Targeted observations based on sensitive areas identified by CNOP to improve the thermal structure predictions in the summer Yellow Sea: operation in the field

Kun Liu[1], Wuhong Guo[1], Lianglong Da[2], Jingyi Liu[1], Huiqin Hu[1], and Baolong Cui[1]

[1]Qingdao National Laboratory for Marine Science and Technology
[2]Navy Submarine Academy

November 24, 2022

Abstract

Targeted observation is an appealing procedure for improving model predictions through the assimilation of additional collected measurements. However, studies on targeted observations in the oceanic field have been largely based on modeling efforts, and there is a need for field validating operations. Here, we report the results of a field program that is designed based on the sensitive areas identified by the Conditional Nonlinear Optimal Perturbation (CNOP) approach to improve the short-range (7 days) summer thermal structure prediction in the Yellow Sea. We found good spatial consistency in the locations of the identified sensitive areas among the hindcast and climatology runs. By introducing the technique of cycle data assimilation and the new concept of time-varying sensitive areas, we designed an observing strategy based on the identified sensitive areas, and conducted a set of Observing System Simulation Experiments prior to assessing the effectiveness of the plan on later observations. On this basis, the impact of targeted observations was investigated by a choreographed field campaign in the summer of 2019. The results of the in-field Observing System Experiments show that compared to conventional local data assimilation, conducting targeted observations in sensitive areas can double the benefits of data assimilation in thermal structure prediction. Furthermore, dynamic analysis demonstrates that the refinement of vertical thermal structures is mainly caused by the changes in the upstream horizontally advected temperature driven by the Yellow Sea Cold Water Mass circulation. This study highlights the effectiveness of targeted observations on reducing the forecast uncertainty in the ocean.

Targeted observations based on sensitive areas identified by CNOP to improve the thermal structure predictions in the summer Yellow Sea: operation in the field

Kun Liu[1,2], Wuhong Guo[1,2], Lianglong Da[1,2,*] , Jingyi Liu[1,2], Huiqin Hu[2], Baolong Cui[1,2]

[1]Navy Submarine Academy, Qingdao, China

[2]Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Corresponding author: Lianglong Da (da_lianglong@126.com)

Key Points:

- We first extend the scope of oceanic targeted observation to summer thermal structure prediction;
- We propose the observation strategy with a new concept of time-varying sensitive areas;
- We validate the benefit of the oceanic targeted observation in the actual field operation guided by the identified sensitive areas.

Abstract

Targeted observation is an appealing procedure for improving model predictions through the assimilation of additional collected measurements. However, studies on targeted observations in the oceanic field have been largely based on modeling efforts, and there is a need for field validating operations. Here, we report the results of a field program that is designed based on the sensitive areas identified by the Conditional Nonlinear Optimal Perturbation (CNOP) approach to improve the short-range (7 days) summer thermal structure prediction in the Yellow Sea. We found good spatial consistency in the locations of the identified sensitive areas among the hindcast and climatology runs. By introducing the technique of cycle data assimilation and the new concept of time-varying sensitive areas, we designed an observing strategy based on the identified sensitive areas, and conducted a set of Observing System Simulation Experiments prior to assessing the effectiveness of the plan on later observations. On this basis, the impact of targeted observations was investigated by a choreographed field campaign in the summer of 2019. The results of the in-field Observing System Experiments show that compared to conventional local data assimilation, conducting targeted observations in sensitive areas can double the benefits of data assimilation in thermal structure prediction. Furthermore, dynamic analysis demonstrates that the refinement of vertical thermal structures is mainly caused by the changes in the upstream horizontally advected temperature driven by the Yellow Sea Cold Water Mass circulation. This study highlights the effectiveness of targeted observations on reducing the forecast uncertainty in the ocean.

1. Introduction

The predictability of oceanic processes is limited since the ocean is an extremely complex dynamic system (Mu et al., 2017). The uncertainty of ocean forecasting can be reduced through the assimilation of observation data (Oke et al., 2015). Unlike observations on land, field-deployed oceanic observations are scarce and expensive. Thus, maximizings the individual impact of these limited measurements is a meaningful pursuit. Targeted observation is believed to be a suitable strategy for solving this problem (Farrara et al., 2013; Lermusiaux, 2007; Li et al., 2014; Majumdar, 2016).

Interest in the field of oceanic targeted observation has accelerated over the past few years (Baehr et al., 2008; Köhl & Stammer, 2004; Krama et al., 2012; Li et al., 2014; Morss & Battisti, 2004; Wang et al., 2013; Zhang et al., 2019). Morss & Battisti (2004) evaluated the effects of different numbers and locations of oceanic observations on the prediction of the El Niño-South Oscillation (ENSO) based on a series of Observing System Simulation Experiments (OSSEs). Baehr et al. (2008) studied the effects of different observing systems on the monitoring of the meridional overturning circulation in the North Atlantic. Krama et al. (2012) investigated the optimal observation locations for improving the predictability of the Kuroshio Extension. Li et al. (2014) reported an improvement in ocean prediction when utilizing targeted observations in the South China Sea (SCS) western boundary current region. Zhang et al. (2019) designed and evaluated a targeted observation network for improving upstream Kuroshio transport prediction. These studies confirmed the effectiveness of oceanic targeted observation; however, most of the relevant studies have been largely based on modeling efforts, and experiments in the field are necessary regarding both method validation and the cost-effectiveness evaluation.

A limited number of oceanic targeted observations in real scenarios have been reported in the literature (Curtin & Bellingham, 2009; Mourre & Alvarez, 2012; Shay et al., 2011). Curtin & Bellingham (2009) implemented the Autonomous Ocean Sampling Network (AOSN) field program in Monterey Bay and demonstrated that proper sampling is critical for both understanding and predicting ocean fields. To predict the local ocean circulation and potential pathways of spilled oil, Shay et al. (2011) carried out oceanographic surveys based on the positions of the exploded oil rig and the loop currents in the Gulf of Mexico. They found that the root-mean-square errors (RMSEs) of the simulated results were reduced by approximately 30% when the additional measurements were assimilated into the hindcast model. Guided by the optimal designed glider trajectory, which sets the trace of the error covariance matrix as criteria (Alvarez & Mourre, 2012), Mourre & Alvarez (2012) found that the data assimilation performance of the adaptive-sampling-driven glider data was better than that of the independent glider data in the same region, with an RMSE reduction of 18%.

However, none of the abovementioned in-field oceanic targeted observations were designed based on identified “sensitive areas”. Given a certain phenomenon, sensitive areas are the specific localized areas that are expected to contribute most in reducing the prediction uncertainties in the target region. In a study of storm tracking prediction, Montani et al. (1999) demonstrated that short-range prediction refinement can be increased from an average of 15% to approximately 37% if the observations are deployed in sensitive areas. Targeted observation studies in the atmospheric field started earlier and are more mature than those in the ocean. Among others, the Atlantic observing-system research and predictability experiment (THORPEX) is a remarkable program that concluded that targeting and assimilating observations in sensitive areas are effective in improving forecasts (Majumdar, 2016). Nevertheless, tests of targeted observations guided by identified sensitive areas in real at-sea scenarios are still lacking.

The identification of the sensitive areas is a crucial step in targeted observations (Majumdar, 2016; Zhang et al., 2017). The sensitive areas for targeted observation can be identified by the Conditional Nonlinear Optimal Perturbation (CNOP) approach proposed by Mu et al. (2003). Utilizing the CNOP approach, the optimal initial errors that cause the largest nonlinear forecast uncertainty can be calculated, and their spatial patterns help to locate the sensitive areas. To date, CNOP-identified sensitive areas have been proven to be quite effective in a number of oceanic applications, such as the prediction of the ENSO (Duan & Hu, 2016), upstream Kuroshio transport (Zhang et al., 2017), Kuroshio intrusion into the SCS (Liang et al., 2019), Kuroshio large meander (Wang et al., 2013), the ocean state in the SCS western boundary current region (Li et al., 2014). However, when focusing on a specific oceanic motion or event, there are many detailed issues to be addressed, e.g., the determination of the objective function and the constraint of the initial errors, the design of the optimal observation strategy and the determination of the effectiveness of the targeted observations in the field operation.

In the present study, we first extend the scope of oceanic targeted observations to summertime thermal structure predictions in a coastal sea, and put them into effect by conducting an oceanographic investigation in the field. We select temperature as the target variable because of its predominant impact on density fields and acoustic propagation (Dushaw et al., 2013). Under the comprehensive impact of the thermodynamic and dynamic oceanic processes and topography, the thermal structures in the coastal sea feature significant spatial and temporal variations, and their forecast uncertainty is generally large (Xia et al., 2006). Identifying the sensitive areas for the selected target region may enhance our understanding of the physical mechanism responsible for the thermal structure variation. In addition, exploring the utility of targeted observations in the thermal structure prediction may help improve the regional forecast system, with an optimally designed monitoring system deployed in the sensitive area. Here, we focus on improving the 7-day thermal structure prediction in the specified target region in the northwest Yellow Sea (YS; see locations in Figure 1). We found that assimilating the observations in the identified sensitive areas is more effective than locally assimilating approximately equal number of measurements inside the target region.

The article is organized as follows: The model configuration, the CNOP approach and the assimilation technique are briefly described in section 2. In section 3, given a specified target region, the sensitive areas for thermal structure prediction are identified. Then, the observation strategies are designed and quantitatively assessed by conducting a series of OSSEs. Section 4 introduces the observational data obtained from the ocean hydrographic survey and presents the improvements in the thermal structure prediction due to the targeted observation through Observing System Experiments (OSEs). The physical mechanism behind signal transport is also discussed. The results are summarized in section 5.

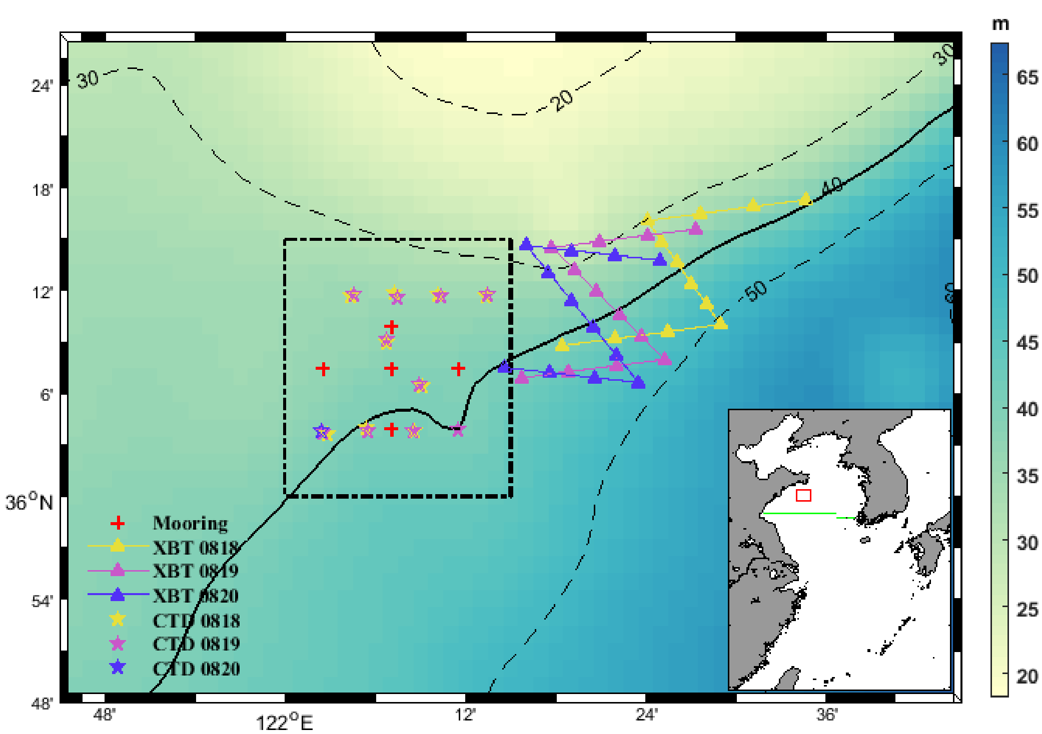


Figure 1. Plan view of the locations of the five temperature profile buoy stations (red crosses), thirty-six XBT stations (triangles), and twenty-one shipboard CTD stations (stars). The differences in the deployment times of the XBT and shipboard CTD observations are distinguished by different colors. The black dashed box indicates the location of the target region. The topography is indicated by shading. The bottom-right insert shows the model area, in which the red box indicates the position of the study area and the green lines indicate the section locations used for vertical thermal structure validation.

2. Methodology

2.1 Numerical model configuration

To investigate the utility of targeted observation in improving the prediction of thermal structures in the shallow YS, the Regional Ocean Modeling System (ROMS) solving the three-dimensional Reynolds-averaged hydrostatic Navier-Stokes equation with the Boussinesq approximation was used (Shchepetkin & McWilliams, 2005). The ROMS utilizes a nonlinear terrain-following vertical coordinate and has been proven to be suitable for regional ocean modeling by an increasing number of studies (Liang et al., 2019; Liu et al., 2019; Yang et al., 2011; Zhang et al., 2017). The K-profile parameterization scheme is used to calculate the vertical eddy viscosity and diffusivity (Large et al. 1994). Harmonic horizontal mixing is employed with constant horizontal eddy viscosity and diffusivity of 10 m^2s^{-1} and 15 m^2s^{-1}, respectively. The bottom stress is parameterized following a quadratic formula with a constant bottom drag coefficient set to 2.5×10^{-3}.

The model region covers the China Seas north of 23.7°N (Figure 1, 23.7-41.3°N, 117-132.5°E) with 1/24° horizontal resolution, and there are 32 vertical levels that are unevenly distributed, with closer spacing within the range of stratification. The model topography is subsampled from ETOPO2 (https://ngdc.noaa.gov/mgg/global/etopo2), and the minimum water depth is set to 10 m. The model initial temperature and salinity are obtained from the multiyear averaged (1998-2018) HYCOM+NCODA reanalysis data (https://www.hycom.org/dataserver) in January. The initial current velocities and sea surface height are set to zero.

First, a climatology run is carried out from a cold start. At the open boundaries, the model is driven by the multiyear averaged monthly HYCOM+NCODA reanalysis data and tidal forcing of eight major tidal constituents (M_2, S_2, K_1, O_1, N_2, K_2, P_1, and Q_1). The tidal forcing is included at the open boundaries by the Flather condition (Flather, 1976) with the tidal elevation and barotropic velocity obtained from the global inverse barotropic tidal model TPXO7.2 (Egbert & Erofeeva, 2002). On the surface, the wind stress, surface heat flux and water exchange are calculated from the multiyear averaged (1998-2018) monthly ECMWF reanalysis data (https://apps.ecmwf.int/datasets/). The climatology run is integrated for 25 years for spin-up.

Thereafter, a hindcast run is conducted from January 2014 to August 2019, starting from the results of the climatology run. Unlike the monthly mean external forcing data used for the climatology run, twelve-hourly surface forcing from the ECMWF reanalysis data and daily boundary forcing from the HYCOM+NCODA reanalysis data are applied to drive the hindcast run. The hindcast run is also forced by tidal forcing (eight major constituents) from TPXO7.2. In this paper, the daily-averaged temperature profiles are used for analysis.

2.2 CNOP approach

In this section, we briefly review the CNOP approach (Mu et al., 2003; 2009). Let M_t be the nonlinear propagator that propagates the value $\mathbf{X}_0$ at initial time t_0 to $\mathbf{X}_t = M_t(\mathbf{X}_0)$ at the end of the forecast time. When adding the initial perturbation $\Delta\mathbf{x}_0$ to the initial state, the nonlinear evolution of the initial perturbation $\Delta\mathbf{x}_t$ can be expressed as

$$\Delta\mathbf{x}_t = M_t(\mathbf{X}_0 + \Delta\mathbf{x}_0) - M_t(\mathbf{X}_0), \tag{1}$$

Following the definition proposed by Mu et al. (2003), the CNOP can be obtained by solving the following nonlinear constraint maximization problem:

$$J(\Delta\mathbf{x}_{0,\sigma}) = \max_{\|\Delta x_0\| \le \sigma} J(\Delta\mathbf{x}_0) = \max_{\|\Delta x_0\| \le \sigma} \left\| M_t(\mathbf{X}_0 + \Delta\mathbf{x}_0) - M_t(\mathbf{X}_0) \right\|, \tag{2}$$

with the constraint condition $\|\Delta\mathbf{x}_0\| \le \sigma$, where $J(\Delta\mathbf{x}_0)$ is the objective function that estimates the nonlinear evolution of the initial perturbation during time t. $\|.\|$ denotes the norm of the vector. $\Delta\mathbf{x}_{0,\sigma}$ is the CNOP-type initial error, which will induces the largest prediction error at the prediction time t.

Generally, CNOP computation relies on the adjoint technique to calculate the gradient of the objective function. However, directly calculating CNOP in a complicated model requires a considerable amount of coding and is computationally expensive (Liang et al., 2019; Zhang et al., 2017). Alternatively, in this study, we use an Empirical Orthogonal Function (EOF) based algorithm proposed by Wang & Tan (2009) to approximate the CNOP without using the adjoint technique (hereafter referred to as the EOF-CNOP method). Wang & Tan (2009) tested the EOF-CNOP method in a typhoon case. They found that the sensitive areas identified by this approximation algorithm are similar to the real CNOP results but require much less computational resources. The calculation process of the EOF-CNOP method is described as follows: First, a set of initial perturbations is added to the initial state to obtain the corresponding prediction increment ensemble by numerical integration. Then, the orthogonal basis of the initial perturbation ensemble is calculated by EOF decomposition. Finally, a statistical relationship is established between the initial perturbations and the associated prediction increment; thus, the gradient of the objective function can be obtained, and the CNOP can be computed.

In practice, the specific form of the objective function and the initial constraint are defined according to the object of study. In the context of the thermal structure of interest in this study, the objective function is defined as the change in the volume-integrated temperature caused by the initial errors in the specified target region, such that

$$J = \left(\int_A \Delta T_t dxdydz\right)^2, \tag{3}$$

where ΔT_t indicates the temperature anomaly at the future time t caused by the initial errors and A denotes the selected target region.

Following the formula of Li et al. (2014), the initial constraint is defined as

$$\left\|\Delta x_0\right\|^2 = \int_D \left(\frac{\Delta T_0}{T_{std}}\right)^2 dxdydz \le \sigma^2, \tag{4}$$

where ΔT_0 indicates the initial temperature perturbation, D denotes the model domain, and T_{std} indicates the regionally averaged temperature standard deviation in the simulated domain, which is calculated from the World Ocean Atlas 2018 (WOA18, https://www.nodc.noaa.gov/OC5/woa18/) in August and set to 0.25℃ in this study. The constraint radius σ is set to 2.5×10^3 to keep the state change in a reasonable range under the perturbation and to ensure the model stability. After completing all these steps, the sequential quadratic programming (Powell, 1983) algorithm is employed to compute the CNOP.

2.3 Optimal interpolation data assimilation

The Optimal Interpolation (OI) technique is utilized to assimilate the targeted observation data to reduce uncertainties in the initial fields, which can be formulated as

$$\begin{cases} x_a = x_b + K(y_{obs} - Hx_b) \\ K = BH^T(HBH^T + R)^{-1} \end{cases}, \tag{5}$$

where x_a and x_b indicate the analysis field and background field, respectively. y_{obs} denotes the observation vector, and H is the matrix of the model background field projections converted into the observational space. K is the weight matrix, which is calculated based on H, the model background field error covariation matrix B, and the observational error covariation matrix R. R is diagonal since all the observational errors are assumed uncorrelated in space. That is,

$$R_{ij} = \sigma_o^2 \sigma_{ij}, \tag{6}$$

where σ_0 is determined by the accuracies of the observations, σ_{ij} is the Kronecker delta, $\sigma_{ij} = 1$ when $i = j$, and $\sigma_{ij} = 0$ when $i \neq j$. The model background field error covariation matrix B at different vertical layers is assumed to be independent. Similar to the estimation used by Zhang (2019), B_{ij} is written as follows:

$$\begin{cases} B_{ij} = \sigma_b^2 \exp(-({d_{ij}}/{L_c})^2) & d_{ij} \le R_0 \\ 0 & d_{ij} > R_0 \end{cases}, \tag{7}$$

where σ_b is determined by the initial model errors, d_{ij} is the distance between two model grid points i and j. Referring to the temperature assimilation study in the YS by Ji et al. (2017), in this paper, the correlation length L_c and the influence radius R_o were set to 60 km and 120 km, respectively.

3. Identification of the sensitive area and observation strategy design

3.1 Vertical thermal structure validation

The simulated monthly averaged temperature along the section of 35°N (see location in Figure 1) in the hindcast years of 2016-2018 is compared with previous observations obtained from the Atlas of Ocean Data in the China Seas (Chen et al., 1992). In August, the water is mixed well in very shallow regions near the coast, and is stratified in the central basin. The simulated vertical distribution of isothermals is generally consistent with observations. Below the thermocline, the Yellow Sea Cold Water Mass (YSCWM) that formed during the previous winter is well reproduced. In the bottom, there are two cold cores inside the YSCWM, which agrees with a previous observational study (Zhang et al., 2008). The vertical thermal structure features interannual variability among the hindcast years, which is closely related to the interannual variability of the YSCWM and surface heating (Hu & Wang, 2004). In summary, the simulated vertical structure shows fairly good agreement with earlier observational and numerical studies. However, there is still a margin for improvement in the accuracy of the simulated thermal structure, especially below the surface mixing layer.

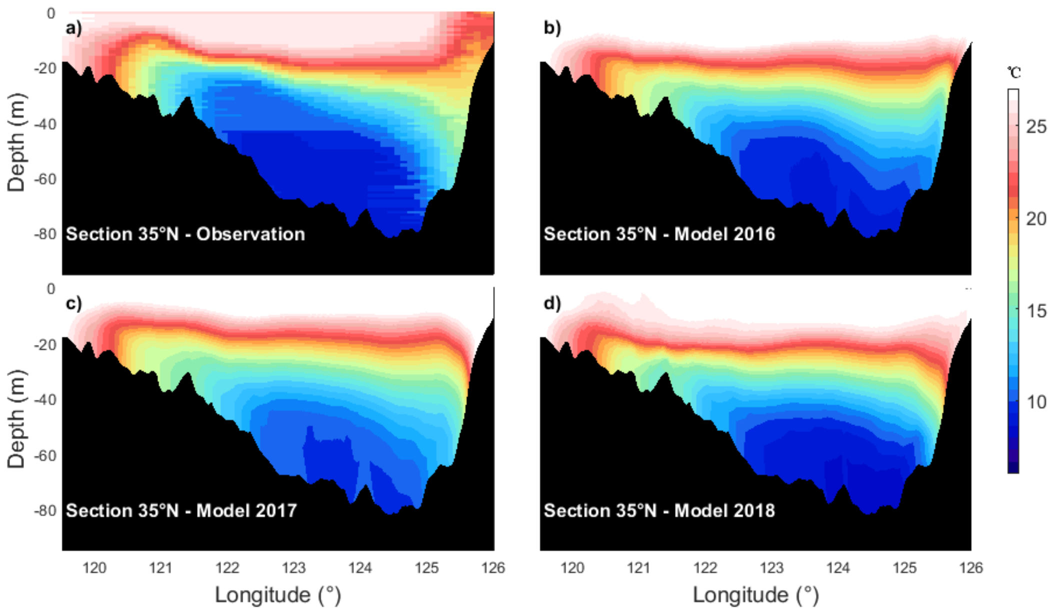


Figure 2. Comparison of the monthly mean (Aug) temperatures along section 35°N between the hindcast simulations (2016, 2017 and 2018) and the observations redrawn from the Atlas of Ocean Data in the China Seas.

3.2 CNOP-identified sensitive areas

To provide guidance for a targeted observation field campaign, a vital step is the identification of the sensitive areas. Several previous studies have utilized CNOP-identified sensitive areas to guide preferred oceanic targeted observations (Li et al., 2014; Wang et al., 2013; Zhang et al., 2017); however, oceanic sensitive area identification has not been studied in the context of thermal structure prediction. Given a selected target region located in the southeastern of the Shandong Peninsula (122-122.25°E, 36-36.25°N, the black dashed box in Figure 1), we aim to improve the short-term (7 days) summer thermal structure prediction by conducting targeted observations in the identified sensitive area. Considering the ships' voyage schedule, the initial prediction time is set to 00:00 on 20 August, and the daily averaged temperature profiles in the target region on 26 August are used for the forecast validation.

Note that the identification of the sensitive areas from the real-time predicted ocean state is not attempted, as this would entail the establishment of a reliable local prediction model with forcing from a larger-scale prediction model as a prerequisite. To provide guidance for the field campaign in August 2019, the sensitive areas in the last three hindcast years (2016-2018) are first identified.

Following Wang & Tan (2009), to identify the sensitive area, in every hindcast year, an ensemble of 20 initial perturbations and a natural run without perturbation is built. For this study of thermal structure prediction, initial perturbations are added to the temperature, which is achieved by taking the discrepancy of the daily averaged HYCOM+NCODA temperature data at the targeting day (20 August) between every two adjacent years during 1998-2018. All the initial temperature perturbations are scaled to the same magnitude of 0.25℃, which is estimated based on the temperature standard deviations within the simulation area from the WOA18 climatology data in August. Then, following the method of Zhang et al. (2017), the CNOP is calculated by employing a vertically integrated temperature scheme based on the 21 sets of initial ensemble conditions and the corresponding 7-day forecast samples. We confine the CNOP-identified sensitive area as the region where the CNOP-type errors have vertically integrated temperatures larger than a certain value τ. τ is determined to obtain a sensitive area which is the same size as the target region, which contains 56 horizontal model grids in this study.

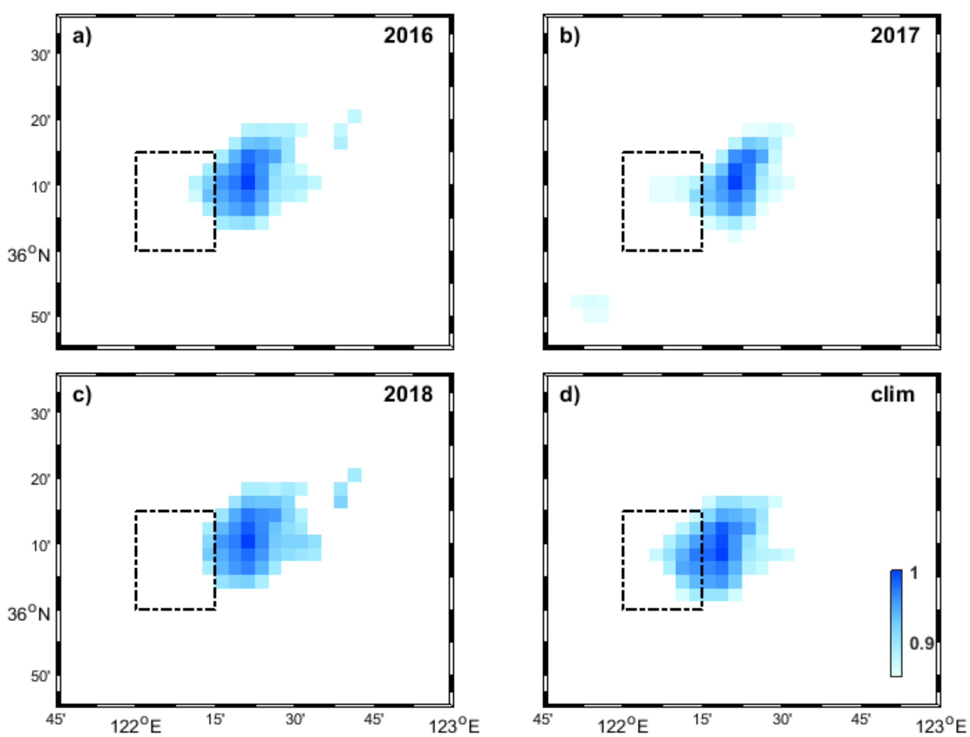


Figure 3. Locations of the identified sensitive areas for a) b) c) the hindcast years 2016-2018 and d) the last climatology year. The CNOPs are all normalized according to their maximum values.

The spatial distributions of the calculated CNOPs for the last three hindcast years of 2016-2018 are shown in Figures 3a-c. The absolute values of the CNOPs are different in every hindcast year, but only the relative values matter in the identification of the sensitive areas; thus, the CNOPs are all normalized according to their maximum values. We find that the sensitive areas are mainly located outside of the target region in the northeast, with only a small fraction of the area overlapping. In every hindcast year, the locations of the maximum values are generally consistent. The discrepancies among the identified sensitive areas are mainly concentrated in the marginal areas.

To validate the sensitivity in more detail, we systematically perturb the temperature fields at the initial time in three different areas (the sensitive area, the target region, and the area northeastern of the target region) and investigate the model responses in the temperature structure simulation (quantified by the regionally averaged temperature profile RMSEs in the target region). The northeast area is regarded as a nonsensitive area outside of the target region, and it is of the same size as the target region for a reasonable comparison. Random temperature perturbations with a normal distribution $N(0, 0.25)$ are added to all three of the above selected regions. The temporal evolution of the temperature prediction errors at a depth of 20 m in 2016 is shown in Figure 4. The development of the temperature perturbations is similar among the three hindcast years (not shown), they move westward and southwestward along with model integration. Adding perturbations in the target region causes the largest RMSEs at the initial time (Figure 4m). When perturbations are added to the sensitive area, the initial RMSEs are small but not zero due to the overlap with the target region. After 7 days of simulation, the RMSEs in the target region become the largest (Figure 5). These results suggest that at the prediction time, the local thermal structures in the target region are mostly affected by the initial perturbations in the sensitive area. Thus, the current method is proven to be effective in identifying the sensitive areas for the vertical thermal structures.

On the basis that the locations of the identified sensitive areas are generally consistent in space in every hindcast year, we try to obtain a multiyear averaged sensitive area to guide the field campaign. Following the same procedures, the sensitive area in the last climatology year is identified. The results show that the location of the identified sensitive area in the last climatology run agrees with that in the hindcast runs (Figure 3d). Perturbation experiments are also conducted in the last climatology run and confirm the effectiveness of the identified sensitive area (Figure 5). Thus, the CNOP-identified sensitive area from the last climatology run is used to guide the observation strategy designment (Figure 3d).

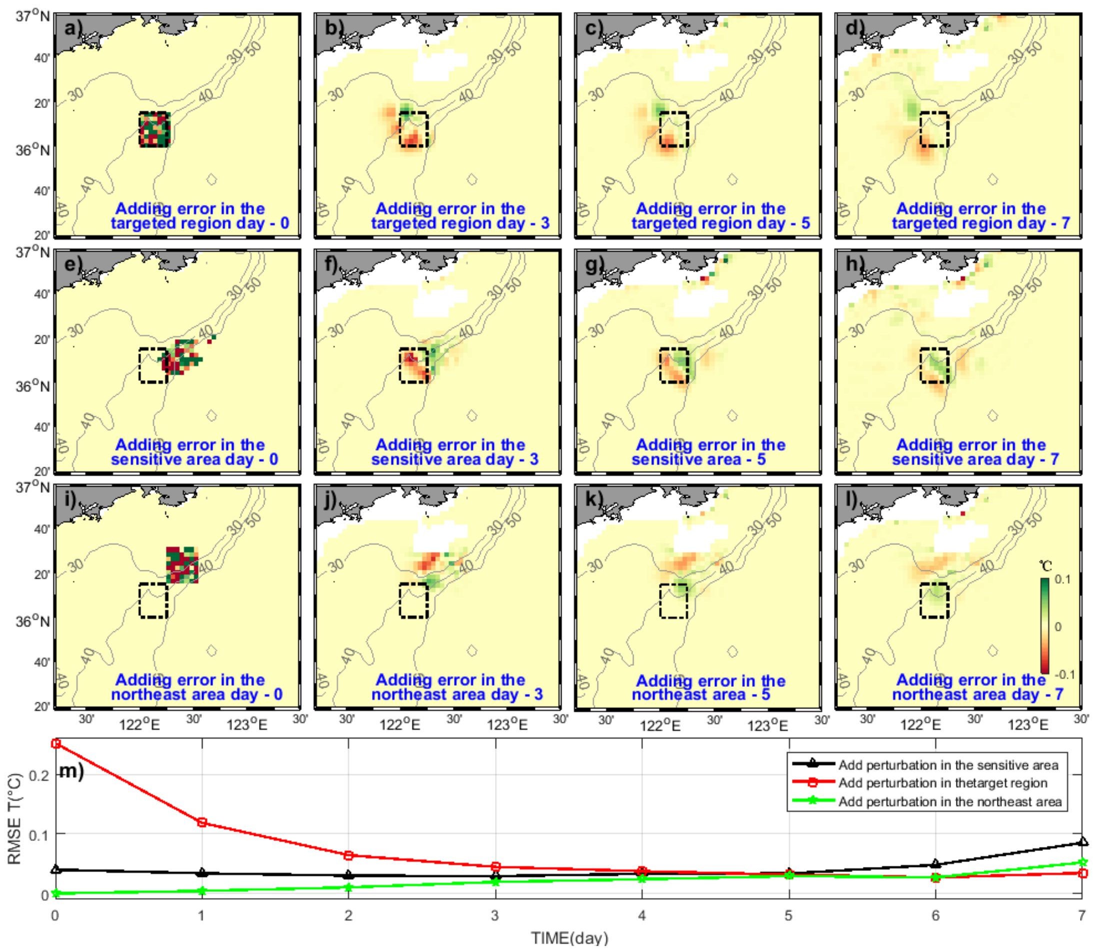


Figure 4. Temporal evolution of the temperature prediction errors at a depth of 20 m during the prediction time in 2016, with initial perturbations added to a-d) the target region, e-h) the sensitive area and i-l) the northeast area, respectively. Daily averaged results of the initial time, the third day, the fifth day and the seventh day are shown. m) Temporal evolution of the temperature RMSEs averaged in the target region.

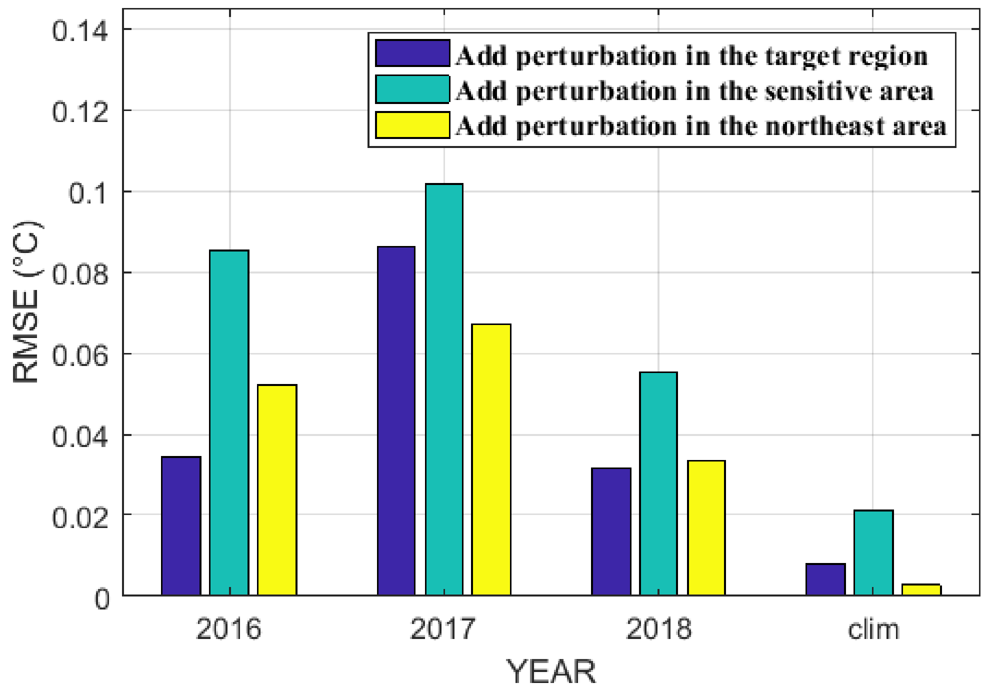


Figure 5. Temperature profile RMSEs in the target region after 7 days of simulation for the hindcast runs and the last climatology run with random perturbations added to different areas.

3.3 Design of observation strategy and benefit assessment with Observing System Simulation Experiments

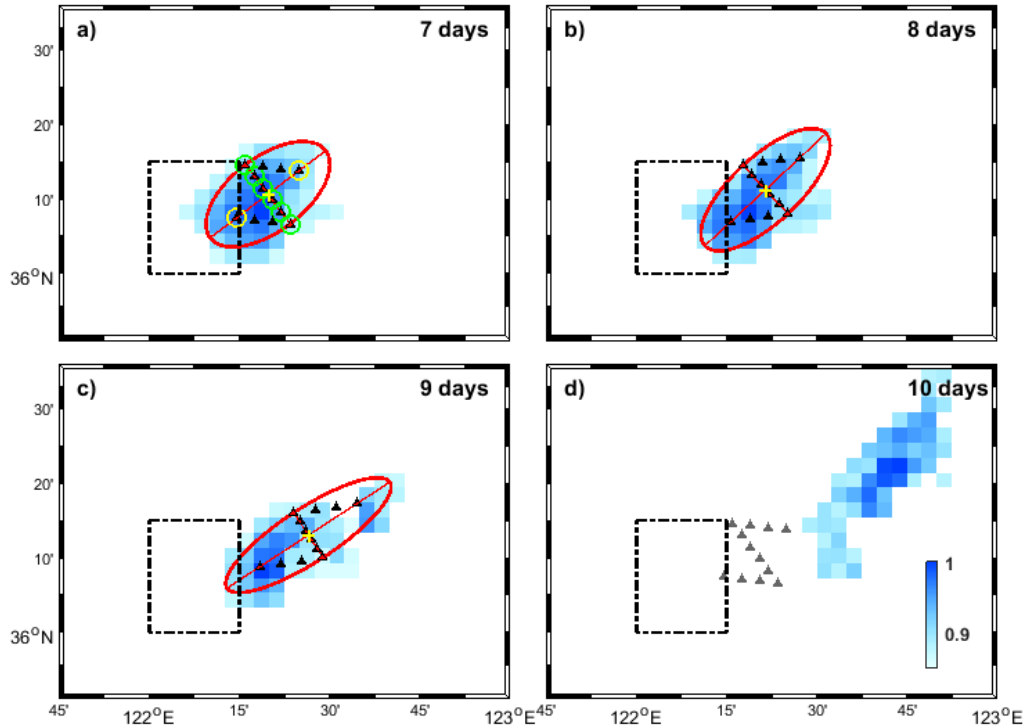
Before actually starting the field campaign, a targeted observation strategy that includes the ship route and the deployment locations should be designed. Moreover, the data assimilation technique (we use OI data assimilation here, section 2.3) should be utilized to maximize the benefit of the limited observation resources.

Three preconditions are assumed before determining the observation stations. First, a “Z” shape route is chosen to maximize the observation coverage in the identified sensitive area after conducting several numerical experiments (not shown here). Second, the daily averaged temperature observations are used for data assimilation to better represent the general vertical thermal structure. To obtain the daily average temperature profiles at each station, the ship route is designed to repeat four times a day (04:30-07:30, 10:30-13:30, 16:30-19:30, and 22:30-01:30).

Considering the observation simultaneity, the ship route length L is limited by the ship's speed (set to 8 knots) and the sailing time to complete each path (set to 3 hours). Third, although the prediction errors are expected to decrease for a higher number of observations, 12 stations are set along each route (approximately 4 km between the adjacent two stations) considering both the horizontal resolution (approximately 5 km) of our model and the observation cost.

Based on these preconditions, the specific ship route and the corresponding deployment locations along it are designed as follows: First, the spatial central point of the route is determined by averaging all the model grid coordinates in the sensitive area (the yellow cross in Figure 6a). Then, an ellipse is fitted with the central point, a major axis A_{long}, a minor axis A_{short} and a dip angle, and is scaled by a certain ratio to represent most of the sensitive area (the red ellipse in Figure 6a). Next, six equally spaced stations are set along the minor axis (green circles in Figure 6a). We assume that the shape "Z" is symmetric and that both ends of "Z" are located on the major axis of the ellipse. Given L and A_{short}, the leading and trailing observation stations on the major axis can be confirmed based on the Pythagorean theorem (yellow circles in Figure 6a). Finally, four equally spaced stations are added along the other two sides of the "Z" based on the above determined stations (Figure 6a). Except for the westernmost station, all the designed observation stations are out of the range of the target region.

It is worth noting that, the settings mentioned above represent a somewhat subjective strategy based on several assumptions and may not be the best solution. Observation optimization strategies for guiding targeting observations are urgently needed but are beyond the scope of this paper and will be investigated in future studies.



355

356 **Figure 6.** a) b) c) Z- shaped observation stations (black triangles) designed based on the
357 time-varying sensitive area (background colors). The black dashed box indicates the target
358 region. The station locations in d) (gray triangles) are the same as those in a), which are
359 completely out of the range of the 10-day sensitive area.

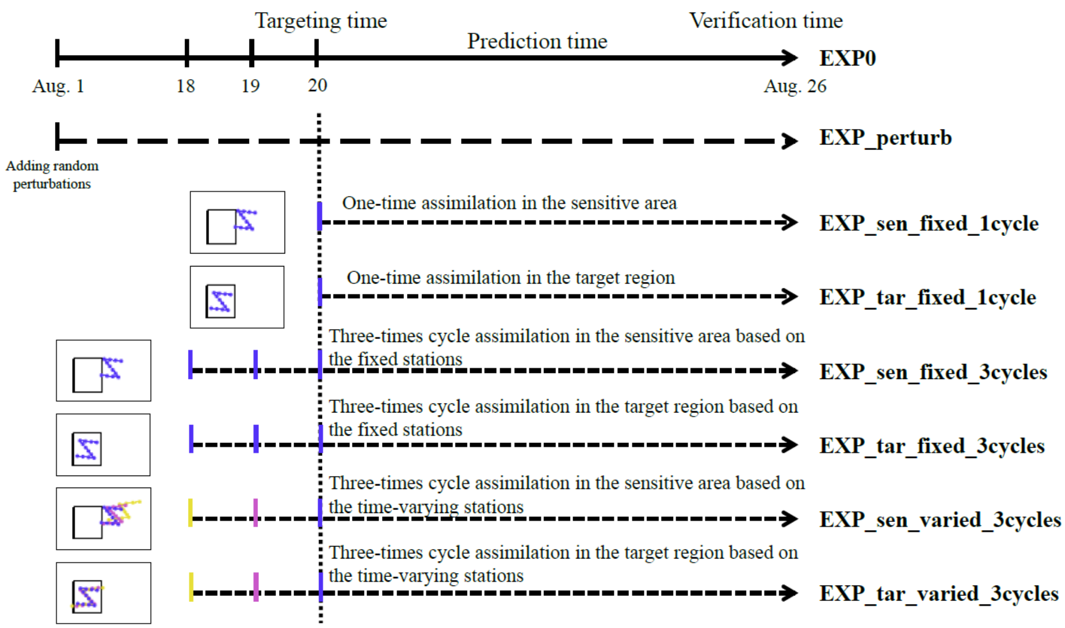


Figure 7. Schematic diagram of the Observing System Simulation Experiments based on the last climatology run. All the assimilation experiments use the results of the natural run as synthetic observations. The assimilated data station locations and the corresponding assimilation times are plotted by the same colors.

Table 1. Design of OSSEs based on the climatology run

Experiments	Data assimilation	Assimilation cycle	Comment
EXP0	no	-	Nature run
EXP_perturb	no	-	Adding perturbation at the targeting time
EXP_sen_fixed_1cycle	yes	1	One-time assimilation in the sensitive area at the stations shown in Figure6a
EXP_tar_fixed_1cycle	yes	1	One-time assimilation at the parallel stations in the target region (Figure 7)

EXP_sen_fixed_3cycles	yes	3	Three-times cycle assimilation in the sensitive area at the fixed stations (Figures 6a and 7)
EXP_tar_fixed_3cycles	yes	3	Three-times cycle assimilation at the parallel fixed stations in the target region (Figure 7)
EXP_sen_varied_3cycles	yes	3	Three-times cycle assimilation in the sensitive area at the time-varying stations (Figures 6a-c, Figure 7)
EXP_tar_varied_3cycles	yes	3	Three-times cycle assimilation at the parallel time-varying stations in the target region (Figures 7 and 10)

To evaluate the performance of the designed observation stations and the assimilation system, we implement a series of OSSEs based on the simulated results of the last climatology year (Table 1 and Figure 7). The original ocean state is denoted by the natural run EXP0, which is considered as the synthetic observation. Then, a control experiment (EXP_perturb) is created by superimposing temperature perturbations to EXP0 at 00:00, 1 August. The perturbation field is chosen among the 20 initial ensemble perturbations created for sensitive area identification, which induces the largest errors after 7 days of simulation. The perturbation magnitude is scaled to 0.35℃, which is larger than the perturbation magnitude of 0.25℃ in sensitive area identification, considering the error attenuation from the beginning (1 August) to the targeting time (20 August). In addition to the natural run and the control run, two assimilation experiments (EXP_sen_fixed_1cycle and EXP_tar_fixed_1cycle) are conducted through the assimilation of the synthetic observations at the targeting time. Stations for EXP_sen_fixed_1cycle are located in the sensitive area along the designed “Z” shape route (Figures 6a and 7). Stations for EXP_tar_fixed_1cycle are located in the target region; these stations are parallel to the stations of EXP_sen_fixed_1cycle, but the center of their route is located in the center of the target region (Figure 7). The regionally averaged temperature profile RMSEs in the target region at the verification time (26 August) between the natural run EXP0 and other experiments are used to evaluate the benefit of the observations for thermal structure prediction.

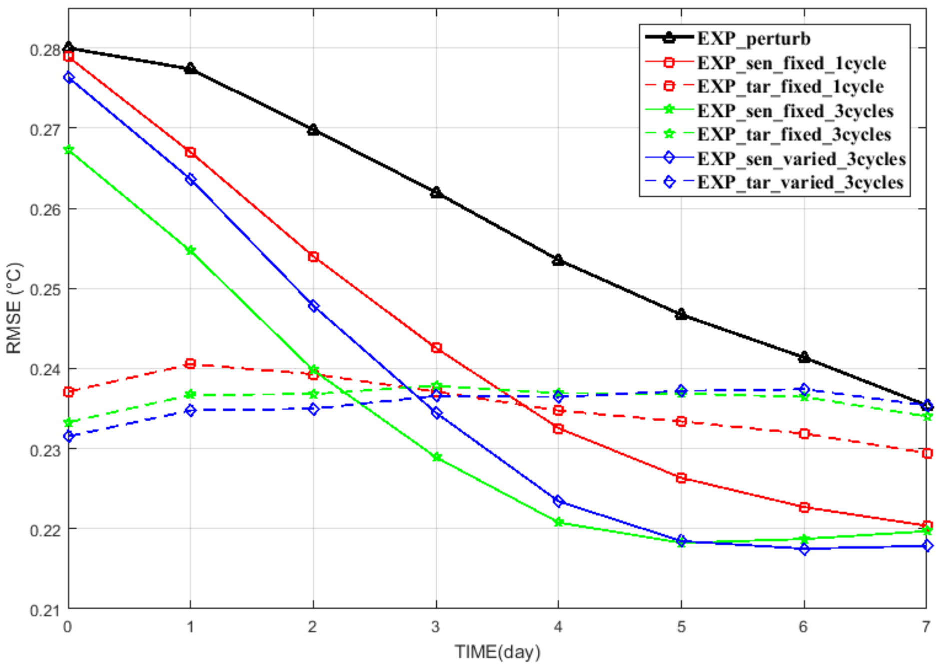


Figure 8. Temporal evolution of the regionally-averaged temperature profile RMSEs in the target region during the prediction time in the Observing System Simulation Experiments.

The temporal evolutions of the temperature profile RMSEs in the OSSEs are shown in Figure 8. At the targeting time, the regionally averaged RMSEs in the target region are approximately 0.28°C for the control experiment (EXP_perturb) and attenuate to approximately 0.235°C at the prediction time (the black line in Figure 8). Although the absolute magnitudes of the RMSEs in the OSSEs are small because of the small initial perturbations, the relative magnitudes of the RMSEs and their temporal evolution can still reflect the effectiveness of the different observation strategies. In the EXP_tar_fixed_1cycle, which represents the conventional observation strategy, the initial RMSEs are greatly reduced after data assimilation in the target region (the red dashed line in Figure 8). After 7 days of integration, the effectiveness of the forecast refinement decreases. In the EXP_sen_fixed_1cycle, the initial RMSEs are only slightly reduced at the targeting time (the red solid line in Figure 8) because only one of the total 12 stations is located inside the target region. However, at the verification time, the forecast errors are smaller than the results of both EXP_perturb and EXP_tar_fixed_1cycle. These results support the effectiveness of our observation strategy and data assimilation system.

To further reduce the forecast error, we explore the possibilities for improving the initial state by utilizing the cycle data assimilation technique. In the EXP_sen_fixed_3cycles and EXP_tar_fixed_3cycles, the data are cycle assimilated three times (00:00, 18, 19, 20 August) at the fixed stations in the sensitive area and the target region, respectively (Figure 6a and Figure 7). Compared to the one-time data assimilation (red solid and dashed lines in Figure 8), at the targeting time, the initial RMSEs are both reduced by the corresponding three cycles of data assimilation (green solid and dashed lines in Figure 8). After 7 days of integration, the forecast errors in the EXP_sen_fixed_3cycles are minimal compared to those of both EXP_tar_fixed_3cycles and EXP_sen_fixed_1cycle. This gives us confidence in the ability of cycle data assimilation to reduce the forecast uncertainty in the identified sensitive area.

We realize that the locations of sensitive areas on 18 and 19 August (9 days and 8 days before the verification time) may be different from that on 20 August (7 days before the verification time). Thus, following the same procedure described in section 3.2, the 8-day, 9-day and 10-day sensitive areas are identified, and these areas are shown in Figures 6b-d. The centralis of the identified sensitive area moves northeastward and becomes oblate with increasing prediction time. The previously designed stations based on the 7-day sensitive area are all out of the range of the 10-day sensitive area (Figure 6d).

Then, new deployment locations based on the identified 8-day and 9-day sensitive areas are designed following the same rule (Figures 6b and c). All the stations based on the 8-day and 9-day sensitive areas are outside of the target region. The impact of the time-varying observation stations is evaluated by conducting two extra experiments, EXP_sen_varied_3cycles and EXP_tar_varied_3cycles. In EXP_sen_varied_3cycles, data are cycle assimilated three times (00:00, 18, 19, 20 August) at the stations in the 7-day, 8-day and 9-day sensitive areas (see station locations in Figures 1 and 7). In EXP_tar_varied_3cycles, the stations of EXP_sen_varied_3cycles are moved parallel to the center of the target region (Figures 7 and 10d). At the targeting time, the RMSE of EXP_sen_varied_3cycles (the blue solid line in Figure 8) is larger than that of EXP_sen_fixed_3cycles because the designed stations based on the 8-day and 9-day sensitive areas are farther away from the target region than those based on the 7-day sensitive area. While the initial RMSE of EXP_tar_varied_3cycles (the blue dashed line in

Figure 8) is less than that of EXP_tar_fixed_3cycles because the designed stations based on the 8-day and 9-day sensitive areas have broader spatial coverage than those based on the 7-day sensitive area. After 7 days of integration, EXP_sen_varied_3cycles performs the best among all the OSSEs in reducing the forecast error.

A two-cycle data assimilation experiment in the sensitive area is also conducted, and the forecast improvement falls between those of EXP_sen_fixed_1cycle and EXP_sen_varied_3cycles (not shown). Four-cycle data assimilation experiments are not implemented considering the actual observation cost in the field campaign. One may argue that, why not triple the observation stations from 12 to 36 in the one-time data assimilation in the sensitive area? In fact, as mentioned above, limited by the ship route length and the horizontal model resolution, a denser observation will not significantly expand the spatial observation coverage.

To further assess the effectiveness of the observation strategy in the subsequent field operation, we conduct additional OSSEs based on the simulated results of the hindcast years 2016-2018 (Table 2). In every hindcast year, the hindcast control experiments are first created following the same procedures as those in EXP_perturb. Then, similar to EXP_sen_varied_3cycles and EXP_tar_varied_3cycles, the benefit of the targeted observation is tested through the assimilation of the synthetic observations at the time-varying stations in the sensitive areas and the target region, respectively. After 7 days of integration, in every hindcast year, assimilating data in the sensitive areas based on the above determined observation strategy can yield more profit than the conventional local data assimilation (Table 3). The results mentioned above support the implementation of the targeted observation campaign in the summer 2019 in the YS.

Table 2. Observing System Simulation Experiments based on the hindcast runs

Experiments	Comment
EXP2016	Nature run
EXP2016_perturb	Control run
EXP2016_tar_varied_3cycles	Three-times cycle assimilation in the target region at the time-varying stations

EXP2016_sen_varied_3cycles	Three-times cycle assimilation in the sensitive area at the time-varying stations
EXP2017	Nature run
EXP2017_perturb	Control run
EXP2017_tar_varied_3cycles	Three-times cycle assimilation in the target region at the time-varying stations
EXP2017_sen_varied_3cycles	Three-times cycle assimilation in the sensitive area at the time-varying stations
EXP2018	Nature run
EXP2018_perturb	Control run
EXP2018_tar_varied_3cycles	Three-times cycle assimilation in the target region at the time-varying stations
EXP2018_sen_varied_3cycles	Three-times cycle assimilation in the sensitive area at the time-varying stations

Table 3. Assessment of the designed observing strategy in the hindcast years of 2016-2018 (RMSEs refinement in percentage)

Year / Experiments	2016	2017	2018
EXP_tar_varied_3cycles	-32.0%	20.3%	59.7%
EXP_sen_varied_3cycles	43.9%	48.2%	70.1%

4. Forecast improvements and discussion

4.1 DATA

A dedicated ocean survey with two synergetic ships is carried out in August 2019 to obtain the targeted observation data in the YS. In the target region, five buoys (red crosses in Figure 10c) are placed from 17 to 27 August for forecast validation and OSEs. The buoys are composed of temperature loggers (SBE56), pressure-temperature loggers (SBE39 and RBRduo[3]) and pressure-temperature-conductivity loggers (RBRconcerto[3]), which can obtain the temperature profiles of nearly the total water volume in approximately 2 m vertical bins. Both ends of the buoys are equipped with pressure sensor instruments to determine the depths of the temperature loggers between them. The sensors collected a sample every 10 mins. During 18-20 August, 21 temperature profiles are obtained by shipboard CTD (stars in Figure 10) to measure the influence of local data assimilation in the target region on forecasts.

In the sensitive areas, temperature profiles are collected by the eXpendable BathyThermographs (XBT) during 18-20 August. Temperature profiles at each XBT station are detected four times a day (16:30-19:30, 22:30-1:30, 4:30-7:30, 10:30-13:30) along the predesigned routes to obtain the daily averaged value, which are used in the cycle data assimilation at 00:00 on 18, 19, and 20 August 2019. Given that the repeated cruises undergo inevitable spatial uncertainty, after performing data quality control, the daily averaged temperature profiles used for data assimilation are obtained by interpolating both the XBT data and the simultaneous buoy data at the standard station locations. All times in the study are referenced to the Chinese Standard Time (UTC+8).

4.2 Forecast improvements

In this section, the performance of the targeted observations in improving the forecast is validated. The daily-averaged temperature profiles of the model results and the observation data at the five buoy stations in 26 August, 2019 are compared (Figure 9, see buoy locations in Figure 10c). The simulated sea surface temperature agrees very well with the observations, but the predicted upper mixed layer thicknesses are slightly thinner. At the bottom, the simulated temperature is higher than the observation, which may be caused by the insufficient cooling in the previous winter. Compared to the upper and bottom layers, the accuracy in the middle water volume is lower as a result of strong seasonal variations in the shallow sea thermocline, and the predicted temperature was relatively lower. At the buoy stations, the RMSEs between the modeled temperature profiles and the observations are approximately 1.66-2.74 ℃ (an average value of 2.11 ℃), and these errors increase to 2.26-3.75 ℃ (an average value of 2.84 ℃) when the depth ranges are restricted to the thermocline of 15-30 m. It should be noted that the temperature RMSEs are only calculated at depths where observations are available. Horizontally, the modeled temperature RMSEs at stations W1 and W4 are apparently higher than those at the other three stations. This indicates that it is more difficult to correctly reproduce the summer vertical thermal structures on the continental slopes with lesser water depths (Figure 10c). Overall, the simulation successfully captures the general vertical thermal structures in the target region, but there is still much room to improve the forecast accuracy, especially within the depth range of the summer thermocline.

Then, observations obtained in the sensitive area are assimilated to quantify the benefits of the targeted observations (EXP2019_sen). As illustrated in Figure 9, there is a marked improvement in the vertical thermal structure simulations after assimilation. Among the five buoy stations, the average RMSEs between the modeled temperature profiles and the observations are reduced from 2.11 to 1.7 ℃, with an average forecast improvement of approximately 18.9% (compared to that of the EXP2019). However, one may argue that, this significant forecast improvement could be attributed to the data assimilation technique rather than the targeted observations. To highlight the contribution of the targeted observations, we conduct a series of OSEs (Table 4) in the following section. The results show that the prediction benefits decrease if the equivalent measurements are deployed locally in the target region.

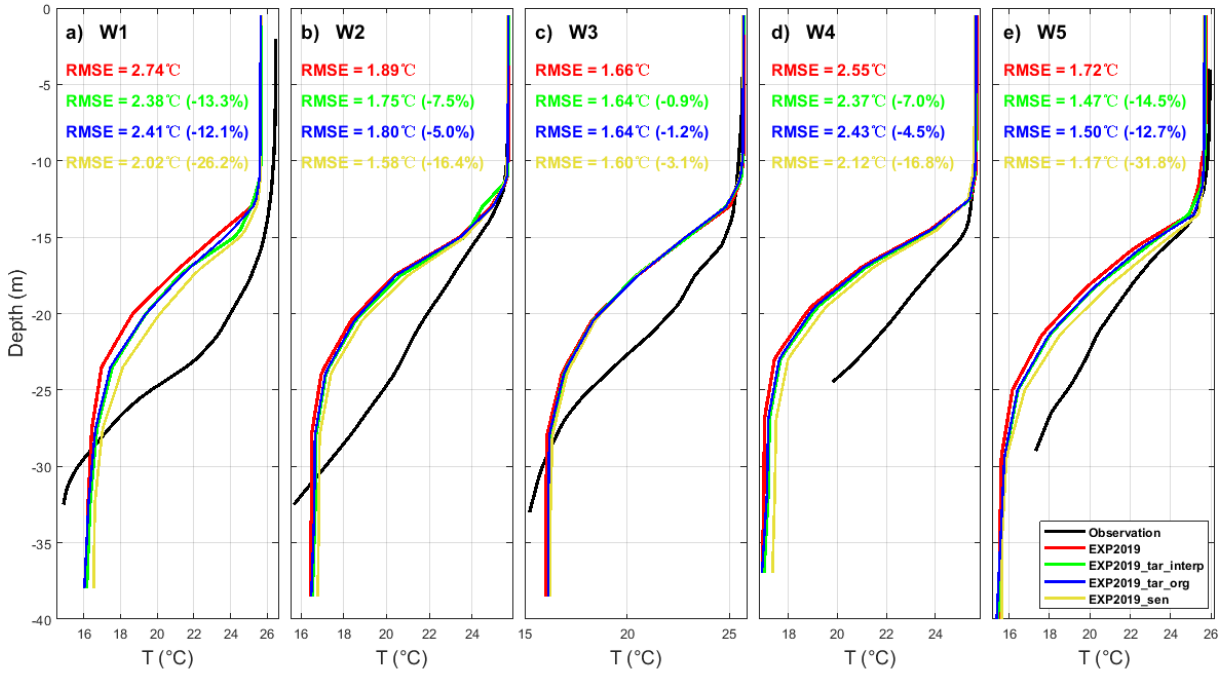


Figure 9. Comparison of the observed and simulated daily-averaged temperature profiles at five buoys (see locations in Figure 10c). The black lines indicate the in-situ observations. The red lines indicate the model results without assimilation. The blue lines and the yellow lines indicate the improvement in the prediction from the assimilation of the observations in the target region and sensitive area, respectively. The green lines indicate the model results of EXP2019_tar_interp for the assimilation of the interpolated data at the synthetic stations in the target region (see locations in Figure 10d).

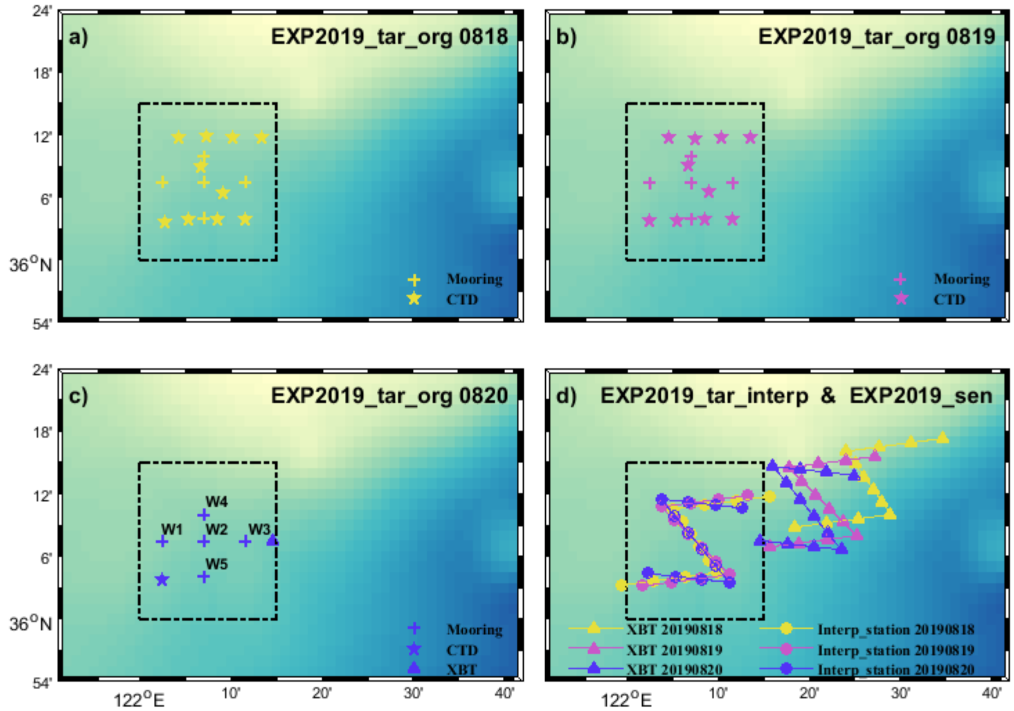


Figure 10. Station locations of the Observing System Experiments. The buoy stations, XBT stations, and shipboard CTD stations are denoted by the crosses, triangles and stars, respectively. The circles inside the target region in d) indicate the synthetic stations. The different deployment times of the observations are distinguished by different colors.

Table 4. Design of Observing System Experiments

Experiments	Data assimilation	Assimilation cycle	Comment
EXP2019	no	-	Control run
EXP2019_tar_org	yes	3	Cycle assimilate the original observations in the target region (15 profiles in 18 and 19 August, respectively, and 7 profiles in 20 August)
EXP2019_tar_interp	yes	3	Cycle assimilate the interpolated data at the synthetic stations in the target region (36 profiles)

EXP2019_sen	yes	3	Cycle assimilate the XBT data at the stations designed based on the identified 7-days, 8-days and 9-days sensitive areas, respectively (36 profiles)

4.3 Observing System Experiments

The benefit of oceanic targeted observations has been tested in some previous studies through a series of OSSEs (Li et al., 2014; Wang et al., 2013; Zhang et al., 2019). However, the effect of oceanic targeted observations guided by sensitive areas has never been tested in OSEs through the use of real data in actual operation. Generally, in the context of standard OSEs designed for atmospheric targeted observation, the experiment assimilating all the available observations is regarded as the control experiment, and the impact of the selected observations is assessed by removing subsets of the measurements or by adding extra measurements and comparing the results with the control experiment (Majumdar et al., 2011). In the oceanic region of this study, the historical observations that we can obtained are sparse. Thus, we set an experiment that does not assimilate any data as the control experiment (EXP2019). Given that the target region is the most representative nonsensitive area for the benefit assessment of OSEs, in addition to experiment EXP2019_sen, we conducted two extra experiments that assimilate approximately equal amounts of measurements inside the target region: EXP2019_tar_org, for which a total of 37 originally observed temperature profiles in the target region are assimilated (see locations in Figures 10a-c), and EXP2019_tar_interp, for which 36 interpolated data in a set of synthetic stations in the target region are assimilated (see locations in Figure 10d). These synthetic stations are parallel to the corresponding stations in the sensitive area, but their daily routes are located in the center of the target region. The daily averaged temperature profiles are obtained by interpolating all the observations available on that day to the synthetic stations. It should be noted that, to take full advantage of the limited observations, the shipboard CTD temperature profiles used in the OSEs are only one-time measurements instead of daily averaged values, which is a flaw of the designed OSEs.

Despite the difference in the spatial locations and numbers of the temperature profiles assimilated every day of the cycle assimilation, the forecast improvements in EXP2019_tar_org and EXP2019_tar_interp are nearly the same. In every buoy station, the simulated temperature profiles in EXP2019_tar_org and EXP2019_tar_interp are refined due to data assimilation in the target region. However, the forecast improvements in EXP2019_tar_org and EXP2019_tar_interp are both less than half of that in EXP2019_sen (average RMSE decreases of 7.1% and 8.6% vs. 18.9%). The results of the OSEs support our initial assumption that conducting data assimilation in the CNOP-identified sensitive area is more effective in forecast improvement than in other areas including the target region itself. However, it should be noted that, the quantitative benefit of targeted observation in the CNOP-identified sensitive area differs from model to model and depends on the initial simulation accuracy and the selected data assimilation scheme.

4.4 Dynamic analysis

To better understand how the local forecast errors are efficiently reduced by conducting targeted observations in the remote sensitive area, it is worth exploring the dynamics behind. We quantitively investigate the physical processes affecting the water temperature in the target region using the model temperature equation

$$\frac{\partial T}{\partial t} = -\nabla \cdot (\vec{v}T) + \nabla_h (A_h \nabla_h T) + \frac{\partial}{\partial z}(A_v \frac{\partial T}{\partial z}), \qquad (8)$$

where T is temperature, $\vec{v}$ is velocity, and A_h and A_v are the horizontal and vertical diffusivity coefficients, respectively. The temperature change in the water is mainly induced by horizontal temperature advection, vertical temperature advection, horizontal temperature diffusion and vertical temperature diffusion. The ocean temperature is also affected by the change in surface heating. However, considering that in this study, we only conducted targeted observations inside the water volume, thus, only the impact of advection and diffusion processes are discussed.

The temporal evolution of the vertically integrated and regionally averaged temperature biases induced by different processes in the target region is shown in Figure 11. Since we are focused on the evolution of different processes during the prediction time rather than the initial refinement of the ocean state, the initial biases are set to zero at the targeting time. The total biases of EXP_perturb and EXP_sen_varied_3cycles against EXP0 are always negative, suggesting a decrease in the temperature discrepancy against the natural run, which is consistent with the previous results of OSSEs (the black and blue descending trend lines in Figure 8). However, the total biases of EXP_tar_varied_3cycles against EXP0 are positive, which suggests that the temperature field in EXP_tar_varied_3cycles becomes worse since the targeting time (the blue dashed line in Figure 8). Due to the effectiveness of the targeted observation, the amplitude of the bias for EXP_sen_varied_3cycles against EXP0 is larger than that for EXP_perturb against EXP0. Thus far, we can conclude that conventional local data assimilation can greatly improve the initial temperature field in the target region, but the effectiveness decreases during model integration. In contrast, the temperature field in the target region may not be significantly refined through the assimilation of the data in the remote sensitive area at the targeting moment, but it will be continuously improved during the prediction time, and reach a more precise state at the verification time.

It is clear that the horizontal advection accounts for the majority of the temperature biases during the prediction time (Figure 11). The temperature biases of vertical advection, horizontal diffusion and vertical diffusion are always positive, indicating a negative effect of these three processes. In Figures 11a and b, the temperature biases of horizontal advection have the largest negative amplitude, suggesting that the horizontal temperature advection contributes the most to improving the prediction. In EXP_tar_varied_3cycles, both the advection and diffusion processes lead to a reduction in the simulation accuracy, within which the horizontal advection contributes the most.

As the baroclinic response of the YSCWM, there exists a cyclonic gyre of approximately 0.2 Sv in the summer YS (Naimie et al., 2001). The identified sensitive area is located northeastward of the target region, which is consistent with the local flow direction of the YSCWM circulation (southwestward). From historical studies, although the YS summer circulation was supposed to feature a complex two-layer or three-layer structure (Xu et al., 2002; Xia et al., 2006), it is widely accepted that most of the middle water volume (4-40 m) is dominated by a basin-scale cyclonic circulation. This interprets our results that the refinement of the vertical thermal structures by targeted observation mainly occurs in the middle water volume rather than in the surface or bottom mixed layer (Figure 9). Through the assimilation of the targeted measurements in the sensitive area, the information is subsequently advectively carried downstream to the target region by the YSCWM circulation. The distance of the identified sensitive area from the target region is associated with the involved prediction time.

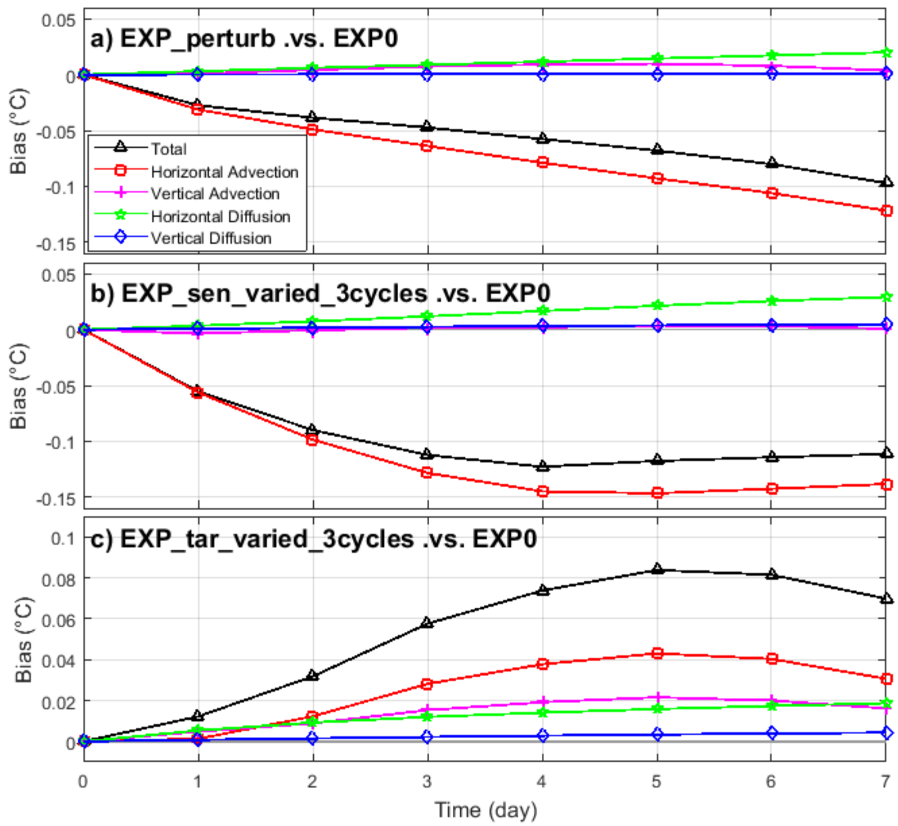


Figure 11. Temporal evolution of the vertically-integrated regionally-averaged temperature biases induced by different processes in the target region for a) EXP_perturb vs. EXP0, b) EXP_sen_varied_3cycles vs. EXP0, and c) EXP_tar_varied_3cycles vs. EXP0 during the prediction time.

5. Conclusion

Targeted observation is believed to be a cost-effective way to decrease forecast uncertainty through the assimilation of additional measurements in the initial state. This study first extends the scope of oceanic targeted observations to the vertical thermal structure predictions. Given a selected target region and a fixed prediction period of seven days, the sensitive areas are identified utilizing the CNOP method and a newly defined objective function. The majority of the sensitive areas are located outside of the target region in the northeast. Through the superimposition of random errors in several selected regions, the initial state of the sensitive area is proven to have the most impact on the thermal structure prediction in the target region. Given that the locations of the identified sensitive areas in the hindcast and climatology runs are generally consistent, guided by the CNOP-identified sensitive area of the last climatology year, we design the observation strategy with the technique of cycle data assimilation and the new concept of the time-varying sensitive area. A series of OSSEs are conducted to assess the observation performance before the field campaign. The results show that, cycle assimilating temperature profiles at the designed stations in the 7-day, 8-day and 9-day sensitive areas can yield the maximum benefits.

A choreographed field campaign is then applied in the summer of 2019 in the YS to evaluate the capabilities of targeted observations to reduce the temperature uncertainty in numerical predictions. Our field experiment applied XBTs to purposefully sample the thermal profiles in the sensitive areas. Inside the target region, an approximately equal number of temperature profiles were gathered by shipboard CTDs and buoys. OSEs were conducted to test the capabilities of targeted observations. The results show that reducing the initial errors in the sensitive area can lead to improvement in the thermal structure prediction (18.9%) greater than that in the target region (7-9%). Compared to assimilating local observations in the target region, assimilating observations in the identified sensitive areas can double the benefit of data assimilation regarding forecast improvements. To explore this further, we investigated the physical dynamics behind. A term-by-term analysis of the model temperature equation indicates that the horizontal temperature advection contributes the most to forecast improvement during the prediction time. After conducting targeted observation in the upstream sensitive area, the physical signals are subsequently carried downstream to the target region by the horizontal temperature advection of the YSCWM circulation.

In this study, we skip the step of establishing a real-time prediction model, on the basis that the locations of the identified sensitive areas in the hindcast and climatology runs are generally consistent. Although this kind of spatial consistency was also found in the optimal precursor study of the Kuroshio intrusion into the SCS (Liang et al., 2019; personal communication), it will not always be applicable if the focused phenomenon or study area changes. Thus, future work should be guided based on a reliable local prediction system. Furthermore, the optimal deployment network should be investigated and extended to the three-dimensional scenarios. A more advanced data assimilation technique is also needed to better exploit the targeted data.

Acknowledgments

This paper benefited from the insightful and constructive comments of the journal editor and the anonymous reviewers. We would like to thank the use of super computer for all the numerical simulations at the Navy Submarine Academy Underwater Marine Environment Institute. We are also gratefully acknowledge to the captains Bing Liu and Xiufeng Chen, the laboratory director Lei Chen, and all the crews of the R/V KeXue No.3 and R/V ChuangXin No.2 for their help in collecting the data. The ETOPO2 topography is available at the website (https://ngdc.noaa.gov/mgg/global/etopo2). The HYCOM+NCODA reanalysis data is available at the website (https://www.hycom.org/dataserver). The ECMWF atmosphere forcing data is available at the website (https://apps.ecmwf.int/datasets/). The TPXO7.2 tidal forcing can be obtained from the website (https://www.tpxo.net/). The observation data can be downloaded from the website (https://figshare.com/s/f486e3452e731f6a8a70). This work was supported by funds from the National Natural Science Foundation of China (Grants No. 41906005 & No. 41705081), National Basic Research Program of China (Grant No. 2019-JCJQ-ZD-149-00), Innovation Special Zone Project (Grant No. 18-H863-05-ZT-001-012-06), and the Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology (Grant NO. 2019A05). K. Liu and G. Wu contributed equally to this work.

References

Alvarez, A., & Mourre, B. (2012). Optimum sampling designs for a glider-mooring observing network. *J. Atmos. Oceanic Technol.*, 29(4), 601-612, doi.org/ 10.1175/JTECH-D-11-00105.1

Baehr, J., Mcinerney, D., Keller, K., & Marotzke, J. (2008). Optimization of an observing system design for the North Atlantic Meridional Overturning Circulation. *J. Atmos. Oceanic Technol.*, 25(4), 625-634, doi.org/ 10.1175/2007jtecho535.1

Chen, G., Niu, Y., Wen, S., Bao, C., Guan, D., Wu, B., et al. (1992). Marine atlas of Bohai Sea, Yellow Sea, East China Sea (Hydrology). China Ocean Press, Beijing, China.

Curtin, T. B., & Bellingham, J. G. (2009). Progress toward autonomous ocean sampling networks. *Deep Sea Res., Part II: Topical Studies in Oceanography*, 56(3-5), 0-67, doi.org/ 10.1016/j.dsr2.2008.09.005

Duan, W., & Hu, J. (2016). The initial errors that induce a significant “spring predictability barrier” for El Niño events and their implications for target observation: results from an earth system model. *Clim. Dynam.*, 46, 3599-3615, doi.org/10.1007/s00382-015-2789-5

Dushaw, B. D., Worcester, P. F., Dzieciuch, M. A., & Menemenlis, D. (2013). On the time-mean state of ocean models and the properties of long range acoustic propagation. *J. Geophys. Res.*, 118(9), 4346-4362, doi.org/ 10.1002/jgrc.20325

Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Oceanic Technol.*, 19, 183-204, doi.org/ 10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2

Farrara, J. D., Chao, Y., Li, Z., Wang, X., Jin, X., Zhang, H., et al. (2013). A data-assimilative ocean forecasting system for the Prince William sound and an evaluation of its performance during sound Predictions 2009, *Cont. Shelf Res.*, 63, S193–S208, doi.org/10.1016/j.csr.2012.11.008

Flather, R. A. (1976). A tidal model of the north-west European continental shelf. *Mem. Soc. Roy. Sci. Liege*, 6, 141-164.

Hu, D., & Wang, Q. (2004). Interannual variability of the southern Yellow Sea Cold Water Mass. *Chin. J. Oceanol. Limn.*, 22(3), 231-236, doi.org/10.1007/BF02842553

Ji, X., Kwon, K. M., Choi, B.-J., Liu, G., Park, K.-S., Wang, H., et al. (2017). Assimilating OSTIA SST into regional modeling systems for the Yellow Sea using ensemble methods. *Acta Oceanolo. Sin.*, 36(3), 37–51, doi: 10.1007/s13131-017-0978-2

Köhl, A., & Stammer, D. (2004). Optimal observations for variational data assimilation. *J. Phys. Oceanogr.*, 34(3), 34-42, doi.org/ 10.1175/2513.1

Kramer, W., Dijkstra, H. A., Pierini, S., & Leeuwen, P. J. V. (2012). Measuring the impact of observations on the predictability of the Kuroshio Extension in a shallow-water model. *J. Phys. Oceanogr.*, 42(1), 3-17, doi.org/10.1175/JPO-D-11-014.1

Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, 32, 363-403, doi.org/10.1029/94rg01872

Lermusiaux, P. F. J. (2007). Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica. D.*, 230(1-2), 172-196, doi.org/ 10.1016/j.physd.2007.02.014

Li, Y., Peng, S., & Liu, D. (2014). Adaptive observation in the South China Sea using CNOP approach based on a 3-D ocean circulation model and its adjoint model. *J. Geophys. Res.*, 119(12), 8973-8986, doi.org/10.1002/2014JC010220

Liang, P., Mu, M., Wang, Q., & Yang, L. (2019). Optimal precursors triggering the Kuroshio Intrusion into the South China Sea obtained by the Conditional Nonlinear Optimal Perturbation approach. *J. Geophys. Res.*, 124, 3941-3962, doi.org/10.1029/2018JC014545

Liu, K., Sun, J., Guo, C., Yang, Y., Yu, W., & Wei, Z. (2019). Seasonal and spatial variations of theM2 internal tide in the Yellow Sea. *J. Geophys. Res.*, 124, doi.org/10.1029/2018JC014819

Majumdar, S. J., Aberson, S. D., Bishop, C. H., Cardinali, C., Caughey, J., Doerenbecher, A., et al. (2011). Targeted observations for improving numerical weather prediction: An overview. WWRP/THORPEX Publ. 15, 37 pp. https://www.wmo.int/pages/prog/arep/wwrp/new/documents/THORPEX_No_15.pdf

Majumdar, S. J. (2016). A Review of Targeted Observations. B. Am. Meteorol. Soc., 97(12), doi.org/ 10.1175/BAMS-D-14-00259.1

Montani, A., Thorpe, A. J., Buizza, R., & Unden, P. (1999). Forecast skill of the ECMWF model using targeted observations during FASTEX. *Q. J. R. Meteoro. Soc.*, 125, 3219-3240, doi.org/10.1002/qj.49712556106

Morss, R. E., & Battisti, D. S. (2004). Evaluating observing requirements for ENSO prediction: experiments with an intermediate coupled model. *J. Climate*, 17(16), 3057-3073, doi.org/10.1175/1520-0442(2004)017<3057:eorfep>2.0.co;2

Mourre, B., & Alvarez, A. (2012). Benefit assessment of glider adaptive sampling in the Ligurian Sea. *Deep Sea Res., Part I*, 68, 68-78, doi.org/ 10.1016/j.dsr.2012.05.010

Mu, M., Duan, W., & Wang, B. (2003). Conditional Nonlinear Optimal Perturbation and its applications. *Nonlinear Proc. Geoph.*, 10(6), 493-501, doi.org/10.5194/npg-10-493-2003

Mu, M., Zhou, F., & Wang, H. (2009). A method for identifying the sensitive areas in targeted observations for tropical cyclone prediction: Conditional Nonlinear Optimal Perturbation. *Mon. Weather Rev.,* 137(5), 1623-1639, doi.org/10.1175/2008mwr2640.1

Mu, M., Wansuo, D., & Tang, Y. (2017). The predictability of atmospheric and oceanic motions: Retrospect and prospects. *Sci. China (Earth Sci.)*, 11, 97-108, doi.org/10.1007/s11430-016-9101-x

Naimie, C. E., Blain, C. A., & Lynch, D. R. (2001). Seasonal mean circulation in the Yellow Sea - a model-generated climatology. *Cont. Shelf Res.*, 21, 667-695, doi.org/ 10.1016/s0278-4343(00)00102-3

Oke, P. R., Larnicol, G., Jones, E. M., Kourafalou, V., Sperrevik, A. K., Carse, F., et al. (2015). Assessing the impact of observations on ocean forecasts and reanalyses: Part 2, Regional applications. *J. Oper. Oceanogr.*, 8:sup1, s63-s79, doi.org/10.1080/1755876X.2015.1022080

Powell, M. J. D. (1983). VMCWD: A FORTRAN subroutine for constrained optimization. *ACM SIGMAP Bulletin*, 32, 4-16, doi.org/10.1145/1111272.1111273

Shay, L. K., Jaimes, B., Brewster, J. K., Meyers, P., McCaskill, E., C., Uhlhorn, E., et al. (2011). Airborne ocean surveys of the loop current complex from NOAA WP-3D in support of the deepwater horizon oil spill. *Geophysical Monograph Series*, 195(10), 131-151, doi.org/10.1029/2011GM001101

Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Model.*, 9, 347-404, doi.org/10.1016/j.ocemod.2004.08.002.

Wang, Q., Mu, M., & Dijkstra, H. A. (2013). The similarity between optimal precursor and optimally growing initial error in prediction of Kuroshio large meander and its application to targeted observation. *J. Geophys. Res.*, 118(2), 869-884, doi.org/ 10.1002/jgrc.20084

Wang, B., & Tan, X. (2009). A fast algorithm for solving CNOP and associated target observation tests. *J. Meteorol. Res.*, 23(4), 387-402, doi.org/ 10.1016/j.agrformet.2009.02.014

Xia, C., Qiao, F., Yang, Y., Ma, J., & Yuan, Y. (2006). Three-dimensional structure of the summer circulation in the Yellow Sea from a wave-tide circulation coupled model. *J. Geophys. Res.*, 111, C11S03, doi.org/10.1029/2005JC003218.

Xu, D., Yuan, Y., & Liu, Y. (2002). The baroclinic current structure of Yellow Sea Cold Mass. *Sci. China Ser. D*, 46(2), 117-126, doi.org/ 10.1360/03yd9011

Yang, D., Yin, B., Liu, Z., & Feng, X. (2011). Numerical study of the ocean circulation on the East China Sea shelf and a Kuroshio bottom branch northeast of Taiwan in summer. *J. Geophys. Res.*, 116, C05015, doi.org/10.1029/2010JC006777.

Zhang, K., Mu, M., & Wang, Q. (2017). Identifying the sensitive area in adaptive observation for predicting the upstream Kuroshio transport variation in a 3-D ocean model. *Sci. China (Earth Sci.)*, (05), 62-71, doi.org/CNKI:SUN:JDXG.0.2017-05-004

Zhang, K., Mu, M., Wang, Q., Yin, B., & Liu, S. (2019). CNOP-based adaptive observation network designed for improving upstream Kuroshio transport prediction. *J. Geophys. Res.*, 124, 4350-4364, doi.org/10.1029/2018JC014490

Zhang, S., Wang, Q., Lü, Y., Cui, H., & Yuan, Y. (2008). Observation of the seasonal evolution of the Yellow Sea Cold Water Mass in 1996-1998. *Cont. Shelf Res.*, 28(3), 442-457, doi.org/10.1016/j.csr.2007.10.002