OPEN

Improving the Real-time Marine Forecasting of the Northern South China Sea by Assimilation of Glider-observed T/S Profiles

Shiqiu Peng[1,2,6]*, Yuhang Zhu [1,7], Zhijin Li[5], Yineng Li[1], Qiang Xie [4], Shijie Liu[3], Yeteng Luo[3], Yu Tian[3] & Jiancheng Yu[3]*

Prediction of marine conditions is notoriously challenging in the northern South China Sea (NSCS) due to inadequate observations in the region. The underwater gliders that were developed during the past decade may provide observing platforms that could produce required observations. During a field experiment, temperature/salinity (T/S) profiles from a set of underwater gliders were assimilated into a real-time marine forecasting system, along with the assimilation of climatological monthly mean Argo data to constrain the basin-wide model biases. The results show that, in addition to the reduction of the basin-wide model biases by the assimilation of the climatological monthly mean Argo data, the assimilation of glider-observed T/S profiles is efficient to reduce the local biases of the NSCS marine forecasting by as much as 28–31% (19–36%) in 24 h to 120 h forecasts for temperature (salinity) from sea surface to a depth of 1000 m. Our results imply that the real-time marine forecasting for the NSCS can largely benefit from a sustainable glider observing network of the NSCS in the future.

As the largest marginal sea in Northwest Pacific, the South China Sea (SCS) takes a key role in connecting the Pacific Ocean to the Indian Ocean. An accurate marine forecast in this region is highly required to guarantee the safety of navigation, fishery, engineering construction, and so on. The SCS enjoys a unique geographical condition. It is controlled by annually-reversing monsoon system and featured with many processes of regional ocean dynamics and physics such as basin-scale circulations[1,2], strong western boundary currents[3], overflows[4,5], upwellings[6,7], energetic mesoscale eddies[8–10], internal tides[11,12] and gravity waves[13], thus its marine forecast remains a great challenge.

Numerical marine forecasting using an ocean circulation model relies on accurate initial conditions. Improving the accuracy of initial conditions is one of the most effective ways to reduce forecasting errors[14,15]. Data assimilation, which incorporates available observations into numerical models to reduce forecasting errors, is a common way at present to generate initial conditions of high accuracy for atmospheric or oceanic models[16–18]. There are different data assimilation algorithms, and currently the widely-used methods of data assimilation includes 3-dimensional/4-dimensional data assimilation (3DVAR/4DVAR)[19–23] and the Ensemble Kalman Filter (EnKF)[24–26]. Each data assimilation approach has its own advantages and disadvantages, and the 3DVAR has still been adopted in most real time high-resolution (~ a few km) marine forecasting systems in the world due to relatively low cost[16,27,28].

A variety of ocean surface observations have been provided by satellite-based remote sensing, but subsurface observations are very limited. Besides mooring buoys, Argo floats have been mainly used for subsurface measurements during the past couple of decades. They are generally located in the open ocean and are extremely sparse

[1]State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China. [2]Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China. [3]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China. [4]Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, China. [5]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA. [6]Institution of South China Sea Ecology and Environmental Engineering, Chinese Academy of Sciences, Guangzhou, China. [7]University of Chinese Academy of Sciences, Beijing, 100049, China. *email: speng@scsio.ac.cn; yjc@sia.cn
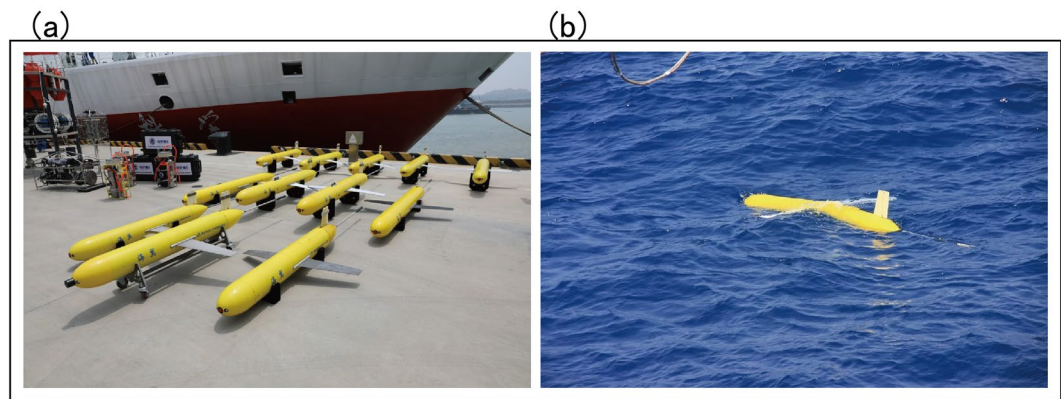
Figure 1. Pictures of Sea-Wing underwater gliders.

in coastal regions or marginal seas such as the SCS. Gliders work in a more controllable and flexible way than Argo floats and thus can be more suitable for coastal regions or marginal seas. In the past decade, gliders have been used increasingly in the world to measure temperature and salinity vertical profiles, especially in the coastal regions[29–31].

In China, gliders have been under intensive development and become available in recent years. To our knowledge, however, no real time assimilation in a real-time marine forecasting system has been carried out yet in the SCS. Beginning in July 2018, a field experiment of observing the upper ocean using China designed and manufactured gliders took place in the northern SCS (NSCS). In this paper, we report the assimilation of T/S profile observations from those gliders into a real-time marine forecasting system and demonstrate their significant impacts on forecasting skills.

The rest of this paper is organized as follows. The next section gives a brief introduction of the real-time forecasting system of marine environment for the SCS and data assimilation method used, followed by a description of the field observation experiment and the obtained T/S profile observations in the NSCS in section 3. Skill assessments are presented in section 4. A summary is given in the final section.

Field Observation Experiment and T/S Profile Observations

The field observation experiment in the NSCS started at August 4, 2018 and lasted for several months. Fourteen Sea-Wing underwater gliders (Fig. 1), developed by the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences[32], were deployed near the Xisha island to measure temperature and salinity profiles from the sea surface down a maximum depth of about 1000 m (Fig. 2). With a compact attitude-regulating unit and optimized steady gliding motion parameters, the Sea-Wing underwater glider was designed to save energy and increase gliding range[32,33]. Its performance has been comprehensively tested and assessed in field observation experiments during the past several years[34–36].

The Sea-Wing underwater gliders dive and rise between the sea surface and the 1000-m depth with a V-like path; one cycle of diving and rising, which crosses a maximum horizontal distance of about 4 km and takes about 4 h, gains a T/S profile with vertical sampling resolution of 1 m. The trajectories of the underwater gliders during one month from Aug. 6 to Sept. 7 and the evolution of daily profile number are shown in Fig. 2, in which 819 T/S profiles were obtained.

The real-time forecasting system and data assimilation method.   The real-time forecasting system, called a New-Generation Real-time Forecasting System for the SCS Marine Environment (NG-RFSSME), is an updated version of the Experimental Platform of Marine Environment Forecasting (EPMEF)[37] which was established in the State Key Laboratory of Tropical Oceanography (LTO), South China Sea Institute of Oceanology (SCSIO), in October 2010. It consists of three main components: an atmosphere model that employs the Weather Research and Forecasting (WRF) model, version 3.6[38], an ocean circulation model that adopts the Princeton Ocean Model (POM), 2002 version[39,40], and a sea wave model that uses the WAVEWATCH III (WWIII) model[41–43]. The WRF model, version 3.6, is a next-generation mesoscale numerical weather prediction system that was developed by the National Center for Atmospheric Research (NCAR) and the National Centers for Environmental Prediction (NCEP) to serve both operational forecasting and atmospheric research needs. A two-domain one-way-nested configuration is adopted for the WRF model with horizontal grid resolutions of 54 km and 18 km respectively and the inner domain covering the entire SCS and southern China (Fig. 2a), and there are 30 layers in the vertical for both domains. The initial and lateral boundary conditions of WRF model are the 6 hourly 1° × 1° output from the Global Forecast System (GFS) maintained by NCEP. The POM, developed in the Princeton University, is a three-dimensional (3D) primitive-equation ocean model embedded with a second-moment turbulence closure model (the Mellor–Yamada level 2.5 scheme)[44], it covers the entire SCS domain (Fig. 2a) with a horizontal resolution of 1/15 and 40 layers in the vertical and is two-way fully coupled to the inner-domain WRF model with a coupler OASIS3 (Ocean Atmosphere Sea Ice Soil III)[45] regarding to the heat and momentum flux exchanges between the ocean and the atmosphere; the climatologically monthly mean SODA (Simple Ocean Data Assimilation)[46,47] data are used for the open boundary conditions of the temperature/
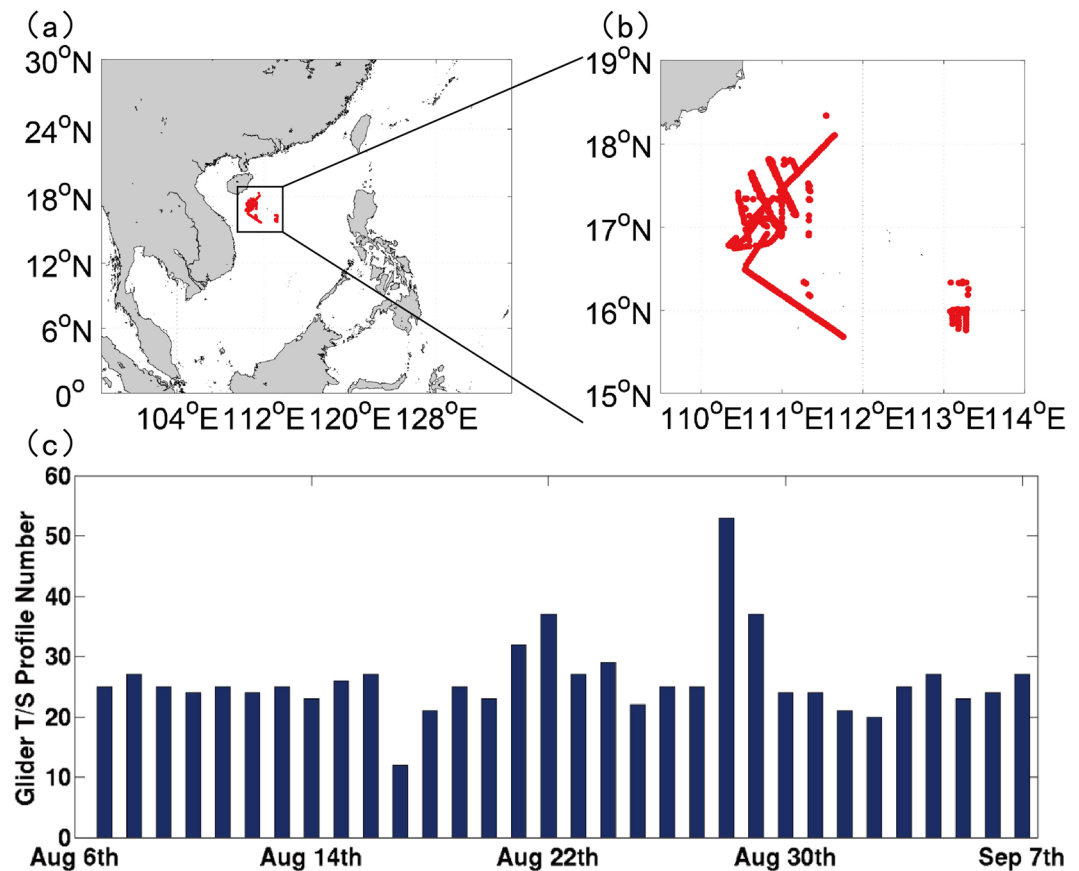
Figure 2. (**a**) Model domain of the NG-RFSSME and the region of the field experiment (denoted by the black square), (**b**) trajectories of the underwater gliders deployed in the field experiment and (**c**) the number of underwater gliders for each day from Aug 6 to Sep 7, 2018.

salinity/currents, and the OTPS (Oregon State University Tidal Prediction Software)[48] provides tidal levels and currents of 13 main tidal constituents at open boundaries. The WWIII model, which was developed at NOAA/ NCEP, is one-way coupled to the inner-domain of WRF model from which the 10-m-height winds are used as the dynamical forcing of the wave model.

Data assimilation modules are constructed separately for the atmospheric component and the oceanic component of the NG-RFSSME: in addition to the built-in WRF-3DVAR data assimilation system, a "scale-selective data assimilation" (SSDA) scheme[49–52] is incorporated in the WRF model, while a multi-scale 3DVAR (MS-3DVAR) data assimilation scheme[27,53–56] is applied to the POM model. In the SSDA scheme, a low pass filter is employed to perform scale separation on the wind fields from both GFS (6 hourly 1° × 1°) and WRF outputs, and then the large-scale component of GFS wind field is assimilated to adjust the large-scale component of the WRF wind field using 3DVAR method; the adjusted large-scale component of the WRF wind field is recombined with the unchanged small scale component of the WRF wind field to be a new wind field for the initial conditions at each forecast cycle. The MS-3DVAR takes into account the large- and small-scale information represented by different sampling density of different observation types, e.g., the sparse T/S profiles and ship-track sea surface temperature (SST) and the dense satellite-derived sea surface height anomaly (SSHA) and SST, and thus can effectively suppress or remove the spurious adjustments caused by the scale mismatch among different observations. Technically, the cost function to be minimized is divided into two parts:

$$J_L(\delta \boldsymbol{x}_L) = \frac{1}{2}\delta \boldsymbol{x}_L^T \mathbf{B}_L^{-1} \delta \boldsymbol{x}_L + \frac{1}{2}(\mathbf{H}\delta \boldsymbol{x}_L - \delta \boldsymbol{y}_L)^T(\mathbf{R} + \mathbf{H}\mathbf{B}_S\mathbf{H}^T)^{-1}(\mathbf{H}\delta \boldsymbol{x}_L - \delta \boldsymbol{y}_L),$$

(1)

$$J_S(\delta \boldsymbol{x}_S) = \frac{1}{2}\delta \boldsymbol{x}_S^T \mathbf{B}_S^{-1} \delta \boldsymbol{x}_S + \frac{1}{2}(\mathbf{H}\delta \boldsymbol{x}_S - \delta \boldsymbol{y}_S)^T(\mathbf{R} + \mathbf{H}\mathbf{B}_L\mathbf{H}^T)^{-1}(\mathbf{H}\delta \boldsymbol{x}_S - \delta \boldsymbol{y}_S),$$

(2)

in which the subscripts L and S represent large- and small-scale, respectively, and the superscript T represents the transpose operator. $\delta \boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^b$ is the increment of optimal values of the model variable vector $\boldsymbol{x}$ relative to their background values $\boldsymbol{x}^b$, while $\delta \boldsymbol{y} = \boldsymbol{y} - \mathbf{H}\boldsymbol{x}^b$ represents the deviation of observations $\boldsymbol{y}$ from their corresponding model output $\mathbf{H}\boldsymbol{x}^b$. $\mathbf{H}$, $\mathbf{B}$ and $\mathbf{R}$ are the Jacobian matrix of the nonlinearly observational operator, the background error covariance matrix and the observational error covariance matrix, respectively. Please be aware that the large-scale cost function (J_L) (1) includes small-scale background error covariance matrix $\mathbf{B}_S$ as the
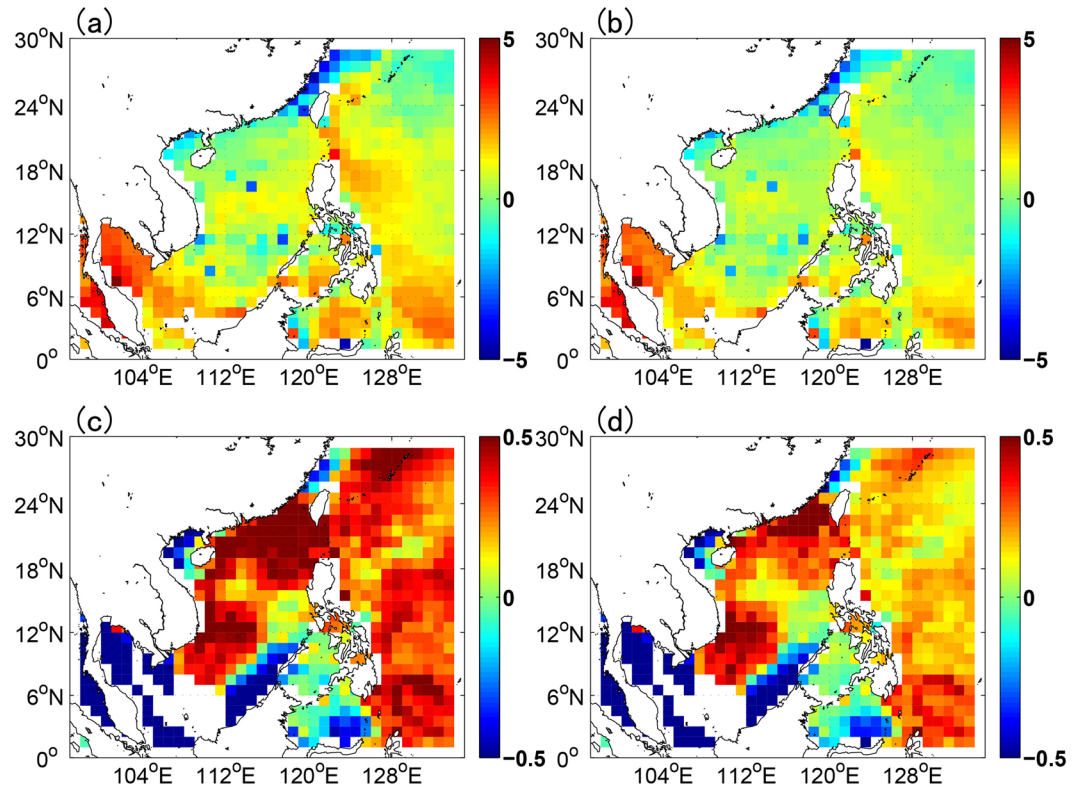
Figure 3. The biases of SST (unit: °C) (**a,b**) and SSS (unit: psu) (**c,d**) from model output before (**a,c**) and after (**b,d**) the assimilation of the climatological monthly mean Argo data at 0000 UTC Aug 7.

representativeness error ($\mathbf{HB}_S\mathbf{H}^T$) of observations, and vice versa; this dealing could help to reduce or eliminate the representativeness errors of observations, and thus improve the performance of the data assimilation with different sampling density of various observation types. Due to the huge dimensions of $\mathbf{B}_L$ (or $\mathbf{B}_S$), the construction of $\mathbf{B}_L$ (or $\mathbf{B}_S$) is a very tough work; to reduce the dimensions for a easy construction, here $\mathbf{B}_L$ (or $\mathbf{B}_S$) is decomposed into the standard deviation matrices and horizontal or vertical correlation matrices. Then the NMC method[57] is used to estimate the standard deviation matrices and the Kronecker product[58] is employed to decompose the 3-D matrices of horizontal and vertical correlation into 1-D matrices in both scales. In practice, the scale separation is achieved by performing the data assimilation procedure (including the calculation of matrix B) on coarse and fine model grids sequentially, in which the fine grid is the same as the original grid of the ocean model, while the coarse grid is 3 times the original grid. The large-scale data assimilation is first conducted to generate the large-scale increment $\delta\boldsymbol{x}_L$, which is then interpolated and added to the small scale background field $\boldsymbol{x}_S^b$ for the small-scale data assimilation. Therefore, $\delta\boldsymbol{y}_S$ is not equal to $\delta\boldsymbol{y}_L$ due to the difference of background fields between the small and large scales. Readers may refer to Li *et al.*[27] or Peng *et al.*[56] for details.

The effect of assimilating glider observations in improving the marine forecasting. Due to the heavily localizing feature of glider observations (Fig. 2) and the sparseness of the real-time Argo profiles in the NSCS, the climatological monthly mean Argo data, generated from the three-dimensional grid dataset (2004–2015) of Argo provided by China Argo Real-time Data Center[59], were assimilated to constrain the basin wide biases of the model in the NSCS at 0000 UTC of each day before the assimilation of glider-observed T/S profiles through the MS-3DVAR scheme. Figure 3 shows the biases of SST and sea surface salinity (SSS) from the model output before and after the assimilation of the climatological monthly mean Argo data; it is found that the biases of both SST and SSS were reduced obviously after the assimilation of the Argo data. Then the real-time T/S profiles obtained by gliders, which were subject to a quality control and then interpolated into the standard vertical layers, were assimilated into the POM of NG-RFSSME every 6 hours (i.e., at 0000 UTC, 0600 UTC, 1200 UTC and 1800 UTC) using the MS-3DVAR scheme, which is denoted as DA_CLIM-ARGO + GLIDER. As a preliminary result, a period of one month from Aug. 7 to Sept. 8 was selected for the assessment of the effect of assimilating the glider observations. For comparison, two paralleling experiments, i.e., one without any data assimilation and one with only climatological monthly mean Argo data assimilated, were performed for the same period, which are denoted as NODA and DA_CLIM-ARGO, respectively. A 5-day forecasting was made 4 times each day at 0000 UTC, 0600 UTC, 1200 UTC and 1800 UTC for all the experiments, and the forecasting results were validated against the "independent" glider observations (i.e., these observations had not been assimilated into the model by the validation time). For a more accurate evaluation on the role of glider observations in improving the marine forecasting skills, the satellite-derived SST/SSH and the real-time Argo T/S profiles were not assimilated in these experiments.
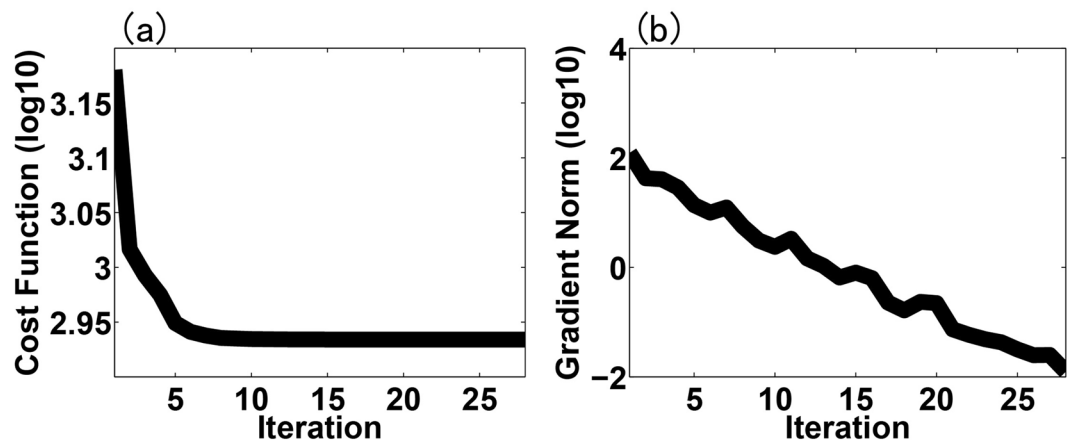
Figure 4. The evolution of (**a**) the small scale cost function (J_S) value and (**b**) its gradient norm in logarithmic scale with iteration number of the minimum procedure during the data assimilation cycle on 0000 UTC Aug 7.
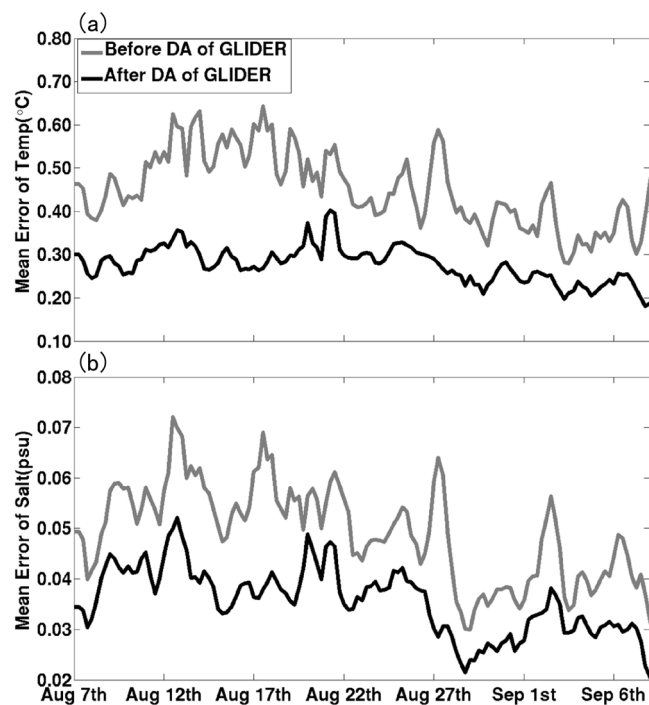

Figure 5. The vertically-integrated mean biases of (**a**) temperature and (**b**) salinity before (grey line) and after (black line) data assimilation of glider T/S profiles at each cycle validated against the glider observations during the one-month period of 7 Aug. to 8 Sept. 2018.

Figure 4 shows the evolution of the small scale cost function (J_S) value and its gradient norm in logarithmic scale with iteration number of the minimum procedure during one data assimilation cycle on 0000 UTC Aug. 7. It can be seen that the gradient norm of the cost function reduced to near zero at 10th iteration with cost function value reducing to about a half of the original value, implying that the MS-3DVAR system works well in the assimilation of glider observations. The vertically-integrated mean biases of temperature and salinity fields before and after data assimilation of glider T/S profiles at each cycle validated against the glider observations during the whole month are shown in Fig. 5, which are found to reduce from about 0.45 °C and 0.05 psu to about 0.28 °C and 0.04 psu that are close to the preset observational errors, respectively, further confirming the well digestion of glider observations by the MS-3DVAR system. Figures 6 and 7 give the mean biases of vertical T/S profiles for 24 h, 48 h, 72 h, 96 h and 120 h forecasts from NODA, DA_CLIM-ARGO and DA_CLIM-ARGO + GLIDER validated against the "independent" glider observations. The results demonstrate that, assimilating only the climatological monthly mean Argo data (DA_CLIM-ARGO) helped to reduce the model biases which could be basin-wide, and the assimilation of the real-time glider-observed T/S profiles was able to achieve an additional bias reduction of about 28–31% (19–36%) for the 24~120 h forecasts of temperature (salinity) from sea surface
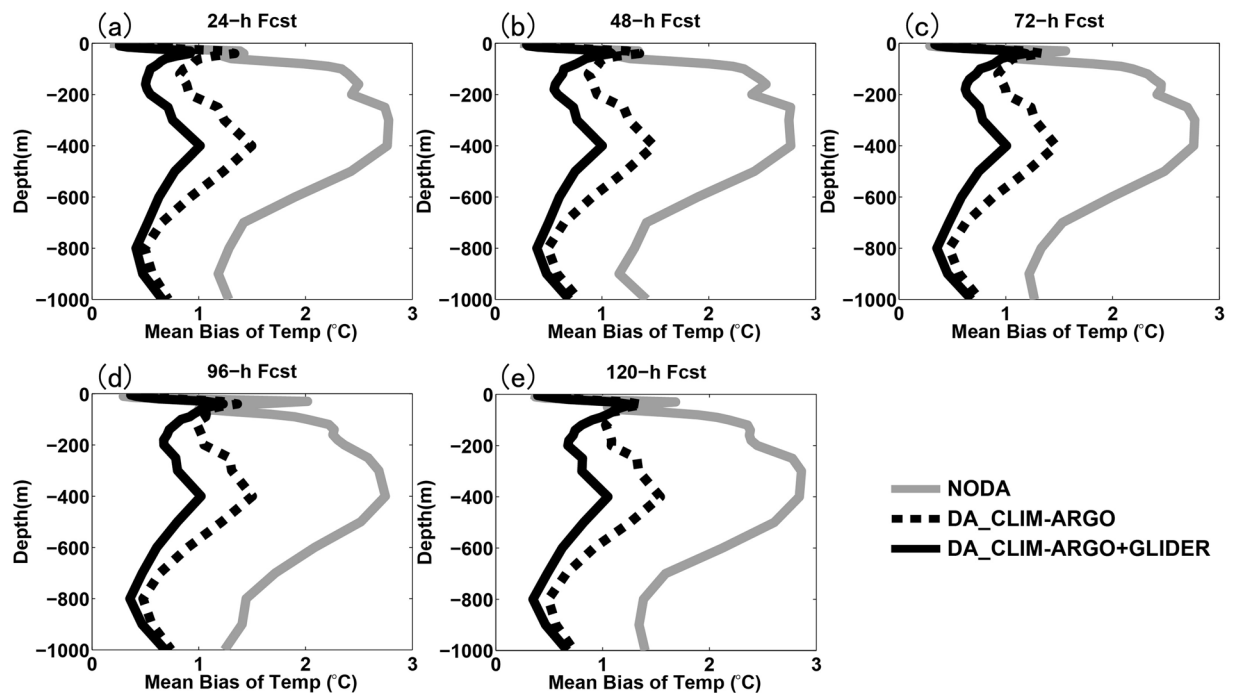
Figure 6. The mean biases of vertical temperature profiles for (**a**) 24-h, (**b**) 48-h, (**c**) 72-h, (**d**) 96-h and (**e**) 120-h forecasts from NODA (grey solid line), DA_CLIM-ARGO (black dotted line) and DA_CLIM-ARGO + GLIDER (black solid line) validated against the "independent" glider observations.
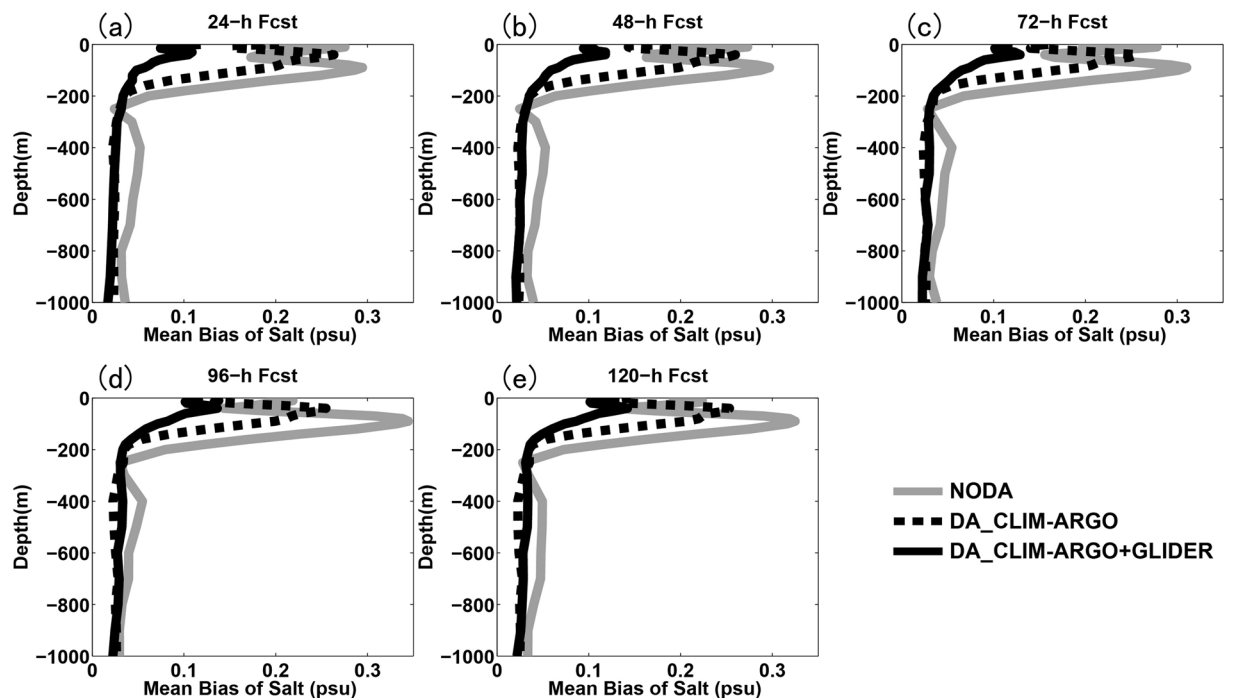
Figure 7. The same as Fig. 6, except for salinity.

to a depth of 1000 m (Table 1). Therefore, the glider observations could be very useful in improving the real time marine forecasting skills in the NSCS, especially for shorter forecasting periods, while it is also much beneficial and necessary to assimilate the climatological monthly mean Argo data for constraining the basin-wide model biases before assimilating these much localized glider observations.

Forecast Period		24 h	48 h	72 h	96 h	120 h	Mean
DA_CLIM-ARGO + GLIDER VS NODA	Temp	67.3%	66.6%	66.4%	65.7%	65.7%	66.3%
	Salt	58.3%	53.9%	51.2%	49.5%	48.7%	52.3%
DA_CLIM-ARGO VS NODA	Temp	52.4%	52.0%	52.3%	52.0%	51.8%	52.1%
	Salt	34.4%	34.1%	33.9%	35.5%	36.0%	34.8%
DA_CLIM-ARGO + GLIDER VS DA_CLIM-ARGO	Temp	31.3%	30.6%	29.6%	28.5%	28.8%	29.8%
	Salt	36.5%	30.1%	26.1%	21.7%	19.9%	26.9%

Table 1. The vertically-integrated mean bias reduction of temperature and salinity for different forecast periods from DA_CLIM-ARGO + GLIDER and DA_CLIM-ARGO compared to NODA, and DA_CLIM-ARGO + GLIDER compared to DA_CLIM-ARGO.

Summary

In this study, the role of assimilating glider observations in NSCS from a field observation experiment into a sophisticated marine forecasting system is evaluated. The Sea-Wing underwater gliders, which were designed and manufactured by the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, were deployed in the field observation experiment and 819 T/S profiles from the sea surface down a maximum depth of about 1000 m obtained during one month of Aug. 6 to Sept. 7 were assimilated using a multi-scale 3-D variation data assimilation method. The climatological monthly mean Argo data were assimilated to constrain the basin-wide model biases of the NSCS before the assimilation of the glider-observed T/S profiles due to the heavily localizing feature of glider observations and the sparseness of the real-time Argo profiles in the NSCS. The results demonstrate that the assimilation of glider-observed T/S profiles was able to improve the forecasting of the marine environment in the NSCS significantly, with a bias reduction of about 28–31% (19–36%) in one-day to five-day forecasting for temperature (salinity) in addition to the bias reduction by the assimilation of climatological monthly mean Argo data. Our results demonstrate that the glider observations are very useful in improving the real time marine forecasting skills in the NSCS, especially for shorter forecasting periods, while it is also much beneficial and necessary to assimilate the climatological monthly mean Argo data for constraining the basin-wide model biases before assimilating glider observations under the situation of rare real-time Argo observations in the NSCS.

Our results imply that a sustainable glider observing network of the NSCS is necessary and valuable to improve the real-time marine forecasting of the NSCS in the future. However, before its operational application in the real-time marine forecasts for the NSCS, some issues still need to be clarified: 1) what is the optimal sampling density of glider observations in both space and time for the data assimilation to improve the marine forecasting in the NSCS? 2) What is the combining effect of assimilating simultaneously glider observations and other types of observations such as the satellite-observed SSH and SST? And 3) besides the temperature and salinity, does the assimilation of glider-observed T/S profiles has any improvement in the forecasting of ocean currents? These issues will be investigated in our future work along with more field observation experiments in the NSCS.

Data availability

The SODA dataset is provided by NCAR, the GFS dataset is provided by NCEP, the tidal constituents used for the lateral boundary of POM is provided by the OTPS developed by Oregon State University, the climatological monthly mean Argo is provided by China Argo Real-time Data Center, and the underwater glider observed T/S profiles are provided by the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, and are available upon cooperation with the Shenyang Institute of Automation.

References

1. Qu, T. D. Upper-layer circulation in the South China Sea. *J. Phys. Oceanogr.* **30**, 1450–1460 (2000).
2. Liu, Q. Y., Kaneko, A. & Su, J. L. Recent progress in studies of the South China Sea circulation. *J. Oceanogr.* **64**, 753–762 (2008).
3. Wang, D. X. *et al.* Progress of regional oceanography study associated with western boundary current in the South China Sea. *Chinese Sci. Bull.* **58**, 1205–1215 (2013).
4. Wang, G. H., Xie, S. P., Qu, T. D. & Huang, R. H. Deep South China Sea circulation. *Geophys Res. Lett.* **38**, 3115–3120 (2011).
5. Lan, J., Zhang, N. N. & Wang, Y. On the dynamics of the South China Sea deep circulation. *J. Geophys. Res. Oceans.* **118**, 1206–1210 (2013).
6. Kuo, N. J., Zheng, Q. & Chung, R. H. Satellite observation of upwelling along the western coast of the South China Sea. *Remote Sens. Environ.* **74**, 463–470 (2000).
7. Xie, S. P., Xie, Q., Wang, D. X. & Liu, W. T. L. Summer upwelling in the South China Sea and its role in regional climate variations. *J. Geophys. Res.* **108**, 3261, https://doi.org/10.1029/2003JC001867 (2003).
8. Li, J. X., Zhang, R. & Jin, B. Eddy characteristics in the northern South China Sea as inferred from Lagrangian drifter data. *Ocean Sci.* **7**, 661–669 (2011).
9. Chen, G. X., Hou, Y. J. & Chu, X. Q. Mesoscale eddies in the South China Sea: Mean properties, spatiotemporal variability, and impact on thermohaline structure. *J. Geophys. Res. Oceans.* **116**, C06018, https://doi.org/10.1029/2010JC006716 (2011).
10. Chen, G. X. *et al.* Eddy heat and salt transports in the South China Sea and their seasonal modulations. *J. Geophys. Res. Oceans.* **117**, C05021, https://doi.org/10.1029/2011JC007724 (2012).
11. Farmer, D., Li, Q. & Park, J. H. Internal wave observations in the South China Sea: The role of rotation and non-linearity. *Atmos. Ocean.* **47**, 267–280 (2009).

12. Jackson, C. R. An empirical model for estimating the geographic location of nonlinear internal solitary waves. *J. Atmos. Oceanic Technol.* **26**, 2243–2255 (2009).

13. Guo, C. & Chen, X. A review of internal solitary wave dynamics in the northern South China Sea. *Prog. Oceanogr.* **121**, 7–23 (2014).

14. Shriver, J. F., Hurlburt, H. E., Smedstad, O. M., Wallcraft, A. J. & Rhodes, R. C. 1/32° real-time global ocean prediction and value-added over 1/16° resolution. *J. Marine Syst.* **65**, 3–26 (2007).

15. Wang, H., Liu, N., Li, B. X., & Li, X. An overview of ocean predictability and ocean ensemble forecast. *Adv. Atmos. Sci.* **29**, 1212–1225 (2014) (In Chinese).

16. Usui, N. *et al*. Meteorological research institute Multivariate Ocean Variational Estimation (MOVE) system: Some early results. *Adv. Space Res.* **37**, 806–822 (2006).

17. Chassignet, E. P. *et al*. US GODAE: Global ocean prediction with the hybrid coordinate ocean model (HYCOM). *Oceanography.* **22**, 48–59 (2009).

18. Martin, M. J., Hines, A. & Bell, M. J. Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. *Q. J. Roy. Meteor. Soc.* **133**, 981–995 (2007).

19. Barker, D. M., Huang, W., Guo, Y. R., Bourgeois, A. J. & Xiao, Q. N. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Weather Rev.* **132**, 897–914 (2004).

20. Peng, S. Q. & Zou, X. Assimilation of NCEP multi-sensor hourly rainfall data using 4D-Var approach: A case study of the squall line on 5 April 1999. *J. Met. Atm. Phy.* **81**, 237–255 (2002).

21. Peng, S. Q. & Zou, X. Assimilation of ground-based GPS zenith total delay and rain gauge precipitation observations using 4D-Var and their impact on short-range QPF. *J. Met. Soc. Japan.* **82**, 491–506 (2004).

22. Huang, X. Y. *et al*. Four-dimensional variational data assimilation for WRF: formulation and preliminary results. *Mon. Weather. Rev.* **137**, 299–314 (2009).

23. Gustafsson, N. *et al*. Four-dimensional variational data assimilation for a limited area model. *Tellus A: Dynamic Meteorology and Oceanography.* **64**, 1, https://doi.org/10.3402/tellusa.v64i0.14985 (2012).

24. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans.* **99**, 10143–10162, https://doi.org/10.1029/94JC00572 (1994).

25. Evensen, G. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynam.* **53**, 343–367 (2003).

26. Houtekamer, P. L. & Mitchell, H. L. Data assimilation using an ensemble kalman filter technique. *Mon. Weather. Rev.* **126**, 796–811 (1998).

27. Li, Z. J., Chao, Y., Mcwilliams, J. C. & Ide, K. A three-dimensional variational data assimilation scheme for the regional ocean modeling system. *J. Geophys. Res. Oceans.* **113**, C05002 (2008).

28. Tolman, H. & Mehra, A. GOVST-III meeting presentations: RTOFS Global Status and Plans. Paris: The 3rd meeting of the GODAE Ocean View Science Team, https://www.godae-oceanview.org/outreach/meetings-workshops/annual-science-team-meetings/govst-iii-meeting/presentations/ (2011).

29. Todd, R. E., Rudnick, D. L. & Davis, R. Monitoring the greater San Pedro Bay region using autonomous underwater gliders during fall of 2006. *J. Geophys. Res. Oceans.* **114**, C06001, https://doi.org/10.1029/2009JC005086 (2009).

30. Todd, R. E., Rudnick, D. L., Mazloff, M. R., Cornuelle, B. D. & Davis, R. Thermohaline structure in the California Current System: Observations and modeling of spice variance. *J. Geophys. Res. Oceans.* **117**, C02008 (2012).

31. Rudnick, D. L. & Cole, S. T. On sampling the ocean using underwater gliders. *J. Geophys. Res. Oceans.* **116**, 1–12, https://doi.org/10.1029/2010JC006849 (2011).

32. Yu, J. C. *et al*. Development and experiments of the sea wing underwater glider. *China Ocean Eng.* **25**, 721–736 (2011).

33. Yu, J. C., Zhang, F. M., Zhang, A. Q., Jin, W. M. & Tian, Y. Motion parameter optimization and sensor scheduling for the sea-wing underwater glider. *IEEE J. Oceanic. Eng.* **38**, 243–254 (2013).

34. Qiu, C. H. *et al*. Sea surface cooling in the Northern South China Sea observed using Chinese sea-wing underwater glider measurements. *Deep-Sea Res. Part I. Oceanographic Research Papers.* **105**, 111–118 (2015).

35. Shu, Y. Q., Xiu, P., Xue, H. J., Yao, J. L. & Yu, J. C. Glider-observed anticyclonic eddy in northern South China Sea. *Aquatic Ecosystem Health & Management.* **19**, 233–241 (2016).

36. Shu, Y. Q. *et al*. Field-observation for an anticyclonic mesoscale eddy consisted of twelve glider and sixty-two expendable probes in the northern South China Sea during summer 2017. *Sci. China Earth Sci.* **61**, 1–8 (2018).

37. Peng, S. Q. *et al*. A real-time regional forecasting system in the South China Sea and its performance in the track forecasts of tropical cyclones during 2011–2013. *Weather Forecast.* **30**, 471–485 (2015).

38. Skamarock, W. C. & Klemp, J. B. A time-split non hydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.* **227**, 3465–3485 (2008).

39. Blumberg, A. F. & Mellor, G. L. A description of a three-dimensional coastal ocean circulation model. *Three-Dimensional Coastal Ocean Models, Coastal and Estuarine Studies Series.* (ed. Heaps, N.). **4**, 1–16, (Amer: Geophys. Union, 1987).

40. Mellor, G. L. Users guide for a three-dimensional, primitive equation, numerical ocean model. *Program in Atmospheric and Oceanic Sciences, Princeton University*, www.aos.princeton.edu (2004).

41. Tolman, H. L. A new global wave forecast system at NECP. *Ocean Wave Measurements and Analysis.* (ed Edge, B. L. & Helmsley, J. M.) **2**, 777–786 (1997).

42. Tolman, H. L. User manual and system documentation of WAVEWATCH-III version 1.18. *NOAA/NWS/NCEP/OMB Tech. Note* **166**, 110 (1999).

43. Tolman, H. L. User manual and system documentation of WAVEWATCH-III version 2.22. *NOAA/NWS/NCEP/MMAB Tech. Note* **222**, 133 (2002).

44. Mellor, G. L. & Yamada, T. Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.* **20**, 851–875 (1982).

45. Valcke, S., Terray, L. & Piacentini, A. The OASIS Coupler User's Guide, Version 2.4. *Tech. Rep. TR/CGMC/00–10, CERFACS, Toulouse, France*, 85pp (2000).

46. Carton, J. A., Chepurin, G., Cao, X. H. & Giese, B. A simple ocean data assimilation analysis of the global upper ocean 1950–95. *Part I: Methodology. J. Phys. Oceanogr.* **30**, 294–309 (2000a).

47. Carton, J. A., Chepurin, G. & Cao, X. H. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part II: Results. *J. Phys. Oceanogr.* **30**, 311–326 (2000b).

48. Egbert, G. D., Bennett, A. F. & Foreman, M. G. G. TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res. Oceans.* **99**, 24821–24852 (1994).

49. Peng, S. Q., Xie, L., Liu, B. & Semazzi, F. Application of scale-selective data assimilation to regional climate modeling and prediction. *Mon. Weather. Rev.* **138**, 1307–1318 (2010).

50. Liu, B. & Xie, L. A scale-selective data assimilation approach to improving tropical cyclone track and intensity forecasts in a limited-area model: A case study of hurricane felix (2007). *Weather Forecast.* **27**, 124–140 (2012).

51. Xie, L., Liu, B. & Peng, S. Q. Application of scale-selective data assimilation to tropical cyclone track simulation. *J. Geophys. Res.* **115**, D17105 (2010).

52. Lai, Z. J. *et al*. On improving tropical cyclone track forecasts using a scale-selective data assimilation approach: a case study. *Nat. Hazards.* **73**, 1353–1368 (2014).

53. Zeng, X. Z., Peng, S. Q., Li, Z. J., Qi, Y. Q. & Chen, R. Y. A reanalysis dataset of the South China Sea. *Sci. Data.* **1**, 140052 (2014).

54. Li, Z. J., McWilliams, J. C., Ide, K. & Farrara, J. D. A multi-scale data assimilation scheme: formulation and illustration. *Mon. Weather. Rev.* **143**, 3804–3822 (2015).
55. Li, Z. J., McWilliams, J. C., Ide, K. & Farrara, J. D. Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme. *Ocean Dynam.* **65**, 1001–1015 (2015).
56. Peng, S. Q., Zeng, X. Z. & Li, Z. J. A three-dimensional variational data assimilation system for the South China Sea: preliminary results from observing system simulation experiments. *Ocean Dynam.* **66**, 737–750 (2016).
57. Parrish, D. F. & Derber, J. C. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **120**, 1747–1763 (1992).
58. Graham, A. Kronecker products and matrix calculus with applications. (Chichester, UK, Ellis Horwood, 1981).
59. Zhang, C. L., Xu, J. P., Liu, Z. H. & Sun, Z. H. User manual of three-dimensional grid dataset (GDCSM_Argo). China Argo Real-time Data Center. 17pp, http://www.argo.org.cn/index.php?m=content&c=index&f=lists&catid=32 (2018) (In chinese).

Acknowledgements

Author contributions

S.P. conceived the study and designed the numerical experiments of data assimilation and forecast, while Y.Z. carried out the numerical experiments and analyzed the results; Z.L. developed the MS-3DVAR system used in the forecast system and helped to design the experiment; Y.L. contributed to the development of the air-sea coupled forecast system and the analysis of the experiment results; J.Y., Y.L., Y.T. and S.L. developed the Sea-Wing underwater gliders and conducted the field observation experiment of underwater gliders; S.P. prepared the manuscript with contributions from all co-authors, and Q.X. contributed to both the field observation experiment and manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.P. or J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.