. We demonstrate the effectiveness of the scheme, which can minimize model–observation mismatch and provides sufficient observations for reconstructing the model’s initial conditions. We investigate the model forecast performance when different types of observational data are assimilated and assess the impacts of assimilating the time series buoy (temporally continuous) data and remote sensing SST (spatially continuous) data. Furthermore, the impacts of the DA system on the hydrodynamics in the PRE are discussed.

## 2. Assimilation algorithms and ocean model

### a. EnOI algorithms

In EnOI, the assumption is the anomaly based on the ensemble can be used to mimic the model error variance, as the previous implementations in the SCS (Xie et al. 2011; Lyu et al. 2014). The EnOI estimates an “optimal” oceanic state at a given time using a numerical model, observations, and assumptions on their respective distribution and uncertainty. The distribution of stochastic errors is assumed to be Gaussian and nonbiased. This relationship is summarized in the following:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \quad (1)$$

where  $\mathbf{y}$  is the  $m \times 1$  vector of observations ( $m$  is the number of available observations),  $\mathbf{x}$  is the  $n \times 1$  model state estimate vector ( $n$  is the number of model state variables ordered by grid points), and  $\mathbf{H}$  is a measurement operator that transforms the model state to the observation space. The superscripts “ $a$ ” and “ $f$ ” refer to the analysis and the forecast model states, respectively. The vector  $\mathbf{y} - \mathbf{H}\mathbf{x}^f$  describes the innovations, and  $\mathbf{K}$  is the gain matrix that weights the observational information based on the model and observation error covariance. Each row of  $\mathbf{K}$  contains the weights used to update one state variable depending on the innovation vector. Assuming that the model

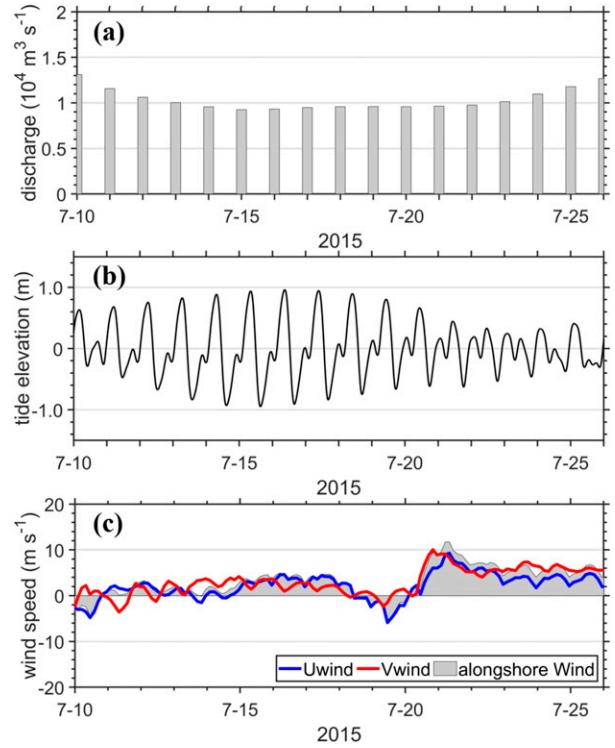


FIG. 2. Time series of (a) river discharge observed at Gaoyao Station, (b) tide elevation at Waglan Island (shown in Fig. 1), and (c) wind stress averaged in the NSCS during the cruise period. The positive values indicated by the red and blue lines in (c) represent westerly and southerly winds, respectively, and the positive value of alongshore wind indicates upwelling-favorable wind (gray area).

and observation errors are nonbiased, independent of each other and that all variables are Gaussian, the gain that leads to a minimum distance to an unknown truth is

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (2)$$

where  $\mathbf{P}^f$  represents the  $n \times n$  prior sample error covariance matrix (approximated by the forecast ensemble),  $\mathbf{R}$  is the  $m \times m$  observation error covariance, and superscript T denotes matrix transpose. As can be seen from the formulas, the quality of the analysis depends on the error statistics  $\mathbf{P}^f$  and  $\mathbf{R}$ . The EnOI is a cost-effective ensemble DA method that assumes a static ensemble is representative of the instantaneous forecast error. The forecast covariance matrix is computed from a collection of model states, as follows:

$$\mathbf{P}^f = \frac{\alpha}{N-1} \mathbf{A}' \mathbf{A}'^T, \quad (3)$$

where the ensemble anomalies are given as

$$\mathbf{A}' = \mathbf{A}^f - \overline{\mathbf{A}^f}, \quad \text{and} \quad (4)$$

$$\mathbf{A}^f = [x_1^f, x_2^f, \dots, x_N^f], \quad (5)$$

TABLE 1. Summary of assimilation experiments.

Expt	DA data	Obs time window for assimilation	Reinitialization/analysis time $T$	Obs time window for validation
NoDA	Control run without DA	No	No	No
EXP1	Only CTD data	$(T - 24\text{ h}, T)$	At 0000 local time	$(T, T + 24\text{ h})$
EXP2	Only CTD data	$(T - 72\text{ h}, T)$	Max flood/ebb	$(T, T + 72\text{ h})$
EXP3	Only OSTIA SST data	$T$	Satellite obs time	$(T, T + 24\text{ h})$
EXP4	Only MUR SST data	$T$	Satellite obs time	$(T, T + 24\text{ h})$
EXP5	Only buoy data	$(T - 0.5\text{ h}, T + 0.5\text{ h})$	At 0000 local time	$(T, T + 24\text{ h})$
EXP6	CTD+buoy+MUR SST	$(T - 72\text{ h}, T)$	Max flood/ebb	$(T, T + 72\text{ h})$

where  $\mathbf{A}^f$  is the ensemble of  $N$  model forecasts and  $\overline{\mathbf{A}^f}$  is the ensemble mean;  $\alpha \in (0, 1]$  is a parameter introduced to adjust the covariance of the ensemble perturbations to capture the model uncertainties. In this study, the  $\alpha$  is set to 0.3.

### b. The coastal ocean model

In this numerical study, the Regional Ocean Modeling System (ROMS) is used to simulate the hydrodynamic environment in the coupled estuary–shelf system off the PRE (Liu and Gan 2020). ROMS is a free-surface, hydrostatic, primitive equation model discretized with a terrain-following vertical coordinate system (Shchepetkin and McWilliams 2005). The model domain has been horizontally discretized into a matrix with  $400 \times 441$  points covering the PRE and the shelves off Guangdong in the NSCS (Fig. 1). An orthogonal curvilinear coordinate system is designed to follow the coastline. The ultrahigh resolution ( $\sim 0.1$  km) resolves the estuary and the inner shelf that neighbors Hong Kong. The grid size gradually increases to  $\sim 1$  km over the shelf at its southern boundary. The model has 30 vertical levels with terrain-following coordinates (Song and Haidvogel 1994) and adopts higher resolutions ( $< 0.2$  m) in both the surface and bottom boundary layers to better resolve the dynamics there. The model is nested within an NSCS model with a coarser resolution ( $\sim 3$  km; Gan et al. 2015). This NSCS model is further downscaled from a hindcast simulation in the China Seas Multiscale Ocean Modeling System (CMOMS) (Gan et al. 2016).

The model is initialized and spun-up with temperature and salinity extracted from the NSCS model on 1 May 2015. A new tidal and subtidal open boundary condition (TST-OBC) scheme developed by Liu and Gan (2016, 2020) is applied to

this limited-area ocean model. Tidal forcing is applied to the open boundary by eight major tidal constituents of  $M_2$ ,  $K_1$ ,  $S_2$ ,  $O_1$ ,  $N_2$ ,  $P_1$ ,  $K_2$ , and  $Q_1$ , which are extracted from the Oregon State University Tidal Inversion Software (Egbert and Erofeeva 2002). Atmospheric fluxes are estimated by the bulk formula using the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis dataset (ERA-Interim). A time-dependent river discharge obtained from the upstream hydrographic monitoring station at Gaoyao is applied as the runoff from the Pearl River. Details about the model implementation and the well-validated model results can be seen in Liu and Gan (2020).

## 3. Observations and experiments

### a. Observations

A hydrographic cruise to collect in situ data on marine environmental parameters in the PRE was implemented from 10 to 25 July 2015. The cruise was divided into two legs: 1) leg 1 from 13 to 18 July and 2) leg 2 from 19 to 25 July. Figure 2 shows the river discharge, tide, and wind conditions during the cruise period. The river discharge was stable during this period, with an average discharge of  $10\,000\text{ m}^3\text{ s}^{-1}$ . The field survey experienced a prominent spring–neap tidal cycle, where a spring tide was seen during leg 1 and a neap tide during leg 2. The winds were weaker during leg 1, while a typical summer southwesterly monsoon prevailed during leg 2. In situ salinity and temperature were measured with a well-calibrated Sea-Bird SBE25 CTD profiling system. The measurement stations are located along the transects in the PRE and over the

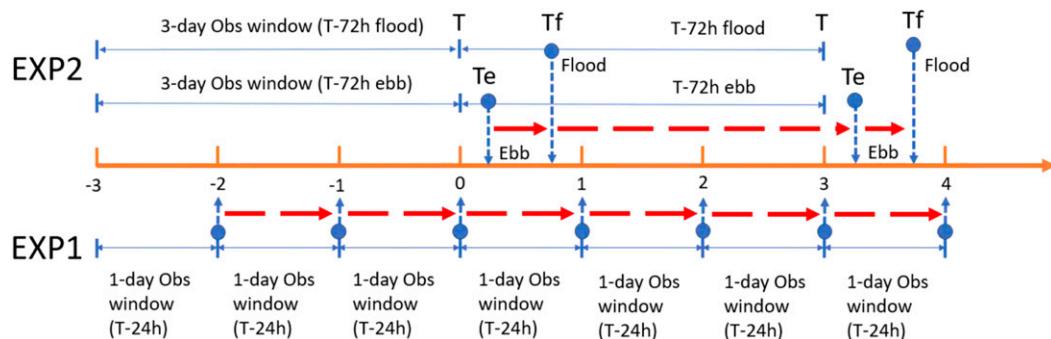


FIG. 3. Analysis–forecast schemes for EXP1 and EXP2 that were used in the model. The blue dot indicates the reinitialization/analysis time of DA.

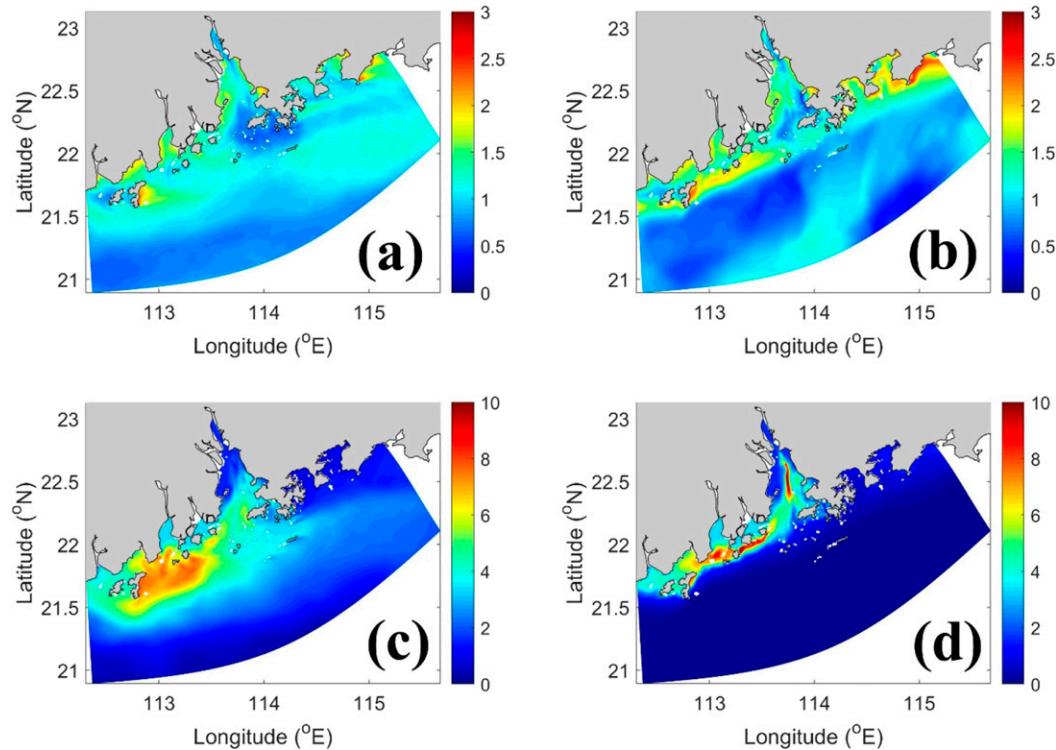


FIG. 4. The structures of SDs of the (a),(b) temperature and (c),(d) salinity in the (left) surface and (right) bottom layers that form the background model state.

adjacent shelf (Fig. 1). The survey period covered the spring–neap cycle and different wind forcing conditions, which aids in the identification of the variability of salinity, temperature, and circulation. At the same time, we conducted time series measurement of temperature, salinity, and current in the surface (1 m), middle (15 m), and bottom (25 m) layers using a buoy mooring located in the south of Hong Kong Island (Fig. 1). Only temperature and salinity data have been assimilated in this study.

Remote sensing SST data are also used in this study. Due to the differences in satellite source data and integration methods, the accuracy of SST products could differ, especially for the shelf and coastal waters. Two kinds of satellite SST products with different spatial resolutions were examined in our DA study. One is the Operational SST and Sea Ice Analysis (OSTIA) SST data (<http://ghrsst-pp.metoffice.gov.uk/ostia-website/index.html>) obtained from the National Centre for Ocean Forecast (NCOF) of the Met Office (Donlon et al. 2012), which is a global daily analysis gridded at a  $0.05^\circ \times 0.05^\circ$  horizontal resolution. The other is the Multi-Scale Ultra High Resolution (MUR) SST daily product (<https://registry.opendata.aws/mur/>) produced at the Jet Propulsion Laboratory (Chin et al. 2017). The MUR SST data are released on a uniform latitude–longitude grid with a sampling resolution of  $0.01^\circ$  or approximately 1 km.

#### b. Data assimilation experiments

We carried out a series of assimilation experiments using different types of observational data and assimilation time

windows to assess the DA effects (Table 1). The control run without DA (NoDA) of the model is used to identify the model anomaly fields and as a benchmark for comparison with the assimilation experiments. The model simulation period is from 10 to 30 July 2015, covering the entire field survey period. Due to the high variability of features in the waters of the PRE and the adjacent shelf, the selection of the assimilation time window is a key factor for successful assimilation. A shorter cycle length of DA has more frequent analyses and initialization of the model, and thus resolves better the highly variable features in the waters. However, the disadvantage of this is that the increasing noise in the initial condition would be accumulated during the model continuous integrations with frequent re-initialization. Although a longer cycle provides better temporal observation coverage, it can introduce a larger analysis increment with temporal representation errors, and observational data might be overfitted (Sandery 2018). To evaluate the influences of the assimilation time window on model forecasts and the effects of assimilating CTD data, two schemes for the observation time window were designed: EXP1 and EXP2 (Fig. 3 and Table 1). EXP1 assimilated the observed CTD data at 0000 (local time  $T$ ) every day, in which the data collected in the previous 24 h ( $T - 24$  h) were assimilated into the model. The model ran a 1-day simulation ( $T + 24$  h) to the next DA time (0000 of the second day) and repeated this assimilation process. The CTD data observed within this 24-h simulation cycle ( $T + 24$  h) were not assimilated into the model

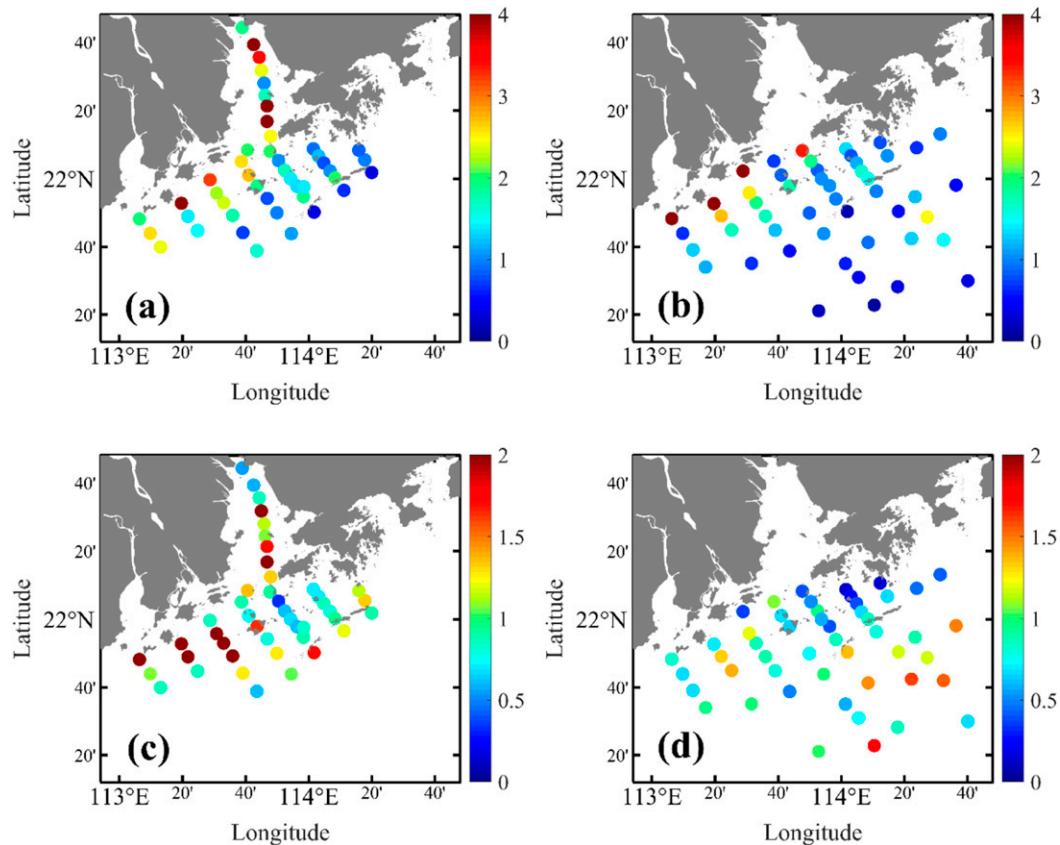


FIG. 5. RMSEs of (a),(b) salinity and (c),(d) temperature of the water column at the observation stations for the model without DA on (left) leg 1 and (right) leg 2.

yet and was to be used for validating the model results of this simulation.

Because the tidal currents with diurnal periods are dominant in the PRE waters, we separate the waters into states of flood and ebb tides, respectively. In EXP2, we used an assimilation time window that agreed with the flood–ebb tidal cycle. The observation data were classified into ebb and flood types based on the observed time. In this scheme, we first assimilate the ebb tide data in the previous 3 days ( $T - 72$  h) (before the DA time) on the maximum ebb tide ( $T_e$ ) of the model as initial condition, and the model ran until the next DA time (maximum flood tide,  $T_f$ ). We then used the flood tide data in the previous 3 days ( $T - 72$  h) to update the initial condition on that maximum flood tide ( $T_f$ ), and subsequently the model ran a 3-day simulation until the next DA time (maximum ebb tide on the third day). We repeated this assimilation process. Compared with EXP1, the advantage of this scheme is that it can reduce the mismatch representation of observation at the analysis time caused by asynchronous observation and provide more observations for better coverage when reconstructing the initial condition.

As the CTD data are sparsely observed, the satellite SST is a significant supplement to get the initial condition of temperature updated over the entire domain. To assess

the influence of remote sensing SST data on the DA, we respectively assimilated the OSTIA SST (EXP3) and MUR SST (EXP4) data. Besides, the time series measurements using buoy mooring provided useful information for the temporal variation in the estuary–shelf system. Assessing the impact of assimilating the near-real-time buoy data (EXP5) not only helps to improve the model simulation but also to provide guidance for where to deploy buoys. We also assessed the outcome of assimilation by using all buoy, CTD, and satellite SST data in EXP6. EXP2 and EXP6 are assimilated at the flood/ebb time, and the rest use a 24-h assimilation time window.

### c. Data assimilation settings

The water properties in the coastal area are highly variable in time and space under multiple forces. The seasonal variability is strong in the region due to the monsoon forcing, and we thus only consider the samples in summer. We generated the ensemble from the model runs without data assimilation in July of 2014 and 2015 and extracted the results every 12 h, in which the samples hold the information of model variability in the temporal scale of semidiurnal tides. These samples were then combined as an ensemble with a size of 120. The structures of standard deviations (SD) of the temperature and salinity in the surface and bottom layers from the background

TABLE 2. RMSE of salinity in the NoDA, EXP1, and EXP2 experiments in region B and transect A on leg 1 and region B and region C on leg 2.

	Expt	Region B on leg 1	Transect A on leg 1	Region B on leg 2	Region C on leg 2
Salinity (PSU)	NoDA	2.27	2.33	2.32	0.87
	EXP1	2.32	2.80	2.17	0.70
	EXP2	1.24	2.09	1.50	0.50
Temperature (°C)	NoDA	1.58	1.08	0.88	0.95
	EXP1	1.31	0.85	0.97	0.99
	EXP2	0.80	0.68	0.63	0.79

model state are shown in Fig. 4. The relatively large SD of temperature occurs around the nearshore waters in both surface and bottom layers (Figs. 4a,b), indicating the strong tidal and river discharge effects. Meanwhile, the larger SD of surface temperature also presents in the eastern shelf region and bottom temperature on the south boundary, suggesting the wind and boundary effects, respectively. For the surface salinity, the larger SD is primarily located in the western coastal region (Fig. 4c), while for the bottom salinity, the larger SD is much closer to the coastline in the west and inside the estuary (Fig. 4d), indicating combined strong tidal and river discharge effects. In basic terms, the background model samples can reflect the temperature and salinity errors from the combined effects of freshwater discharges, tides, winds, and open boundary conditions.

To increase the rank of the background error covariance matrix and decrease the spurious correlations of ensemble covariances, horizontal localization has been applied in the DA experiments. Localization ensures that measurements will only affect states in the waters within a certain distance from the measurement location. The covariance localization is a procedure where the covariance is multiplied point by point by a fifth-order function (Gaspari and Cohn 1999). The in situ data are scattered with different spatial distances, but the distances between the stations are generally less than 20 km. Here, a uniform horizontal correlation scale of 20 km is adopted based on sensitivity experiments. In the vertical direction, to avoid the assimilation of more than one measurement for the same position, observations averaged in model layers were used for every profile. In general, the satellite remote sensing SST and the data of temperature/salinity collected from CTD profilings are different and they have different observation error variances. For observational errors in the temperature/salinity profiles, we adopted the functional shapes for observation errors as those from Xie and Zhu (2010) and Stammer et al. (2002), which were based on uncertainty estimations from analyses of global observation data of Levitus et al. (1994) and Levitus and Boyer (1994), but with a larger SD because of the higher representativity errors in the coastal zone. We assumed that the observational errors are Gaussian with zero mean and uncorrelated between temperature and salinity and between different levels. Thus, the SDs of observational errors of temperature and salinity are assumed to be

represented as a function of the depth  $d$  (m) as the following equations:

$$\sigma_{\text{temp}} = 0.05 + 0.45 \exp\left(\frac{-d}{500}\right) \quad \text{and} \quad (6)$$

$$\sigma_{\text{salt}} = 0.3 + 0.2 \exp\left(\frac{-d}{125}\right). \quad (7)$$

Xie et al. (2008) showed that the satellite remote sensing SST products near the coastal waters have a larger uncertainty in the SCS, in which the uncertainty has an increasing trend with decreasing water depth. To well consider the representative error, we exaggerated the raw observation error from the data provider by a factor of 3. Thus, the SD of SST is represented as this equation:

$$\sigma_{\text{SST}} = 3\sigma_o + 0.5 \exp\left(\frac{-d}{30}\right). \quad (8)$$

Here,  $\sigma_o$  represents the estimated error SD of SST.

#### 4. Results

In the following sections, we explore the impact of different assimilation schemes on the forecast of the water in the estuary–shelf off Hong Kong. The root-mean-square error (RMSE) and the correlation coefficient between model results and observations are used to evaluate the DA improvement. All the model results from the assimilative experiments are compared with the not-yet-assimilated observations. For instance, in the EXP1 and EXP2, we assimilated the CTD data observed 24 or 72 h ( $T - 24$  h or  $T - 72$  h) before the analysis time  $T$  and ran the model simulation for 24 or 72 h ( $T + 24$  h or  $T + 72$  h). And then the CTD data observed within this period were not assimilated into the model yet and will be used to validate the model results of this simulation (Table 1). Thus, these observations are semi-independent (Liu et al. 2009).

The RMSE represents the average magnitude of the model–observation differences, defined as

$$\text{RMSE} = \left[ \frac{1}{N} \sum_{i=1}^N (\eta_{\text{mo}} - \eta_{\text{ob}})^2 \right]^{1/2}. \quad (9)$$

The correlation coefficient is defined as

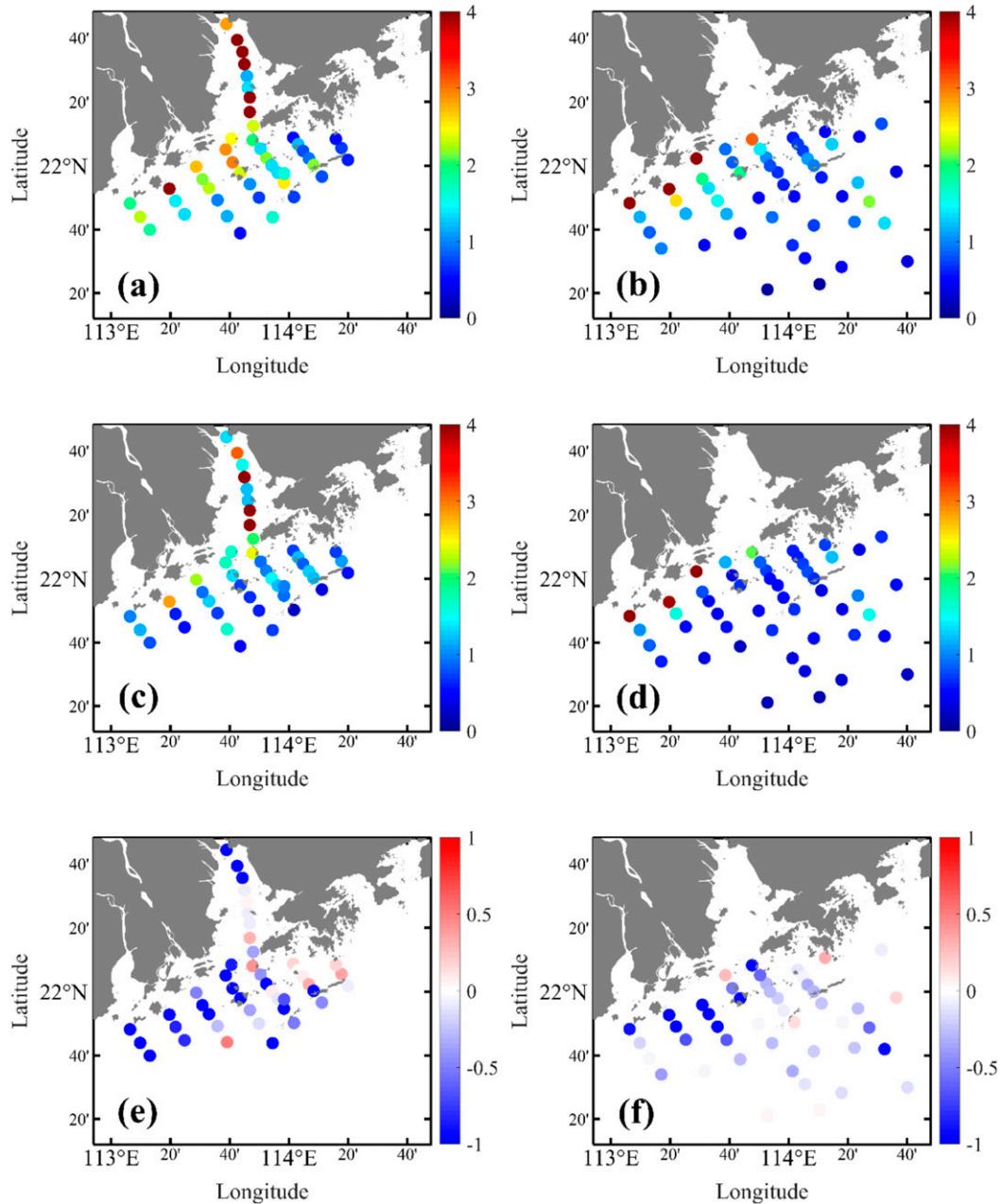


FIG. 6. For (left) leg 1 and (right) leg 2, RMSEs of salinity of the water column at the observation stations for (a),(b) EXP1 and (c),(d) EXP2, along with (e),(f) the differences in temperature RMSEs between EXP2 and EXP1.

$$R = \frac{\sum_{i=1}^N [(\eta_{mo} - \bar{\eta}_{mo})(\eta_{ob} - \bar{\eta}_{ob})]}{\left[ \sum_{i=1}^N (\eta_{mo} - \bar{\eta}_{mo})^2 \sum_{i=1}^N (\eta_{ob} - \bar{\eta}_{ob})^2 \right]^{1/2}}, \quad (10)$$

where  $\eta_{mo}$  and  $\eta_{ob}$  represent the modeled and observed variables, respectively; and the overbar denotes the temporal average;  $N$  is the number of records.

A model run without DA is first used to identify the simulation bias. We discuss the results for leg 1 and leg 2 separately,

to identify the responses to different tidal and wind conditions during the survey period. Figure 5 presents the RMSEs of salinity and temperature of the water column at the observation stations for two legs in NoDA. On leg 1 with a spring tide and weak wind, the larger RMSEs of salinity and temperature mainly occur inside the estuary (transect A) and the western coastal waters (region B) over the shelf (Figs. 5a and 5c; transect A and region B shown in Fig. 1), where relatively strong tidal and river discharge effects occur in the nearshore waters. Relatively small RMSEs are found in the offshore

waters in region C (Fig. 1). On leg 2 with a neap tide and strong wind, the RMSEs of salinity at the stations are still larger in the western coastal waters (region B) and smaller in the offshore region (Fig. 5b). Despite the neap tide and strong wind during this leg, the RMSE of salinity in region B of the tidally affected zone is very close between leg 1 and leg 2, while there is a much smaller RMSE (0.87 PSU, where PSU is a “practical salinity unit”; Table 2) in offshore waters (region C). As for temperature, the larger RMSEs exist mainly in the western coastal waters on leg 1 (Fig. 5c), although the RMSE of 0.88°C on leg 2 during neap tide is much smaller than that of 1.58°C on leg 1 during spring tide (Table 2). We also see a slight increase in the RMSE of temperature (0.95°C) in the offshore waters (region C) from that (0.88°C) in the coastal region (region B). Overall, the larger model simulation errors occur primarily in the shallow coastal regions where dynamical instabilities are most intense under combined forcings of freshwater discharges, tides, wind, and boundary conditions (external flux).

#### a. Flood–ebb assimilation time windows

As discussed above, the tidal currents play an important role in changing the salinity and temperature structures in the PRE and over the shallow shelf region. As a result, the observed salinity and temperature data have large variability during the flood–ebb tidal cycle. We first conduct CTD DA in EXP1 with a fixed 24-h assimilation time window followed by adding a varied assimilation time window based on the flood–ebb tidal cycle in EXP2 (Table 1). Figure 6 shows the RMSEs of salinity for the water column and their differences at the sampling stations in the PRE and adjacent shelf during leg 1 and leg 2 for EXP1 and EXP2. The patterns of RMSE in salinity of EXP1 are similar to those in the NoDA case (Figs. 5a,b) for both leg 1 and leg 2 (Figs. 6a,b). The RMSE of salinity in region B and transect A are 2.32 and 2.80 PSU in EXP1, which are even larger than those (2.27 and 2.33 PSU) in the NoDA case (Table 2), implying that the model–observation mismatch at the analysis time can amplify the errors in the innovation vector and undermine the model simulation. During leg 2 with a neap tide, the RMSEs of salinity in EXP1 in regions B and C are slightly decreased compared with the NoDA case.

In EXP2 with the consideration of the tidal variation in the assimilation window, the RMSEs of salinity are greatly improved in the western coastal region and outside the estuary on leg 1 as compared with EXP1, in which most of the salinity RMSEs decrease to less than 1 PSU (Fig. 6c). For leg 2, the RMSEs of salinity at most stations are less than 1 (Fig. 6d). The differences in salinity RMSE between EXP2 and EXP1 are negative at most stations (Figs. 6e,f), suggesting a greater improvement in EXP2. Overall, the RMSEs of salinity at all stations are 1.57, 1.54, and 1.07 PSU in NoDA, EXP1, and EXP2 (Table 3), respectively. This indicates the RMSE of salinity is improved from 1.9% in EXP1 to 31.9% in EXP2. The correlation coefficient of salinity between the DA experiments and observations increases from 0.83 in NoDA to 0.91 in EXP2.

TABLE 3. RMSE and correlation coefficients of salinity and temperature in the NoDA, EXP1, EXP2, EXP4, and EXP6 experiments.

Variable	Expt	RMSE	RMSE	Correlation
			improvement (%)	
Salinity (PSU)	NoDA	1.57	—	0.83
	EXP1	1.54	1.9	0.83
	EXP2	1.07	31.9	0.91
	EXP6	1.09	30.6	0.91
Temperature (°C)	NoDA	1.03	—	0.81
	EXP1	0.99	3.9	0.85
	EXP2	0.69	33.0	0.89
	EXP6	0.69	33.0	0.89

For temperature, the RMSEs of EXP1 in region B and transect A (1.31° and 0.85°C) on leg 1 are slightly lower than those (1.58° and 1.08°C) in NoDA (Table 2). However, the RMSEs of temperature in EXP1 become even larger (0.97° and 0.99°C) in both regions B and C than those (0.88° and 0.95°C) in NoDA during leg 2. As a result, the patterns of RMSE in the temperature of EXP1 are also similar to those in NoDA (Figs. 5c,d) for both leg 1 and leg 2 (Figs. 7a,b). By adopting the flood–ebb tidal assimilation window in EXP2, the reduction in RMSE is remarkable, in which the RMSEs of temperature are lower than 1°C at most sampling stations on both leg 1 and leg 2 (Figs. 7c,d). The differences in temperature RMSE between EXP2 and EXP1 are negative at most stations, indicating the RMSEs are reduced in EXP2 (Figs. 7e,f). The RMSE of temperature at all stations is 1.03°, 0.99°, and 0.69°C in NoDA, EXP1, and EXP2 (Table 3), respectively. The RMSE of temperature has been reduced by 33.0% in EXP2, but only by 3.9% in EXP1. The correlation coefficient of temperature at all stations increases from 0.81 in NoDA to 0.89 in EXP2.

The results demonstrate that the DA scheme with the variable flood–ebb assimilation time window minimizes the mismatch between assimilated observations and model forecasts at the analysis time, reducing the error in the estimation of the innovation vectors. The longer assimilation cycle and more accurate assimilation frequency or windows in EXP2 lead to high-quality analysis. The quality of simulations obtained by DA is largely determined by the coverage and quality of observations (She et al. 2007). Overall, the DA experiment with a varied assimilation time window based on the intratidal variation of the observational data shows a much greater improvement in model performance in the estuary and coastal regions than that with a fixed and shorter assimilation time window.

#### b. Assimilation of remote sensing SST data

The remote sensing SST data can potentially provide a valuable constraint by updating the initial condition systematically due to its better spatiotemporal coverage. Two experiments (EXP 3 and EXP4, Table 1) with SST data of different spatial resolutions are implemented to assess the

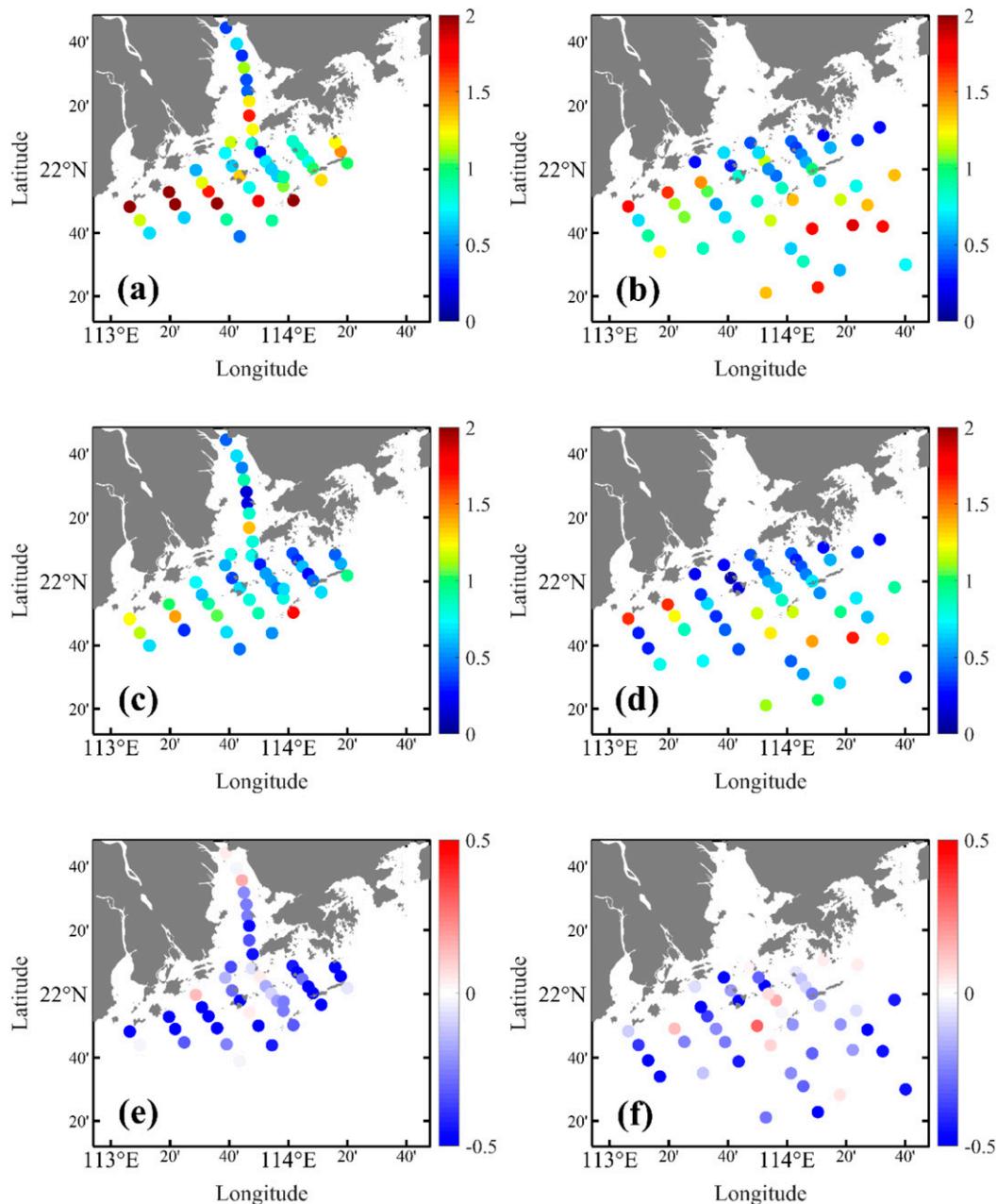


FIG. 7. As in Fig. 6, but for temperature.

influences of assimilating satellite SST data. The overall RMSE of surface temperature is  $0.75^{\circ}\text{C}$  for the MUR data, while it is  $1.04^{\circ}\text{C}$  for the OSTIA data (Table 4). The correlation coefficient between the MUR SST and CTD data is also greater than that of the OSTIA data. This suggests MUR SST has a better agreement with the CTD data as compared with OSTIA SST data. Meanwhile, the MUR SST data have better coverage near the coastal region and inside the PRE. Figure 8 illustrates SST differences between the CTD observation and models of NoDA, EXP3 assimilating OSTIA SST data, and EXP4 assimilating MUR SST data. Apparently, the SST is warmer in

the estuary and western coastal region and colder outside the estuary than the observations in NoDA on leg 1 (Fig. 8a), in which the deviations of SST exceed  $1^{\circ}\text{C}$  at most stations. On leg 2 under the strong upwelling-favorable wind, the model without DA overestimates the SST on the west coast and underestimates outside the estuary (Fig. 8b). As a result of less coverage of OSTIA SST data inside the estuary and nearshore region, the SST deviation in EXP3 shows a similar pattern to that in NoDA on leg 1 (Fig. 8c). On leg 2 of EXP3, the SST is decreased at most stations with SST deviations within  $1^{\circ}\text{C}$  by assimilating OSTIA SST data (Fig. 8d). For the EXP4, the SST

TABLE 4. RMSE and correlation coefficients of surface temperature for the OSTIA data, MUR data, and NoDA, EXP3, and EXP4 experiments.

Source	RMSE (°C)	RMSE improvement (%)	Correlation coef
OSTIA SST	1.04	—	0.72
MUR SST	0.75	—	0.82
NoDA	0.82	—	0.72
EXP3	0.72	12.2	0.74
EXP4	0.57	30.5	0.83

deviations are reduced within 1°C at most stations on both leg 1 and leg 2 by assimilating higher-resolution MUR SST data (Figs. 8e,f). Even under the strong upwelling-favorable wind, the SST simulation errors can be reduced effectively by assimilating OSTIA and MUR SST data on leg 2. Table 4 shows the RMSE and correlation coefficients of SST at all stations in NoDA, EXP3, and EXP4. The RMSE of SST has been reduced by 30.5% in EXP4, decreasing from 0.82° to 0.57°C, but only by 12.2% in EXP3 (Table 4). The correlation coefficient between the model and observation increases from 0.72 in NoDA to 0.83 in EXP4. Therefore, the higher quality and better coverage of MUR SST data lead to a greater improvement in the surface temperature of the model simulation.

Figure 9 displays the time-averaged SST for leg 1 and leg 2 from the remote sensing observation, NoDA, EXP2, and EXP4. An average of the MODIS, OSTIA, and MUR SST data is considered as the remote sensing observation. On leg 1, warmer water appears on the western side and in the offshore region where the temperature is higher than 30°C. Meanwhile, a colder-water pool appears around Hong Kong and extends to the eastern side (Fig. 9a). Without DA, the pattern of SST is consistent with the observation. However, the SST is warmer in the western and offshore regions and colder around Hong Kong and over the eastern shelf in NoDA (Fig. 9c). On leg 2 under the upwelling-favorable wind, the SST drops to less than 29°C over the entire domain (Fig. 9b). The relatively warm water is located on the eastern side and the range of colder SST around Hong Kong is enhanced by the upwelling water. Without DA, a warmer-water belt lies on the western side and extends eastward (Fig. 9d). Relatively warm water occupies in the estuary and the western shelf due to the river discharge. After assimilating the CTD data, the time-averaged SST in EXP2 shows a similar pattern to that in NoDA (Figs. 9e,f), indicating the DA effects are unsustainable when only sparse CTD profiles are assimilated and will be dissipated by the strong multiple forcings of winds, river discharges, tides, and open boundary conditions. In EXP4, the SST has reproduced two warmer water regions and the intensity of the colder water region is reduced around Hong Kong on leg 1 (Fig. 9g). For leg 2, the warmer-water belt is smoothed, and the SST is consistent with the observation (Fig. 9h). Thus, assimilating the MUR SST data can provide sustainable improvement of SST in the simulation.

### c. Assimilation of buoy data

We examine the effect of assimilating time series buoy data only in EXP5 with a daily assimilation frequency. During the analysis of the assimilation cycle, observations that fall within a 30-min window on either side of the analysis time are used. The time series temperature and salinity in the surface, middle, and bottom layers for the buoy observation, NoDA, and EXP5 at the buoy location are shown in Fig. 10.

The surface water has lower salinity and higher temperature during spring tide on leg 1 due to the spreading of freshwater by the strong tidal current, while it has a higher salinity and lower temperature during neap tide with the upwelling-favorable wind on leg 2 (Figs. 10a,b). The model without DA and EXP5 can reproduce these variabilities caused by the spring–neap tidal cycle. The RMSE of surface salinity has further been reduced by 53.9% in EXP5, decreasing from 3.30 PSU in NoDA to 1.52 PSU in EXP5 (Table 5). Meanwhile, the RMSE of temperature reduces from 0.89°C in NoDA to 0.59°C in EXP5 with a 33.7% improvement. For the middle and bottom layers, the water salinity and temperature are also influenced by the spring–neap tide cycle and the upwelling-favorable wind. The model in EXP5 reproduces two lower salinity periods at subsurface on 14 and 21 July. In particular, the model also captures an increasing salinity due to the upwelling-favorable wind after 21 July during leg 2. However, the model without DA shows smaller variabilities than those from EXP5 in the middle and bottom layers. As a result, the RMSEs of salinity in the middle and bottom layers have been reduced by 50.9% and 16.7%, respectively (Table 5). Similar to the salinity, the DA experiment has greater improvements in the temperature in the middle and bottom layers, reducing 53.6% and 66.7% of the temperature RMSE, respectively. After a long-time simulation, the errors of the simulation without DA are increased in the last few days of July, while EXP5 assimilating the time series data can constrain the model dispersion and reduce forecast errors.

Figure 11 shows the differences in RMSEs of the salinity and temperature of the water column at the observation stations within a 40-km radius (twice of the localization radius) of the location of the buoy between EXP5 and NoDA on leg 1 and leg 2. A prominent reduction in RMSE of salinity (negative value) occurs around the buoy location when the wind is weak on leg 1 (Fig. 11a). When the strong southwesterly wind prevails during leg 2, the reduction zone of salinity RMSE is extended alongshore outside the estuary (Fig. 11b). For the temperature, assimilating buoy data also reduce the RMSEs around the buoy location. (Fig. 11c). Meanwhile, the RMSEs of temperature are obviously reduced around the buoy location and on the eastern shelf in EXP5 during leg 2 (Fig. 11d). Apparently, the improvements of both salinity and temperature are more significant on leg 2 than on leg 1. The spatial impact by assimilating buoy data may well be linked with the prevailing wind during different legs of the research cruise.

The depth-averaged horizontal currents are relatively weak under the weaker upwelling-favorable wind on leg 1, while strong eastward currents under strong upwelling-favorable wind dominate the entire domain on leg 2 (Figs. 12a,b).

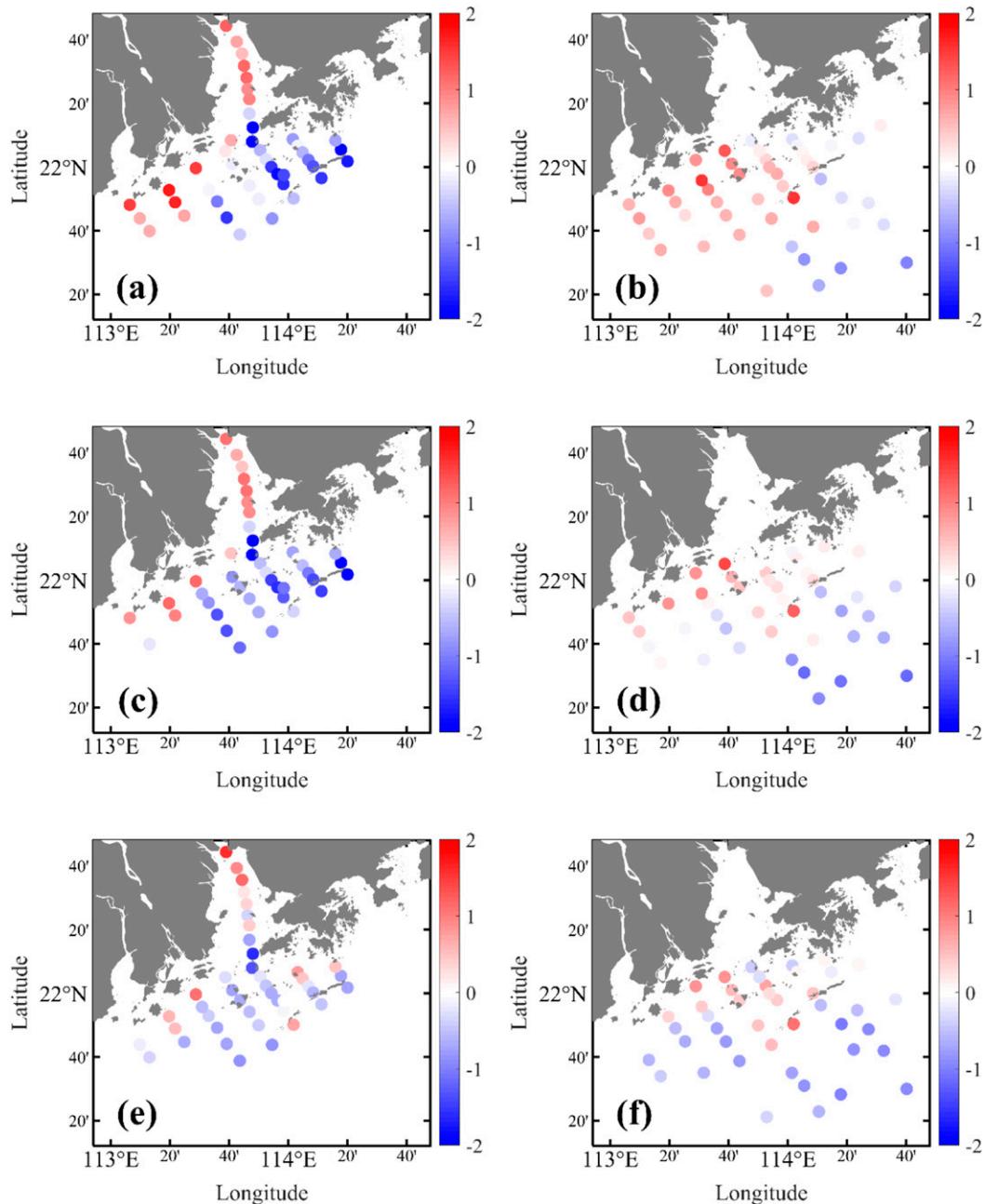


FIG. 8. Sea surface temperature differences with respect to in situ data from (left) leg 1 and (right) leg 2 for the (a),(b) NoDA; (c),(d) EXP3; and (e),(f) EXP4.

The spatial correlations of the simulated salinity and temperature at the buoy location indicate the footprint, or region of influence, by assimilating data from the buoy. The cross-correlation structures for the surface temperature and salinity show that the maximum correlation occurs around the buoy location and in the direction of the wind-driven upwelling current (Figs. 12c,d). The fluctuation of upwelling wind, that is, upwelling and relaxation from upwelling, may also extend the buoy influence upstream when upwelling-favorable wind relaxes (Gan and Allen 2002).

For the bottom layer, the maximum correlation regions of salinity and temperature run approximately parallel to the isobaths outside the estuary with a strong cross-shelf gradient on the shelf (Figs. 12e,f), implying the influence of bottom boundary conditions.

#### d. Combined buoy, CTD, and remote sensing SST data

To investigate whether assimilating all three kinds of observational data can achieve the best forecast results, a DA

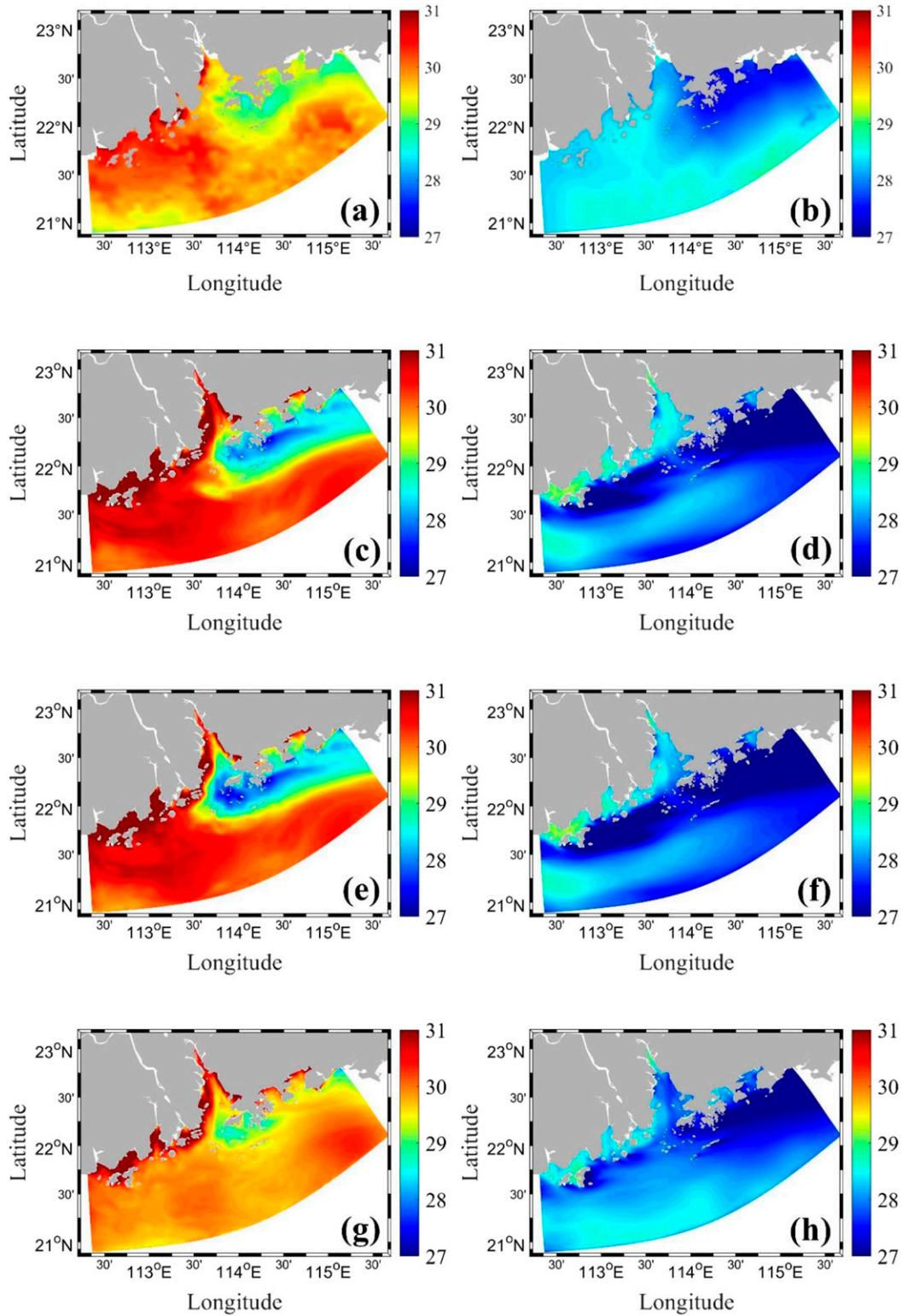


FIG. 9. Time-averaged SST for (left) leg 1 and (right) leg 2 from (a),(b) satellite observation; (c),(d) NoDA; (e),(f) EXP2; and (g),(h) and EXP4.

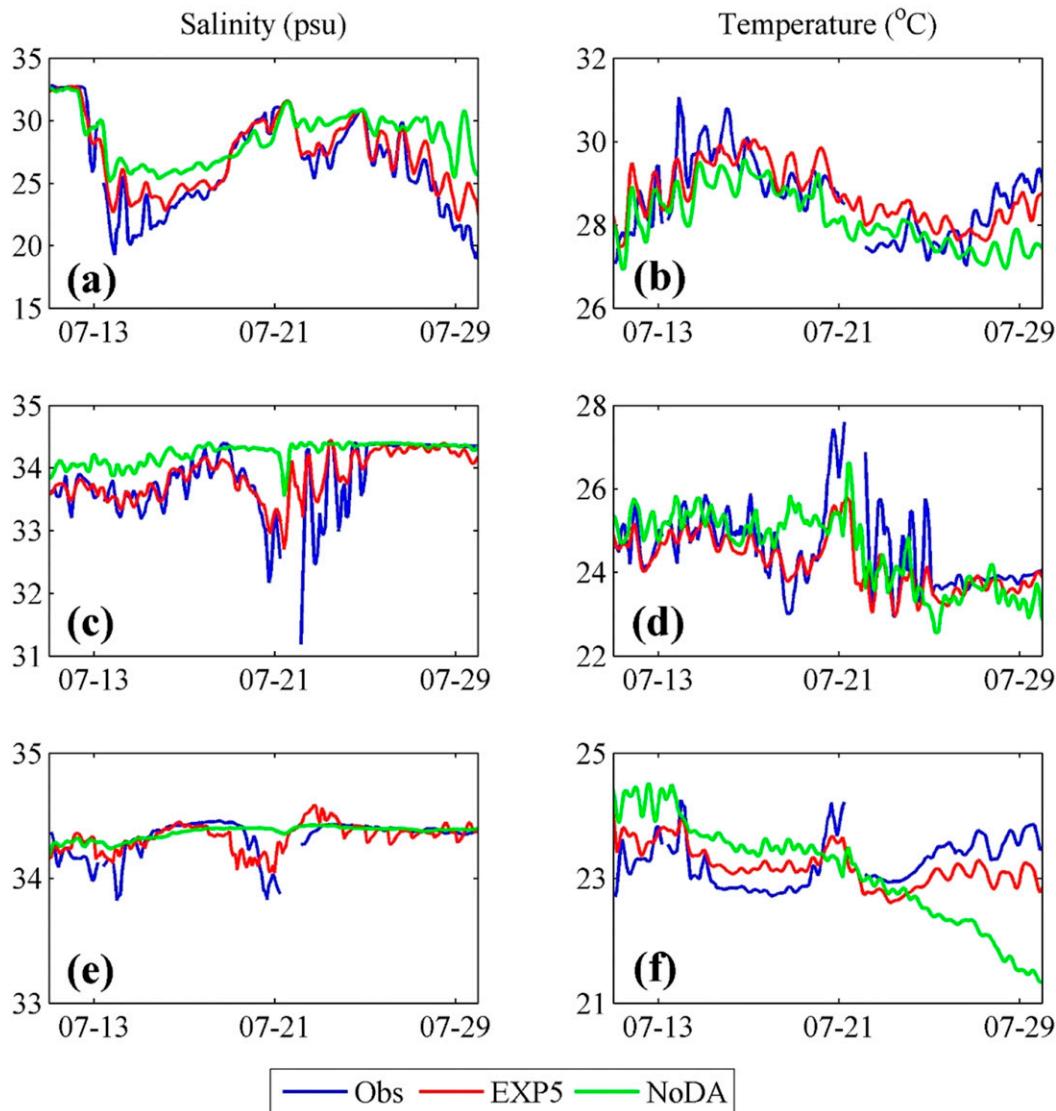


FIG. 10. Time series (left) salinity and (right) temperature in the (a),(b) surface (1 m); (c),(d) middle (15 m); and (e),(f) bottom (25 m) layers for the observation (blue), NoDA (green), and EXP5 (red) at the buoy location (shown in Fig. 1).

experiment EXP6 combining buoy, CTD, and remote sensing SST data is conducted. The RMSEs of salinity and temperature and the time-averaged SST for leg 1 and leg 2 in EXP6 are shown in Fig. 13. The RMSEs of salinity and temperature in EXP6 are greatly improved on the west coast and outside the estuary for both leg 1 and leg 2, in which most of the RMSEs of salinity decreased to less than 1 PSU (Figs. 13a,b). Similar to salinity, the RMSEs of temperature are lower than 1°C (Figs. 13c,d). In short, the patterns of RMSEs for both salinity and temperature in EXP6 are similar to those in EXP2 (Figs. 6c,d and 7c,d). For the SST, EXP6 has similar structures to those of EXP4 assimilating MUR SST for both leg 1 and leg 2 (Figs. 13e,f). Therefore, the combined assimilation can take advantage of both the vertical improvement from the assimilation of CTD data and the improvement of the

upper-layer temperature from the assimilation of remote sensing SST.

To illustrate the overall performance of these assimilating schemes, the vertical profiles of the RMSEs of salinity and

TABLE 5. RMSE of salinity and temperature in the surface, middle, and bottom layers in the NoDA and EXP5 experiments.

Variable	Expt	Surface	Middle	Bottom
Salinity (PSU)	NoDA	3.30	0.57	0.12
	EXP5	1.52	0.28	0.10
	Improvement (%)	53.9	50.9	16.7
Temperature (°C)	NoDA	0.89	1.25	1.26
	EXP5	0.59	0.58	0.42
	Improvement (%)	33.7	53.6	66.7

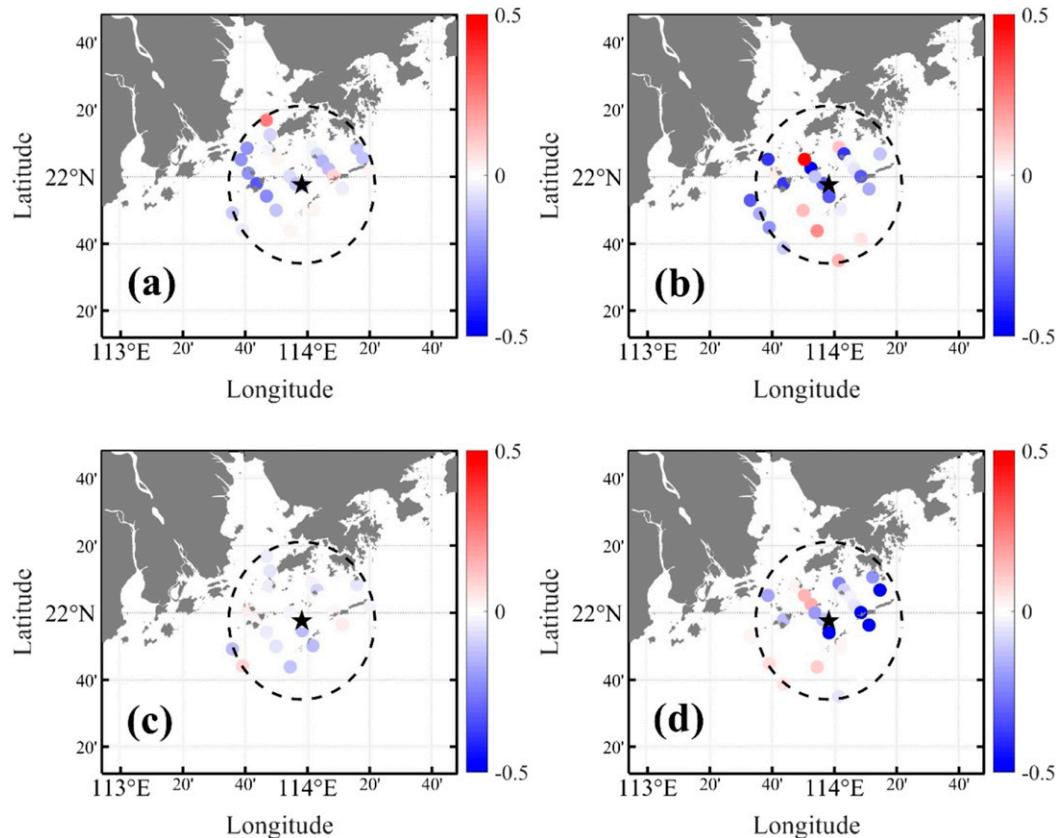


FIG. 11. Differences in RMSEs of the (a),(b) salinity and (c),(d) temperature of the water column at the observation stations within a 40-km radius of the location of the buoy between EXP5 and NoDA on (left) leg 1 and (right) leg 2. The star indicates the buoy location, and the dashed circle indicates the range of a 40-km radius.

temperature calculated from the data at the same depths of all stations in the experiments of NoDA, EXP1, EXP2, EXP4, and EXP6 are presented in Fig. 14. The RMSE profile of salinity in EXP6 is nearly overlaid on that in EXP2, where the improvements are remarkable from near the surface to a depth of 30 m. For the water below a depth of 30 m, the RMSEs of salinity are less than 0.1 PSU in all experiments since the deep water is relatively stable with high salinity (Fig. 14a). For EXP4 only assimilating the MUR SST data, the RMSEs of salinity are worse than those in NoDA in the whole water column, indicating that assimilating remote sensing SST data has a negative impact on salinity simulation. That is because the salinity is largely controlled by river discharge in this coastal water near PRE, and less linked with temperature as compared with that in the open ocean. This is also the reason why we assimilate both salinity and temperature from CTD, rather than SST from remote sensing data alone, in the coastal waters such as PRE. The RMSEs of temperature are reduced in the whole water column in EXP6 relative to the NoDA experiment (Fig. 14b). It is interesting to note that only assimilating MUR SST can dramatically decrease the RMSE of temperature in surface waters. However, in the deep water (below 30 m), the RMSEs of temperature in EXP4 are larger than those in

NoDA (Fig. 14b). Assimilating SST data can have a significant improvement in the surface layer, because of better spatiotemporal coverage and high-quality of SST data at the surface. However, temperature data have generally poor spatiotemporal coverage and larger uncertainties in the deep water.

The RMSE of salinity in EXP6 is 1.09 PSU, which is close to the RMSE in EXP2 of 1.07 PSU (Table 3) and has been reduced by 30.6% relative to the NoDA experiment. The RMSE of temperature in EXP6 is 0.69°C, in which the improvement reaches 33.0% in the experiment period (Table 3). The correlation coefficient of temperature between EXP6 and observation is 0.89. The RMSE and correlation coefficient of temperature are both the same as those in EXP2. Although the RMSE of salinity in the combined experiment is a little larger than that in EXP2 due to the negative effect of assimilating remote sensing SST data, it is necessary to assimilate remote sensing SST into the ocean model as doing so can significantly improve the surface temperature systematically.

## 5. Discussion and summary

We have performed a series of DA experiments to investigate the favorable DA scheme to improving numerical simulation in

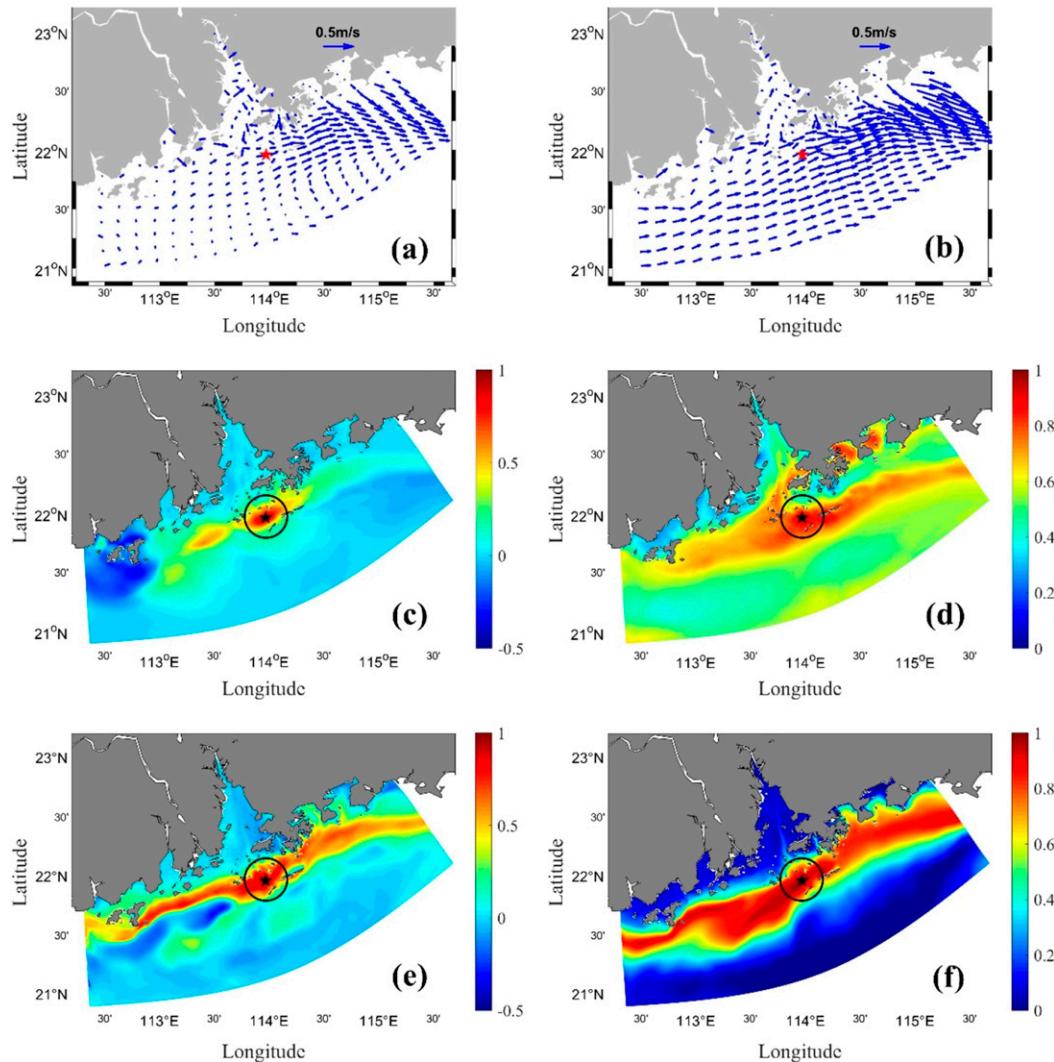


FIG. 12. Depth-averaged horizontal velocity vectors on (a) leg 1 and (b) leg2, and distribution of spatial correlation for surface (c) salinity and (d) temperature and bottom (e) salinity and (f) temperature with the buoy location during the cruise period. The black circle indicates the range of a 20-km localization radius.

the dynamically active coastal ocean around PRE under winds, tides, and freshwater forcing. For the estuarine circulation, the buoyancy forcing associated with the freshwater runoff from river inlets generates a horizontal baroclinic pressure gradient that increases with depth. With the improvement of temperature and salinity, the correction in the vertical stratification will yield horizontal density gradients that will be balanced by a correction in the baroclinic currents. Figure 15 displays the salinity and temperature along the transect A (Fig. 1) on 13 July for the observation, NoDA, and EXP6 combining buoy, CTD, and remote sensing SST data. The freshwater upstream flows to the mouth of the estuary, forming a strong mixing in the upper estuary and stratification in the lower estuary. Vertical differences in salinity of 8–12 PSU appear in the

lower estuary (Fig. 15a). Relatively warm water is located in the near-surface layer and cold water in the bottom layer, in which the vertical temperature differences can reach 5°C (Fig. 15b). The salinity in the upper estuary is overestimated in the model without DA, while that in the lower estuary is underestimated (Fig. 15c). The temperature in NoDA is higher than the observation in the upper estuary and lower in the bottom layer (Fig. 15d). For the case EXP6, the salinity in the upper estuary is decreased and the bottom salinity is increased forming a 10 PSU vertical difference in salinity (Fig. 15e), which is close to the observations. The temperature turns colder after DA in the bottom layers (Fig. 15f). The water column in the upper estuary is mainly controlled by thorough mixing resulting from the tidal mixing and freshwater. For the

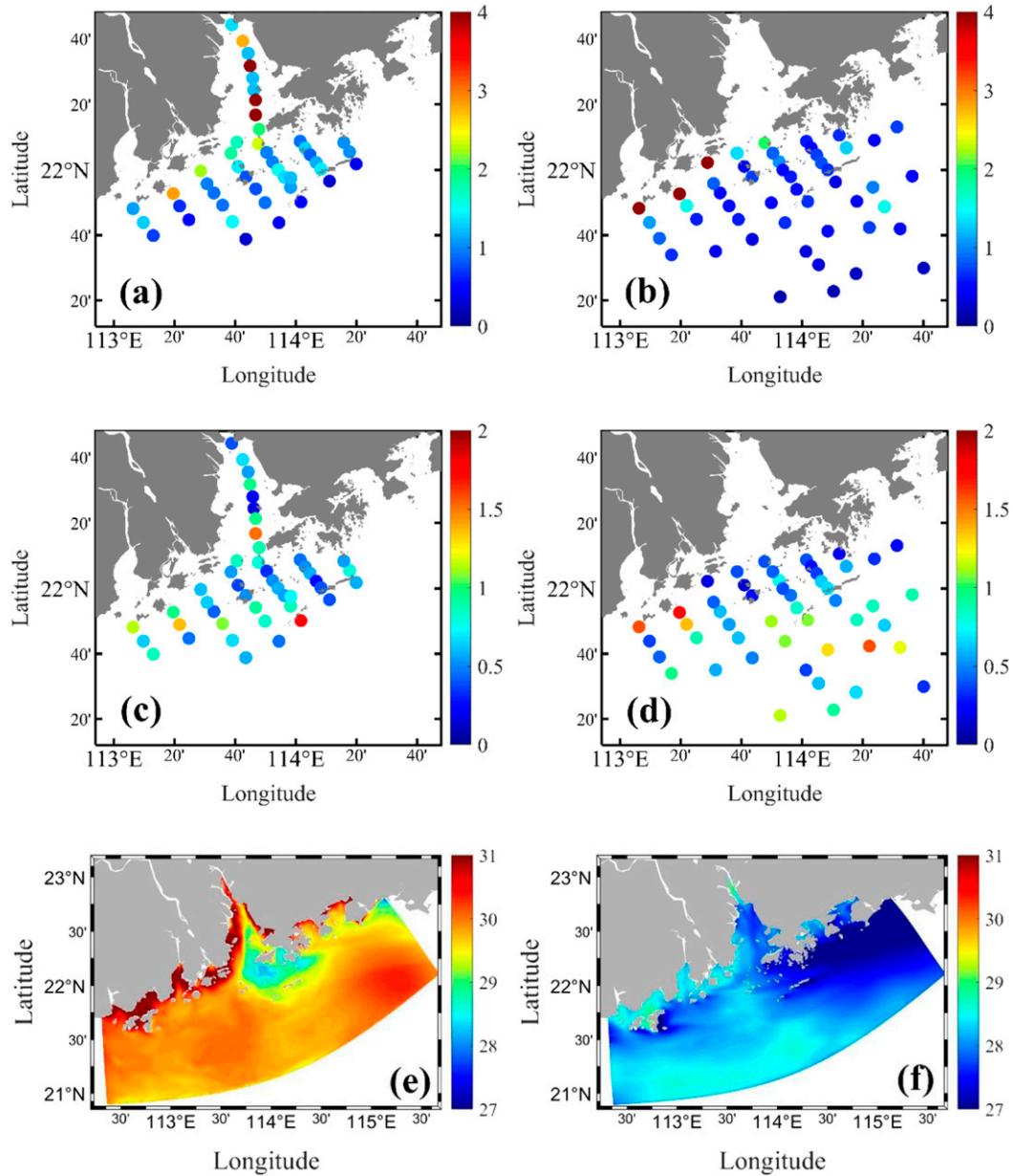


FIG. 13. For EXP6 for (left) leg 1 and (right) leg 2, RMSEs of (a),(b) salinity and (c),(d) temperature of the water column at the observation stations, along with (e),(f) time-averaged SST.

tidally averaged along-estuary circulation, a prominent two-layer exchange prevails with a seaward current in the surface layer and a landward current in the bottom layer (Fig. 15g). The intensity of the surface current is strengthened after DA and the bottom current intrudes landward more so than that without DA (Fig. 15h). The improvement to stratification induced by the assimilation can significantly enhance stratification and decrease mixing and thus strengthen the two-layer estuarine circulation after DA in the central estuary.

Another important issue is the covariance localization radius, which essentially controls the impact area of the

buoy observation. Localization can effectively ameliorate the spurious long-range correlations between the background and observations. Oke et al. (2007) suggested that the localization radius should be larger than the decorrelation length scale (i.e., where the correlation becomes insignificant) of the variables being updated. The localization radius is appropriate in the sense that it includes the significantly correlated region, while small or negative correlations away from the reference point are forced to have zero correlation. However, the shape of the significantly correlated region is not always a circle. For instance, the high-correlation region in the western waters is in the

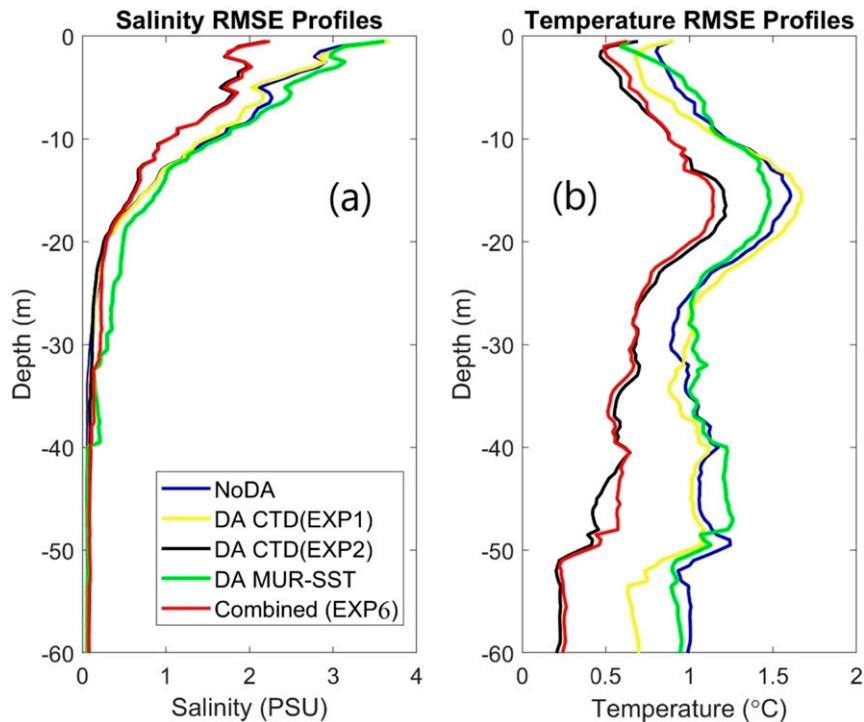


FIG. 14. Vertical profiles of the RMSEs of (a) salinity and (b) temperature calculated from the data at all stations at the same depths in the experiments of NoDA, EXP1, EXP2, EXP4, and EXP6.

shape of a long belt. Therefore, when we construct the localization radius of the influenced region, we need to consider a creative function for adapting the different shapes of the correlated region.

In this study, the EnOI algorithm is implemented to assimilate the CTD, buoy, and remote sensing SST data into a coastal ocean model in the PRE. We have assessed the impacts of assimilating these three kinds of data on the coastal water simulation. We found that the DA experiment with a flood–ebb assimilation time window based on the intratidal variation of the observation data reduces the errors in the estimation of the innovation vector caused by the model–observation mismatch at the analysis time and improve the simulation particularly in the estuary and shallow coastal regions. Statistically, the overall RMSEs of temperature and salinity between the DA forecasts and not-yet-assimilated observations have been reduced by 33.0% and 31.9% in the experiment period, respectively. Moreover, when the monsoon is relatively weak, a 3-day observation window provides better observation coverage, resulting in greater improvement. The selection of the DA time window associated with regional physics and the number of observation samples needed to spatially cover the variability in the simulated region are two key factors for successful assimilation in the dynamic estuary–shelf system.

The assimilation of remote sensing SST data appears to give a great improvement in the temperature field in shallow

water and the upper layer outside the estuary. Only assimilating remote sensing SST tends to have a negative effect on the salinity simulation. Because of the higher quality and better coverage of MUR SST data, assimilating higher-resolution MUR SST data has better performance than assimilating OSTIA SST data, which increases the percentage of surface temperature improvement from 12.2% to 30.5%. The near-real-time and continuous buoy data can further provide continuous correction to the model around the buoy location and beyond with the aid of coastal current. Analyses of the cross-correlation structures and the regions of influence of assimilating time series buoy data can significantly guide the distribution of buoy location in the field observation.

Assimilation of the combined buoy, CTD, and MUR SST data can take advantage of both the improvement in the water column from the assimilation of CTD profiles and the improvement of the upper-layer temperature from the assimilation of remote sensing SST. Assimilating CTD data, on the other hand, corrects the salinity bias that remains after assimilating the remote sensing SST data. Assimilating both the buoy and remote sensing SST data is essential in the model forecasting as it can continuously and systematically improve the model simulation. The RMSEs of temperature and salinity have been remarkably reduced by 33.0% and 30.6% in the experiment period, respectively. With the improvements of temperature and salinity, the errors of stratification and pressure gradient

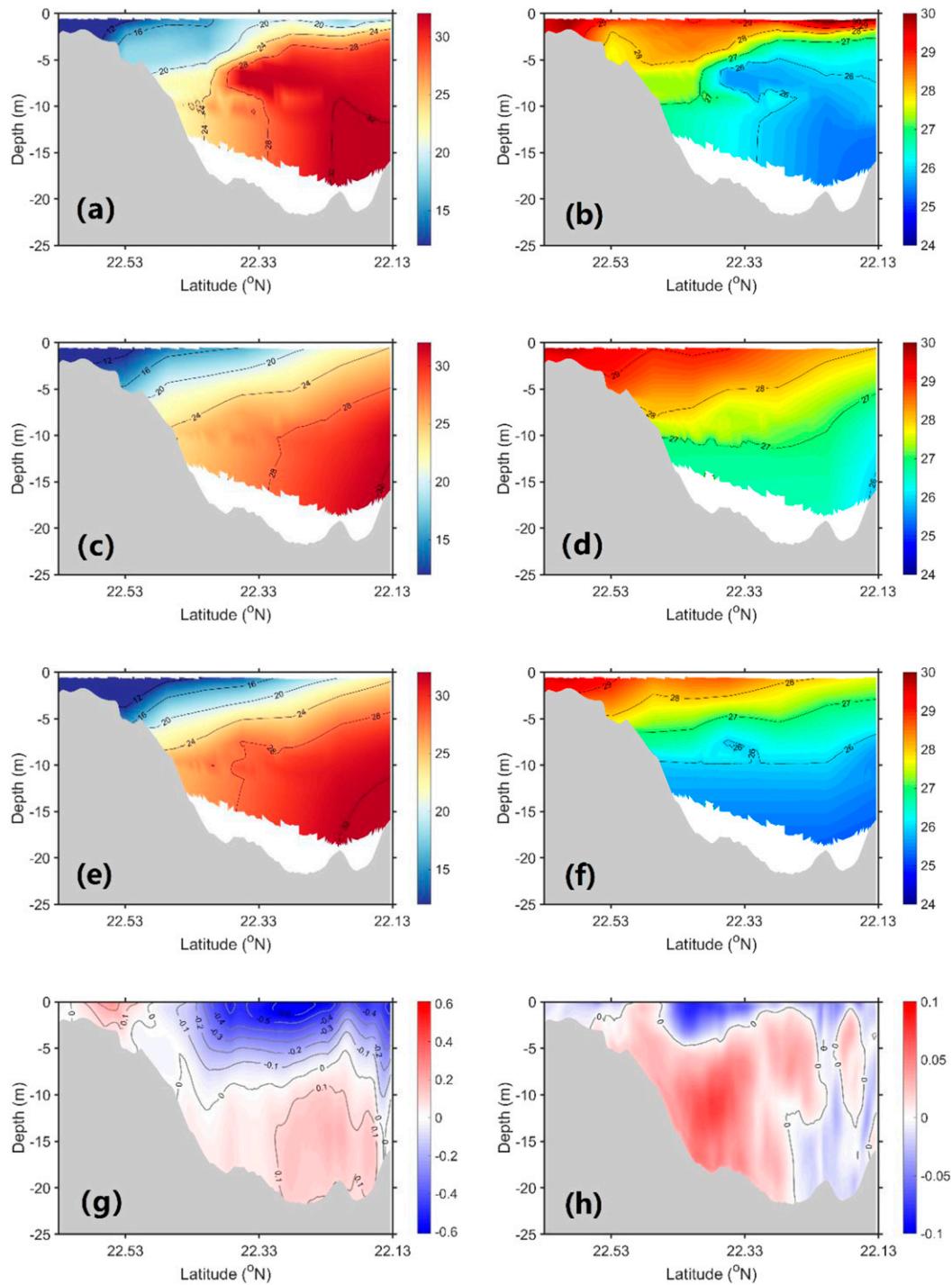


FIG. 15. Salinity and temperature distributions along transect A (Fig. 1) for (a),(b) the observation, (c),(d) NoDA, and (e),(f) EXP6 on 13 Jul; (g) the tidally averaged along-estuary velocity  $U$  ( $\text{m s}^{-1}$ ) on transect A for NoDA, and (h) the difference of the along-estuary velocity  $U$  between EXP6 and NoDA on 13 Jul.

are decreased, improving the model physics and thus estuarine circulation simulation.

This study provides new insight into the DA scheme in the active dynamics coastal waters and indicates that a

dynamics-oriented assimilation scheme is essential for the model forecasting in the dynamically active estuary–shelf system under the multiple forcings of winds, tides, and river discharges.

*Acknowledgments.* This work was supported by the Theme-based Research Scheme (T21-602/16-R, OCEAN\_HK project) of the Hong Kong Research Grants Council. The Center for Ocean Research in Hong Kong and Macau (CORE) is a joint ocean research center established by the Qingdao National Laboratory for Marine Science and Technology and the Hong Kong University of Science and Technology. We are grateful to the National Supercomputer Center of Tianhe-1 (Tianjin) and Tianhe-2 (Guangzhou). The model data for this study are produced by community model ROMS (<https://www.myroms.org/>). The data used in the study can be found online (<https://ocean.ust.hk/data/>).

## REFERENCES

- Backeberg, B., F. Counillon, J. A. Johannessen, and M. I. Pujol, 2014: Assimilating along-track SLA data using the EnOI in an eddy resolving model of the Agulhas system. *Ocean Dyn.*, **64**, 1121–1136, <https://doi.org/10.1007/s10236-014-0717-6>.
- Barth, A., A. Alvera-Azcarate, J. M. Beckers, J. Staneva, E. V. Stanev, and J. Schulz-Stellenfleth, 2011: Correcting surface winds by assimilating high-frequency radar surface currents in the German Bight. *Ocean Dyn.*, **61**, 599–610, <https://doi.org/10.1007/s10236-010-0369-0>.
- Botto, A., E. Belluco, and M. Camporese, 2018: Multi-source data assimilation for physically based hydrological modeling of an experimental hillslope. *Hydrol. Earth Syst. Sci.*, **22**, 4251–4266, <https://doi.org/10.5194/hess-22-4251-2018>.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2).
- Chin, T. M., J. Vazquez-Cuervo, and E. M. Armstrong, 2017: A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.*, **200**, 154–169, <https://doi.org/10.1016/j.rse.2017.07.029>.
- Counillon, F., and L. Bertino, 2009: High-resolution ensemble forecasting for the Gulf of Mexico eddies and fronts. *Ocean Dyn.*, **59**, 83–95, <https://doi.org/10.1007/s10236-008-0167-0>.
- Crosby, S. C., B. D. Cornuelle, W. C. O'Reilly, and R. T. Guza, 2017: Assimilating global wave model predictions and deep-water wave observations in nearshore swell predictions. *J. Atmos. Oceanic Technol.*, **34**, 1823–1836, <https://doi.org/10.1175/JTECH-D-17-0003.1>.
- De Mey, P., E. Stanev, and V. H. Kourafalou, 2017: Science in support of coastal ocean forecasting—Part 1. *Ocean Dyn.*, **67**, 665–668, <https://doi.org/10.1007/s10236-017-1048-1>.
- Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens. Environ.*, **116**, 140–158, <https://doi.org/10.1016/j.rse.2010.10.017>.
- Edwards, C. A., A. M. Moore, I. Hoteit, and B. D. Cornuelle, 2015: Regional ocean data assimilation. *Annu. Rev. Mar. Sci.*, **7**, 21–42, <https://doi.org/10.1146/annurev-marine-010814-015821>.
- Egbert, G. D., and S. Y. Erofeeva, 2002: Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Oceanic Technol.*, **19**, 183–204, [https://doi.org/10.1175/1520-0426\(2002\)019<0183:EIMOBO>2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2).
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, <https://doi.org/10.1029/94JC00572>.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Gan, J., and J. S. Allen, 2002: A modeling study of shelf circulation off Northern California in the region of the coastal ocean dynamics experiment: Response to relaxation of upwelling winds. *J. Geophys. Res.*, **107**, 3123, <https://doi.org/10.1029/2000JC000768>.
- , L. Li, D. Wang, and X. Guo, 2009: Interaction of a river plume with coastal upwelling in the northeastern South China Sea. *Cont. Shelf Res.*, **29**, 728–740, <https://doi.org/10.1016/j.csr.2008.12.002>.
- , J. Wang, L. Liang, L. Li, and X. Guo, 2015: A modeling study of the formation, maintenance, and relaxation of upwelling circulation on the northeastern South China Sea shelf. *Deep-Sea Res. II*, **117**, 41–52, <https://doi.org/10.1016/j.dsr2.2013.12.009>.
- , Z. Liu, and L. Liang, 2016: Numerical modeling of intrinsically and extrinsically forced seasonal circulation in the China seas: A kinematic study. *J. Geophys. Res. Oceans*, **121**, 4697–4715, <https://doi.org/10.1002/2016JC011800>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, <https://doi.org/10.1002/qj.49712555417>.
- Lai, W. F., J. Y. Pan, and A. T. Devlin, 2018: Impact of tides and winds on estuarine circulation in the Pearl River Estuary. *Cont. Shelf Res.*, **168**, 68–82, <https://doi.org/10.1016/j.csr.2018.09.004>.
- Levitus, S., and T. P. Boyer, 1994: *Temperature*. Vol. 4, *World Ocean Atlas 1994*, NOAA Atlas NESDIS 4, 117 pp.
- , R. Burgett, and T. P. Boyer, 1994: *Salinity*. Vol. 3, *World Ocean Atlas 1994*, NOAA Atlas NESDIS 3, 99 pp.
- Liu, D., J. Zhu, Y. Shu, D. Wang, W. Wang, and S. Cai, 2018: Model-based assessment of a northwestern tropical Pacific moored array to monitor the intraseasonal variability. *Ocean Modell.*, **126**, 1–12, <https://doi.org/10.1016/j.ocemod.2018.04.001>.
- Liu, Y., J. Zhu, J. She, S. Y. Zhuang, W. W. Fu, and J. Gao, 2009: Assimilating temperature and salinity profile observations using an anisotropic recursive filter in a coastal ocean model. *Ocean Modell.*, **30**, 75–87, <https://doi.org/10.1016/j.ocemod.2009.06.005>.
- , H. E. M. Meier, and L. Axell, 2013: Reanalyzing temperature and salinity on decadal time scales using the ensemble optimal interpolation data assimilation method and a 3D ocean circulation model of the Baltic Sea. *J. Geophys. Res. Oceans*, **118**, 5536–5554, <https://doi.org/10.1002/jgrc.20384>.
- Liu, Z., and J. Gan, 2016: Open boundary conditions for tidally and subtidally forced circulation in a limited-area coastal model using the Regional Ocean Modeling System (ROMS). *J. Geophys. Res. Oceans*, **121**, 6184–6203, <https://doi.org/10.1002/2016JC011975>.
- , and —, 2020: A modeling study of estuarine-shelf circulation using a composite tidal and subtidal open boundary condition. *Ocean Modell.*, **147**, 101563, <https://doi.org/10.1016/j.ocemod.2019.101563>.
- , —, and X. Wu, 2018: Coupled summer circulation and dynamics between a bay and the adjacent shelf around Hong Kong: Observational and modeling studies. *J. Geophys. Res. Oceans*, **123**, 6463–6480, <https://doi.org/10.1029/2018JC013830>.
- Losa, S. N., S. Danilov, J. Schroter, L. Nerger, S. Massmann, and F. Janssen, 2012: Assimilating NOAA SST data into the BSH operational circulation model for the North and Baltic Seas:

- Inference about the data. *J. Mar. Syst.*, **105–108**, 152–162, <https://doi.org/10.1016/j.jmarsys.2012.07.008>.
- Lyu, G., H. Wang, J. Zhu, D. Wang, J. Xie, and G. Liu, 2014: Assimilating the along-track sea level anomaly into the regional ocean modeling system using the ensemble optimal interpolation. *Acta Oceanol. Sin.*, **33**, 72–82, <https://doi.org/10.1007/s13131-014-0469-7>.
- Mao, Q., P. Shi, K. Yin, J. Gan, and Y. Qi, 2004: Tides and tidal currents in the Pearl River Estuary. *Cont. Shelf Res.*, **24**, 1797–1808, <https://doi.org/10.1016/j.csr.2004.06.008>.
- O’Dea, E. J., and Coauthors, 2012: An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf. *J. Oper. Oceanogr.*, **5**, 3–17, <https://doi.org/10.1080/1755876X.2012.11020128>.
- Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, and P. M. Kosro, 2002: Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res.*, **107**, 3122, <https://doi.org/10.1029/2000JC000511>.
- , P. Sakov, and S. P. Corney, 2007: Impacts of localisation in the EnKF and EnOI: Experiments with a small model. *Ocean Dyn.*, **57**, 32–45, <https://doi.org/10.1007/s10236-006-0088-8>.
- , G. B. Brassington, D. A. Griffin, and A. Schiller, 2008: The Bluelink ocean data assimilation system (BODAS). *Ocean Modell.*, **21**, 46–70, <https://doi.org/10.1016/j.ocemod.2007.11.002>.
- Pan, J., W. Lai, and A. T. Devlin, 2020: Channel-trapped convergence and divergence of lateral velocity in the Pearl River Estuary: Influence of along-estuary variations of channel depth and width. *J. Geophys. Res. Oceans*, **125**, e2019JC015369, <https://doi.org/10.1029/2019JC015369>.
- Sandery, P., 2018: Data assimilation cycle length and observation impact in mesoscale ocean forecasting. *Geosci. Model Dev.*, **11**, 4011–4019, <https://doi.org/10.5194/gmd-11-4011-2018>.
- Shchepetkin, A. F., and J. C. McWilliams, 2005: The Regional Oceanic Modeling System (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modell.*, **9**, 347–404, <https://doi.org/10.1016/j.ocemod.2004.08.002>.
- She, J., J. L. Hoyer, and J. Larsen, 2007: Assessment of sea surface temperature observational networks in the Baltic Sea and North Sea. *J. Mar. Syst.*, **65**, 314–335, <https://doi.org/10.1016/j.jmarsys.2005.01.004>.
- Shu, Y., J. Zhu, D. Wang, and X. Xiao, 2011a: Assimilating remote sensing and in situ observations into a coastal model of northern South China Sea using ensemble Kalman filter. *Cont. Shelf Res.*, **31**, S24–S36, <https://doi.org/10.1016/j.csr.2011.01.017>.
- , D. Wang, J. Zhu, and S. Peng, 2011b: The 4-D structure of upwelling and Pearl River plume in the northern South China Sea during summer 2008 revealed by a data assimilation model. *Ocean Modell.*, **36**, 228–241, <https://doi.org/10.1016/j.ocemod.2011.01.002>.
- Song, Y., and D. Haidvogel, 1994: A semi-implicit ocean circulation model using a generalized topography-following coordinate system. *J. Comput. Phys.*, **115**, 228–244, <https://doi.org/10.1006/jcph.1994.1189>.
- Sperrevelik, A. K., K. H. Christensen, and J. Rohrs, 2015: Constraining energetic slope currents through assimilation of high-frequency radar observations. *Ocean Sci.*, **11**, 237–249, <https://doi.org/10.5194/os-11-237-2015>.
- Stammer, D., and Coauthors, 2002: Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.*, **107**, 3118, <https://doi.org/10.1029/2001JC000888>.
- Tanajura, C. A. S., A. N. Santana, D. Mignac, L. N. Lima, K. Belyaev, and J. Xie, 2014: The REMO ocean data assimilation system into HYCOM (RODAS\_H): General description and preliminary results. *Atmos. Ocean. Sci. Lett.*, **7**, 464–470, <https://doi.org/10.1080/16742834.2014.11447208>.
- Xie, J., and J. Zhu, 2010: Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model. *Ocean Modell.*, **33**, 283–298, <https://doi.org/10.1016/j.ocemod.2010.03.002>.
- , —, and Y. Li, 2008: Assessment and inter-comparison of five high-resolution sea surface temperature products in the shelf and coastal seas around China. *Cont. Shelf Res.*, **28**, 1286–1293, <https://doi.org/10.1016/j.csr.2008.02.020>.
- , F. Counillon, J. Zhu, and L. Bertino, 2011: An eddy resolving tidal-driven model of the South China Sea assimilating along-track SLA data using the EnOI. *Ocean Sci.*, **7**, 609–627, <https://doi.org/10.5194/os-7-609-2011>.
- Zu, T. T., D. X. Wang, J. P. Gan, and W. B. Guan, 2014: On the role of wind and tide in generating variability of Pearl River plume during summer in a coupled wide estuary and shelf system. *J. Mar. Syst.*, **136**, 65–79, <https://doi.org/10.1016/j.jmarsys.2014.03.005>.