

Estimation and prediction of the upper ocean circulation in the Bay of Bengal



Ganesh Gopalakrishnan ^{a,*}, Aneesh C. Subramanian ^b, Arthur J. Miller ^a, Hyodae Seo ^c, Debasis Sengupta ^d

^a Climate, Atmospheric Science and Physical Oceanography, Scripps Institution of Oceanography, USA

^b University of Colorado Boulder, USA

^c Woods Hole Oceanographic Institution, USA

^d Indian Institute of Science, India

A B S T R A C T

The upper ocean stratification and circulation in the Bay of Bengal (BoB) plays a key role in the northward propagating monsoon intraseasonal oscillation during the months of June–August. This region is highly influenced by strong, seasonal atmospheric forcing and the oceanic circulation is characterized by dominant mesoscale variability and strong horizontal gradients in salinity and temperature during the monsoon period. Given the role of the ocean in the monsoon circulation, it is important to investigate accurate ocean state estimates and forecasts of the BoB ocean circulation in preparation for coupled ocean-atmosphere modeling and predictions. Hence, we use a mesoscale-permitting regional implementation of Massachusetts Institute of Technology general circulation model (MITgcm) and its adjoint-based four-dimensional variational (4DVAR) system to assimilate satellite-derived Sea Surface Height (SSH) and Sea Surface Temperature (SST) data in the BoB for a period of one month (June 1–30, 2017). It is shown that the MITgcm-BoB 4DVAR assimilation system is able to significantly improve the model consistency with the assimilated observations in the BoB region, reducing the model-data misfit by 50% and provided a dynamically-consistent BoB ocean circulation for the one-month hindcast period. We performed forecasting experiments using the state estimate to initialize two forecasts for a period of 30-days (July 1–30, 2017) from the end of the hindcast period. These forecasts used either atmosphere reanalysis and ocean analysis forcings or monthly climatology of atmosphere reanalysis and ocean analysis forcings. They therefore do not represent a “true” regional ocean forecast, forced using actual atmosphere and ocean forecasts, but bound the performance between climatological and nearly perfect forecasts. The model forecast is a cross-validation against future observations and showed that the initial conditions from the state estimate improves the prediction of the three-dimensional circulation in the BoB. The model hindcast and forecasts were also cross-validated against independent Argo temperature and salinity observations in the BoB. Additional state estimation and forecast experiments for other periods showed similar model performance with improved hindcasts and forecasts for the BoB region.

1. Introduction

Intraseasonal variability (ISV) is a dominant mode of variability in the tropical ocean-atmosphere system and has timescales between the chaotic weather systems and the long-term interannual and decadal climate modes of variability. The ISV influences many temporal and spatial phenomena that include the diurnal cycle of tropical convection (Tian et al., 2006; Oh et al., 2013; Seo et al., 2014), tropical cyclone activity (Bessafi and Wheeler, 2006; Kim et al., 2008), synoptic disturbances over the monsoon trough (Goswami et al., 2003; Neena and Goswami, 2010), Asian and Australian monsoons (Sikka and Gadgil, 1980; Hendon and Liebmann, 1990; Webster et al., 1998), and the El Nino Southern Oscillation (ENSO) (Lau and Chan, 1988; Takayabu et al., 1999; Roundy, 2008).

While the Madden-Julian Oscillation (MJO) is more active during

boreal winter (Madden, 1986), the Monsoon Intra-Seasonal Oscillation (MISO) is most active during boreal summer. The MISO manifests as a quasi-oscillatory mode embedded within the seasonal variability of the Asian summer monsoon (Webster et al., 1998). The MISO is seen as a strong northward propagating band of convection and precipitation over the Indian Ocean region and critically influences the weather and ocean state over the entire Bay of Bengal (BoB) region during the season.

Recent field campaigns in the BoB region as part of the Air-Sea Interactions in the Northern Indian Ocean (ASIRI) international research effort (2013–2017) have collected a unique set of observations to discern coupled atmosphere-ocean dynamics in the region (Wijesekera et al., 2016). ASIRI combined mature and new observational platforms to study the BoB region from submesoscale to regional scale dynamics and thermodynamics (Lucas et al., 2014). These analyses showed rich frontal features, submesoscale variability, shallow mixed-layer

* Corresponding author.

E-mail address: ganeshgopal@ucsd.edu (G. Gopalakrishnan).

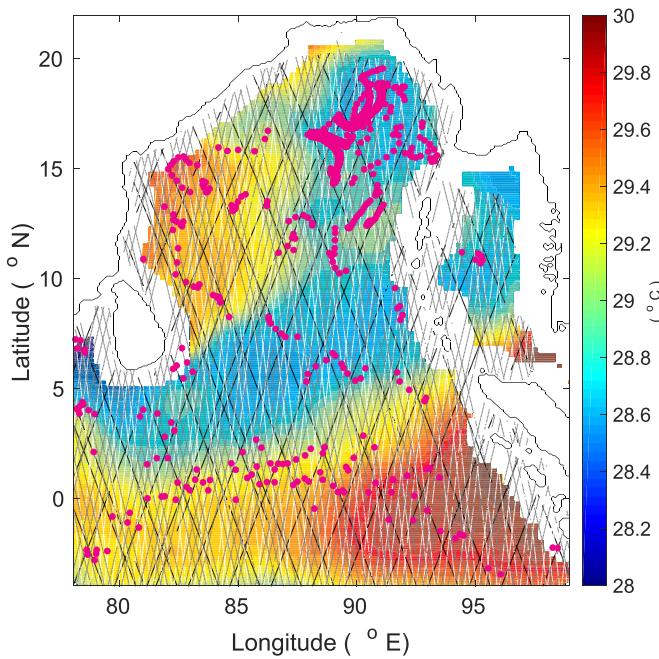


Fig. 1. Mean Sea Surface Temperature for the hindcast-forecast (June 1 – July 30, 2017) period. Superposed satellite tracks for J2 (black), C2 (white) and Sa (gray) satellites are shown. Argo profile locations are also marked by magenta filled circles.

dynamics with weak mixing and other dynamic features of the region, all of which were less known previously.

The Northern Indian Ocean is different from all other oceans of the world primarily due to its seasonally reversing monsoon system and its impact on billions of lives in one of the most densely populated regions in the world. The Indian Ocean is bound in the north at about 25° N by land. Hence the currents in the Indian Ocean cannot transport and discharge heat from the equatorial region to higher latitudes unlike the Gulf stream in the Atlantic Ocean or the Kuroshio Current in the Pacific Ocean. The unique feature of the North Indian Ocean is its two basins split into the Arabian Sea and BoB which is a recipe for its unique oceanographic features. These and many other unique features of the Indian Ocean have made it a region of considerable interest to oceanographers and atmospheric scientists alike.

Analyzing and predicting the Indian Ocean state can play an important role in improving the coupled ocean-atmosphere forecasts for the region. The availability of good quality synoptic datasets provided by satellites and also the increasing availability of in situ measurements from profiling floats deployed in the oceans provides an opportunity to meld observations with numerical ocean models to obtain an accurate estimate of the ocean circulation. This has prompted the oceanographic modeling community to adopt data assimilation (DA) in ocean models. Ocean observations are now commonly used to constrain numerical model solutions for ocean state estimates as well as forecasts. Efficient combination of numerical models and observations using the framework of DA is now recognized as the most effective way to forecast the state of the ocean (Ghil and Malanotte-Rizzoli, 1991; Wunsch, 1996; Edwards et al., 2015). Ocean models still have many deficiencies in reproducing the ocean state accurately due to erroneous atmospheric forcings, sub-grid scale parameterizations, and inaccurate initial and open-ocean boundary conditions. The theoretical framework of DA for oceanic problems is now well established and two separate methodologies are usually followed; the deterministic variational approach and the statistical filtering approach (Edwards et al., 2015). A detailed review of the developments and recent advances of variational and sequential methods with a focus on regional ocean assimilation systems is provided

in Edwards et al. (2015).

There is an increasing number of operational ocean analysis and forecasting systems worldwide, both regional and global. A comprehensive list of ocean analysis and forecasting systems is provided by the Global Ocean Data Assimilation Experiment (GODAE) OceanView (<https://www.godae-oceanview.org/science/ocean-forecasting-systems/>). Few of these systems are running in real-time operational/near-real-time mode. These systems use a variety of Ocean General Circulation Models (OGCMs) including Ocean Forecasting Australia Model (OFAM), Hybrid Coordinate Ocean Model (HYCOM), Massachusetts Institute of Technology general circulation model (MITgcm), Regional Ocean Model System (ROMS), Nucleus for European Modelling of the Ocean (NEMO), Meteorological Research Institute Community Ocean Model (MRI.COM), and Modular Ocean Model (MOM), and assimilate satellite and in situ ocean observations using both variational (three-dimensional, four-dimensional: 3DVAR, 4DVAR), and sequential (Kalman filter, smoother) methods. A detailed description of model, assimilated observations, and assimilation methods with their respective references are provided in the GODAE OceanView website. Although one can use these ocean analyses for regional/global ocean studies, the focus of this work is on improving regional-ocean predictability for high-resolution coupled ocean-atmosphere forecasts in the BoB and Indian Ocean.

We present a regional implementation of the MITgcm and its adjoint-based 4DVAR ocean assimilation system for the BoB. The MITgcm-BoB 4DVAR system is tested by assimilating satellite-derived Sea Surface Height (SSH) and Sea Surface Temperature (SST) observations for a period of one month (June 1 – 30, 2017) during the monsoon season. This time range is chosen to effectively utilize the data from the Office of Naval Research (ONR)'s ongoing MISO-BoB Departmental Research Initiative (DRI) observational campaigns. As a first step, this work describes the MITgcm-BoB model and the assimilation system using satellite SSH and SST from the global observing system and evaluates the model hindcast and forecast performance. This time period is during the southwest monsoon in the region (June–September), when the “cold dome” southeast of Sri Lanka, known as the Sri Lanka dome (SLD; Vinayachandran and Yamagata, 1998) often develops. The state estimate is cross-validated by forecasting the ocean state in the region for the following month (July 1 – 30, 2017), verifying the forecasts against the independent future observations in the region. We present results on the improved state estimates as well as forecasts. Additional state estimation experiments are also performed for other periods and their results are presented.

The MITgcm-BoB 4DVAR state estimation experiments are designed to explore both how long the model evolution can remain consistent with a variety of observations, and for how long it can skillfully predict the future ocean state. A longer ocean predictability lead time is desirable for coupled ocean-atmosphere modeling and its application to climate systems. The MITgcm-BoB 4DVAR state estimation experiments use HYCOM 1/12° global daily analysis (<http://hycom.org/data-server/glb-analysis>, hycom.org/dataserver/glb-analysis) (Chassignet et al., 2007) as the initial conditions and open-ocean boundary conditions for the “first-guess” (called the reference solution: REF). The HYCOM global analysis uses Navy Coupled Ocean Data Assimilation (NCODA) (Cummings, 2005). The MITgcm-BoB 4DVAR system changes this starting solution to reduce the differences between the model and the assimilated observations by adjusting the uncertain model controls using iterative optimization to obtain a regional ocean state estimate for ocean analysis, initializing ocean forecasts, and for coupled ocean-atmosphere modeling and prediction studies.

We present the MITgcm-BoB ocean model in Section 2, along with Subsection 2.1 where we discuss the solutions from a non-assimilated model integration. We then describe the assimilation methodology in Section 3, and state estimation experiments in Subsection 3.3 and forecast experiments in Subsection 3.4. The evaluation of the state estimate and forecast against observations is presented in Section 4, with

Table 1
MITgcm-BoB model description.

Description	Value/Source
Longitude	78° E – 99° E
Latitude	4° S – 22° N
Topography	Two-minute gridded global topography (ETOPO-2), with maximum depth of 6000 m.
Horizontal grid	1/12° × 1/12° (~ 9 km) spherical grid.
Vertical grid	50 vertical z - levels, with level spacing gradually increasing with depth from 2.5 m at the surface to a maximum of 300 m near the bottom.
Sub-grid scale parameterization	K-profile parameterization (KPP; Large et al., 1994).
Vertical diffusivity and viscosity	$2 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$ (fwd), $5 \times (2 \times 10^{-5} \text{ m}^2 \text{ s}^{-1})$ (adj)
Horizontal diffusivity and viscosity (second order)	$1 \times 10^{-2} \text{ m}^2 \text{ s}^{-1}$ (fwd), $5 \times (1 \times 10^{-2} \text{ m}^2 \text{ s}^{-1})$ (adj)
Horizontal diffusivity and viscosity (fourth order)	$1 \times 10^{10} \text{ m}^4 \text{ s}^{-1}$ (fwd), $5 \times (1 \times 10^{10} \text{ m}^4 \text{ s}^{-1})$ (adj)

* (fwd): forward model simulation.

* (adj): adjoint model simulation.

results from the hindcast verification in Subsection 4.1 and forecast verification in Subsection 4.2. Finally, we summarize the results and conclude in Section 5.

2. Ocean model and domain description

The MITgcm - Bay of Bengal (MITgcm-BoB) model is a regional implementation of the MITgcm (Marshall et al., 1997). The MITgcm solves the Navier-Stokes equations on a sphere under the Boussinesq approximation. The equations are written in z - coordinates and discretized using the third-order direct space time advection scheme in a staggered Arakawa C-grid. The MITgcm is equipped with automatic generation of its adjoint model using the Transformation of Algorithms in Fortran (TAF) (Giering and Kaminski, 1998; Heimbach et al., 2002). The model has been applied to numerous state estimation and sensitivity studies at global as well as regional scales (e.g. Stammer et al., 2002; Fukumori et al., 2004; Menemenlis et al., 2005; Hoteit et al., 2009, 2010, and 2013; Köhl et al., 2007; Mazloff et al., 2010; Gopalakrishnan et al., 2013a, b, and 2019; Verdy et al., 2017).

The MITgcm-BoB domain extends from 4° S to 22° N and from 78° E to 99° E (Fig. 1). The maximum bottom depth is at 6000 m and the bathymetry is derived from the 2-min gridded global topography (ETOPOV2, <https://www.ngdc.noaa.gov/mgg/global/etopo2.html>). The model is integrated on a 1/12° × 1/12° spherical grid, with 50 vertical z - levels. The vertical z - level spacing is 2.5 m at the surface, and the spacing gradually increases to a maximum of 300 m near the bottom. The sub-grid scale mixing is parameterized by K-profile parameterization (KPP; Large et al., 1994) and the mixing parameter values used in this setup are provided in Table 1. The model is operated in hydrostatic mode with an implicit free surface.

The lateral open-ocean boundaries are set at 4° S, 78° E, and 99° E. Temperature, salinity and the horizontal velocities (zonal and meridional) from the HYCOM/NCODA 1/12° global daily analysis were sampled every 7 days and are spatially interpolated to model grid and specified along the open-ocean boundaries. The HYCOM/NCODA horizontal velocities were prescribed at the grid points just outside the open-ocean boundary and the model solution is relaxed to these values within a “sponge” zone of 1° thickness, over time scales varying linearly from 1 day at the boundary to 5 days at the inner-edge of the zone. The HYCOM/NCODA normal velocity fields have been further adjusted by computing weighted-area transport across the open-ocean boundaries to exactly balance the net volume flux into the domain.

The atmospheric forcings were obtained from the Japanese 55-year reanalysis (JRA-55 reanalysis, https://jra.kishou.go.jp/JRA-55/index_en.html). The atmospheric forcing fields include air temperature,

specific humidity, zonal and meridional wind speed at 10 m from the ground, total precipitation, and short and long wave radiative fluxes. The model uses a bulk formulation (Large and Pond, 1981) for the computation of the atmospheric fluxes including the latent and sensible heat flux from the prescribed atmospheric forcing fields and simulated SST fields. The JRA-55 reanalysis forcing fields were sampled every 3 h on a global grid of 0.5° × 0.5°. Annual climatology of continental runoff (Fekete et al., 2002) were also used in this model. The MITgcm-BoB model is initialized using the HYCOM/NCODA 1/12° global daily analysis on January 1, 2009 and integrated over a 9-year period from 2009 to 2017 using the above model forcings.

2.1. Free model solution

The free (non-assimilated) model solution is archived as 6-day averaged fields and is compared with Archiving Validation and Interpretation of Satellite Oceanographic (AVISO) gridded Sea Surface Height (SSH) and Optimally Interpolated - Sea Surface Temperature (OI-SST). The HYCOM/NCODA 1/12° global analysis SSH and SST are also compared with AVISO-SSH and OI-SST, respectively, as an additional benchmark. The AVISO-SSH, OI-SST, and HYCOM/NCODA SSH and SST daily fields are sub-sampled as 6-day averaged fields prior to comparison. The daily gridded SSH fields from AVISO analysis are obtained from the Ssalto/Duacs altimeter product on a 1/4° longitude × 1/4° latitude grid, produced and distributed by the Copernicus Marine and Environment Monitoring Service (CMEMS) (<http://marine.copernicus.eu/>). The OI-SST data are obtained from the daily optimally interpolated product derived from the Tropical Rainfall Measuring Mission's (TRMM) Microwave Imager (TMI) and the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) instruments produced by Remote Sensing Systems Inc. (<http://www.remss.com/>) on a 1/4° longitude × 1/4° latitude grid.

The MITgcm-BoB SSH shows a close comparison with AVISO for both the mean and the standard deviation (Fig. 2). The positive values in the north and northeastern bay and negative values in the south and southwestern bay of MITgcm-BoB mean SSH shows a good comparison with AVISO than HYCOM/NCODA SSH. Both MITgcm-BoB and HYCOM/NCODA show a negative mean SSH representing a cyclonic circulation southeast (at 7° N and 84° E) of Sri Lanka marking the SLD, but this is not clearly seen in AVISO mapped SSH, likely due to its coarser 1/4° resolution. Another cyclonic circulation is found in MITgcm-BoB in the northern bay (18–19° N and 86 – 87° E) with an anticyclonic circulation south of it, both of which are again seen in HYCOM/NCODA SSH field, and are somewhat muted in the AVISO SSH fields. This northern cyclonic/anticyclonic feature corresponds to maximum SSH variability as shown by all SSH fields, and the variability spreads offshore and extends southward along the east coast of India to the southern tip of Sri Lanka. This southeastward advection is likely due to the seasonal, monsoon reversal of the coastal currents. The HYCOM/NCODA SSH field shows anticyclonic features east of the Andaman and Nicobar island chain with some variability, which is absent in the MITgcm-BoB and AVISO SSH fields. These differences near the island chain are hypothesized to depend on the model topography, and on how each model resolves the island group.

Both the MITgcm-BoB and HYCOM/NCODA SST means (Fig. 3) show cooler temperatures in the northern bay similar to OI-SST, while the latter shows more pronounced cooling than other SST fields and the cooling also extend southward towards the center of the bay. The northern bay also corresponds to stronger SST variability, as shown by all fields, mainly due to extreme precipitation and river-runoff. The MITgcm-BoB SST also shows cooler temperatures around Sri Lanka, which is seen in HYCOM/NCODA SST, but not clearly shown by OI-SST probably due to its coarser resolution of 1/4° longitude × 1/4° latitude. The HYCOM/NCODA SST time mean shows an intense cooling along the western and northern coasts, and along the eastern shelf-slope, which is absent in both MITgcm-BoB and OI-SST fields. This intense coastal

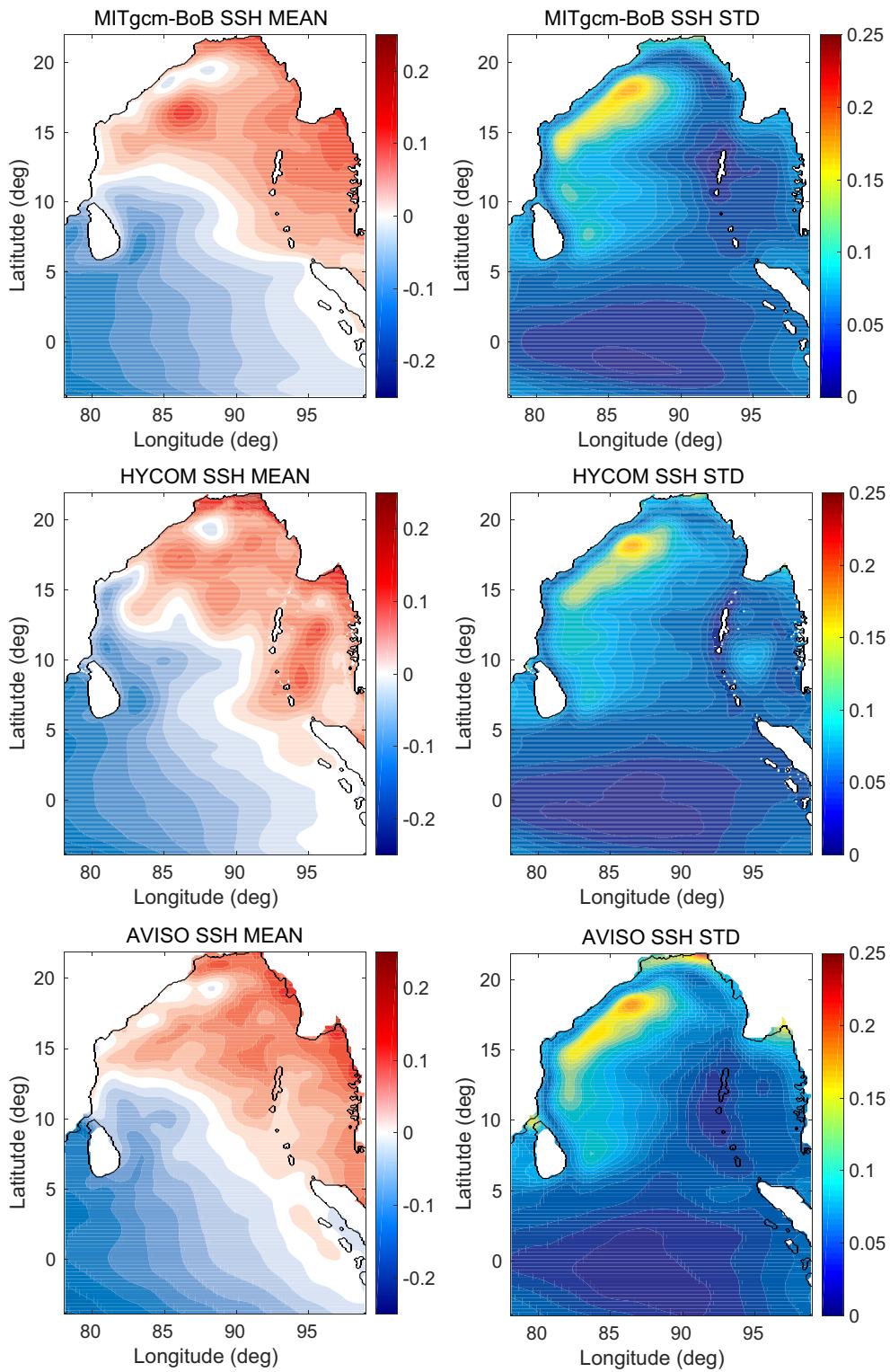


Fig. 2. Sea Surface Height mean (left panels) and standard deviation (right panels) for MITgcm-BoB (top panels), HYCOM/NCODA (middle panels), and AVISO (bottom panels). The SSH values are in m. The mean and standard deviation for the SSH fields are computed over the period of nine years (2009 – 2017), sampled as 6-day averaged fields.

cooling also corresponds to high SST variability in HYCOM/NCODA. The MITgcm-BoB SST shows warmer temperatures west of Indonesia than other SST fields. This region is close to the southern and eastern boundaries of the model, where open-ocean boundary condition errors may contribute to this mismatch.

Overall, the MITgcm-BoB free model solution qualitatively

reproduces the large scale SSH and SST features as observed by AVISO and OI-SST, and by HYCOM/NCODA 1/12° global daily analysis. This builds confidence in the model set-up and forcings and justifies its application for the ocean state estimation and prediction experiments discussed in the following sections.

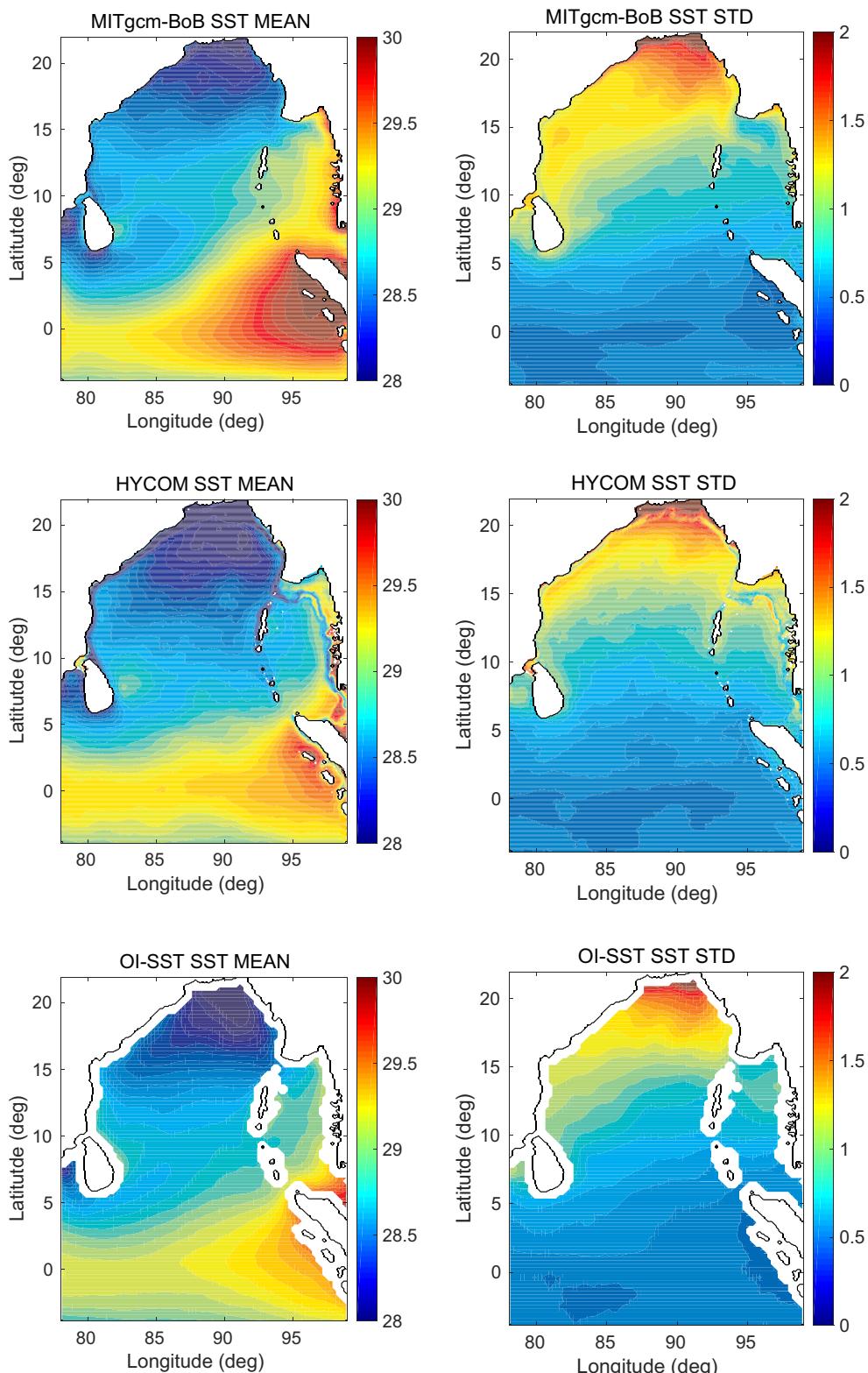


Fig. 3. Sea Surface Temperature mean (left panels) and standard deviation (right panels) for MITgcm-BoB (top panels), HYCOM/NCODA (middle panels), and OI-SST (bottom panels). The SST values are in °C. The mean and standard deviation for the SST fields are computed over the period of nine years (2009 – 2017), sampled as 6-day averaged fields.

3. Ocean state estimation

The state estimation methodology used in this study is a regional implementation of the MITgcm- Estimating the Circulation and Climate

of the Ocean (ECCO) 4DVAR system (Stammer et al., 2002; Wunsch and Heimbach, 2013; Forget et al., 2015). The state estimation is a minimization problem of a “cost function”, which is defined as a weighted sum of quadratic norms of both model-data misfit and changes to the

Table 2
Cost descent for June 1–30, 2017 assimilation experiment.

		Iteration 1	Iteration 5	Iteration 9
Total cost		27262	15765	13705
Observations	SSH (21162)	229	185	169
	MDT (20028)	7368	5521	4559
	SST (1451460)	19666	8761	6480
Controls	T0 (2403819)	0	715	1447
	S0 (2403819)	0	66	175
	Zonal wind (376764)	0	273	362
	Meridional wind (376764)	0	135	246
	Specific humidity (376764)	0	44	100
	Short-wave down (376764)	0	50	115

* The number of data/control points are provided in the parenthesis.

* The cost contribution of the controls for the first iteration are zero by definition.

control variables between the initial time and the final time of the assimilation period. The cost function is minimized subject to the nonlinear model equations by adjusting the control variables (Le Dimet and Talagrand, 1986; Wunsch, 1996). Here we use a “strong-constraint 4DVAR” state estimation, where we assume the ocean model to be perfect with no controls for errors in the model physics.

The gradient of the cost function, obtained by integrating the adjoint of the tangent linear model backward in time (Le Dimet and Talagrand, 1986), determines the descent directions toward the minimum cost function. The MITgcm-BoB 4DVAR system uses iterative optimization to reduce the cost function toward the minimum via a variable-storage Quasi-Newton M1QN3 algorithm (Gilbert and Lemarechal, 1989).

The following subsections briefly describe the observational constraints and uncertainties, and control variables and background uncertainty covariances used in this MITgcm-BoB 4DVAR system. A more detailed description with formulations of the state estimation procedure and iterative optimization, including cost function for controls and observations, similar to those used in this study can be found in Gopalakrishnan et al. (2013b).

3.1. Observations and uncertainties

SSH: SSH anomalies are obtained from the Radar Altimeter Database

Table 3
Model hindcast and forecast abbreviations.

	IC	OBCS	AF
Hindcast	SE	Adjusted IC	Adjusted OBCS
	REF	HYCOM/ NCODA	Adjusted AF JRA-55
Forecast	PER	Adjusted IC	
	F-SE	SE end state	HYCOM/NCODA JRA-55
	F-REF	REF end state	HYCOM/NCODA JRA-55
	F-HY	HYCOM/ NCODA	HYCOM/NCODA JRA-55
	F-PER	SE end state	
	F-SE-C	SE end state	HYCOM/NCODA monthly climatology JRA-55 monthly
	F-REF-C	REF end state	HYCOM/NCODA monthly climatology JRA-55 monthly
	F-HY-C	HYCOM/ NCODA	HYCOM/NCODA monthly climatology JRA-55 monthly

* IC: Initial conditions, OBCS: Open-ocean boundary conditions, AF: Atmospheric forcings.

* SE: State estimate, REF: Reference solution (first-guess).

* PER: Persistence solution (no time evolution, keeping initial state constant).

* All hindcasts and forecasts used annual climatology of continental runoff.

System (RADS (Scharroo et al., 2013), <http://rads.tudelft.nl/rads/index.shtml>). The along-track observations are obtained from three satellites: Jason-2 (J2), Cryosat-2 (C2), and SARAL/AltiKa (Sa) with respect to the time-mean dynamic ocean topography (MDT), calculated from the difference between the Danish National Space Center Mean Sea Surface 2008 (DNSCMSS08) and Earth Gravity Model 2008 (EGM08): DNSCMSS08-EGM08 (Andersen and Knudsen, 2009; Pavlis et al., 2012). These are computed over each day and spatially bin-averaged onto the model grid. Since the current configuration of the MITgcm-BoB did not include tidal or atmospheric pressure forcing, tidal corrections and inverted barometer correction are applied to the along-track SSH observations. The SSH data undergoes rigorous quality control prior to assimilation. The MDT constraint used in this setup comes from the MDT-CNES-CLS13, which is an estimate of the ocean MDT for the 1993–2012 period, produced by CLS and distributed by AVISO, with support from CNES (<https://www.aviso.altimetry.fr/>).

The model SSH time mean and anomalies are separately fit to the observed SSH time mean and anomalies using different uncertainties in the cost functions, isolating the generally larger uncertainties associated with the time-independent geoid from the generally smaller uncertainties in the anomalies. A spatially-constant SSH uncertainty of 5 cm is assigned for the SSH anomalies from satellite J2 which has a repeating orbit. The uncertainty for satellites C2 and Sa with non-repeating orbits are increased to 10 cm to account for the additional error caused by spatial binning of SSH observations taken at different locations into the same model grid cell. The MDT constraint is assigned with a spatially-constant uncertainty of 10 cm for all satellites to account for errors in the geoid estimate (Mazloff et al., 2014).

SST: The SST data are obtained from the OI-SST (Subsection 2.1). The daily gridded SST data are spatially interpolated onto the model grid prior to assimilation for computational convenience. The SST uncertainty is set at five times the standard deviation (over time) of the modeled SST (Subsection 2.1). The uncertainties range from 3 °C near the southern boundary to 8 °C near the northern boundary. This relatively high SST observational uncertainty is used to account for two factors: 1) the redundancy of observations due to spatial interpolation of coarsely gridded (1/4°) OI-SST to finer model grid (1/12°) cells, where the number of SST observations are duplicated to 3 × 3 “independent SST observations” per SST grid point, and 2) the model representational errors, especially near the surface, from sources including expected error in the atmospheric forcing fields that determine the mixed-layer depth, coarse time resolution of atmospheric forcings, and expected errors in the model mixing parameterizations, all of which affect the model mixed-layer physics. These representational errors are assumed to be high near the coast, especially in the shallow waters, and the extreme values of SST uncertainty are assigned only to those regions so that they would not affect the assimilation.

In addition to the above observations which are used in the state estimates, daily gridded SSH fields from AVISO analysis (Subsection 2.1), and temperature and salinity observations from Argo profiles are also used for evaluating the model hindcasts and forecasts. Argo profiling floats (<http://www.argo.ucsd.edu>) are obtained from the US-GODAE GDAC web site (<https://www.usgoda.org/argo/argo.html>), which are adjusted real-time profile data (“A” files). Although AVISO SSH analysis is not a perfect representation of the true ocean SSH, it is a widely accepted benchmark for global ocean SSH analysis.

The time mean of the OI-SST data for the hindcast-forecast period June 1 - July 30, 2017 is shown in Fig. 1 along with the superposed satellite ground tracks from three satellites: J2, C2 and Sa for the same period. Locations of Argo profiles for the same period are also marked. In order to ensure the quality of SSH and SST data and to account for expected model representational error, observations near the coast (at water depth < 500 m) are not used for assimilation.

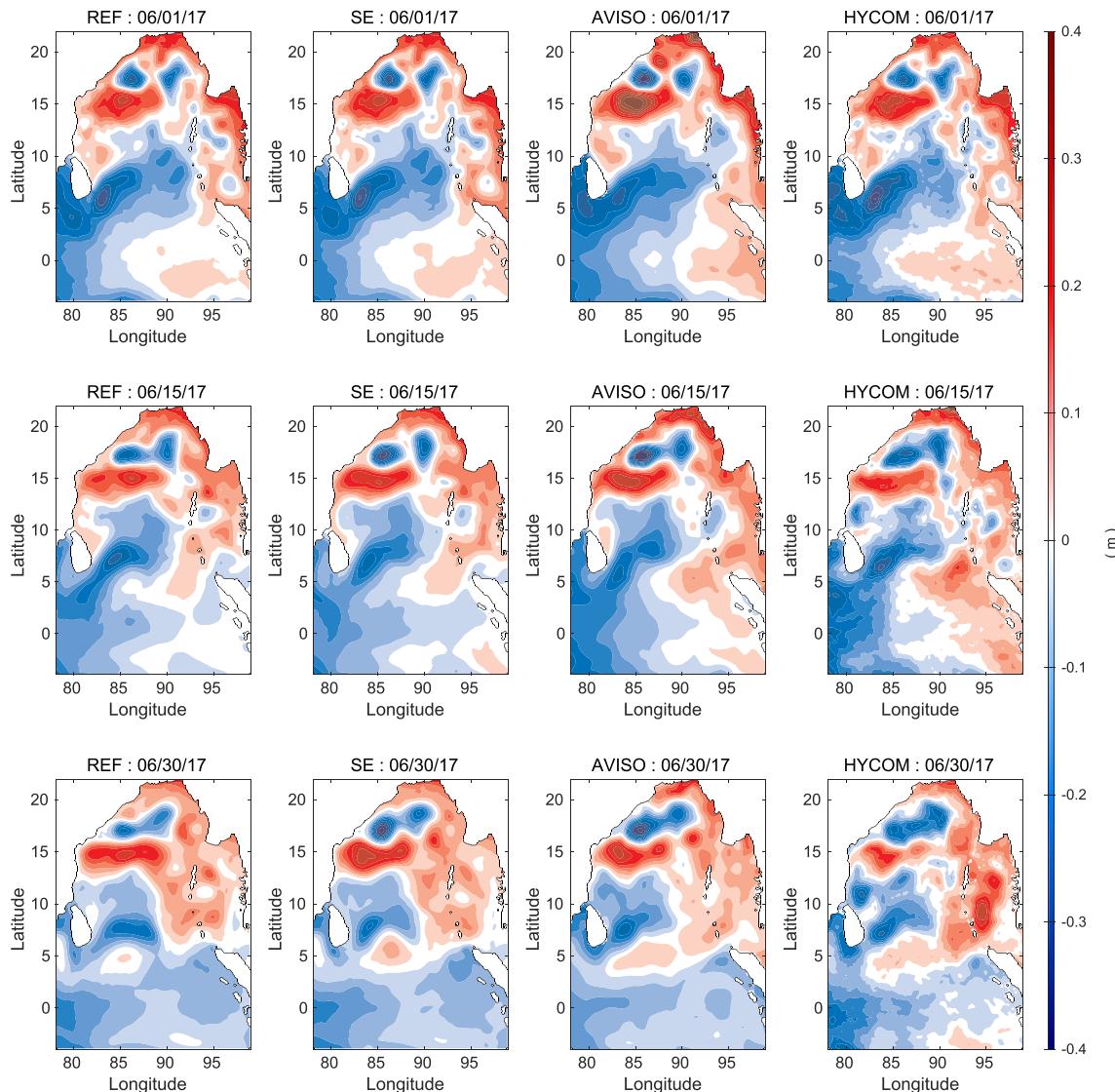


Fig. 4. Hindcast: Sea Surface Height snap shots for the day-1 (top panels), day-15 (middle panels), and day-30 (bottom panels) of the hindcast period for REF (first column), SE (second column), AVISO (third column) and HYCOM/NCODA (last column). The SSH values are in m.

3.2. Controls and uncertainties

The iterative optimization adjusts the model controls including initial conditions for temperature and salinity, open-ocean boundary conditions for temperature, salinity, and horizontal velocities, and atmospheric forcing fields, to bring the model solution into consistency with observational constraints.

In order to ensure smoothness of the control adjustment fields, the MITgcm-BoB 4DVAR system enforces 2D and 3D smoothness of control variables following Forget (2010) and Forget et al. (2015). The spatial correlation scale for the smoothed controls are assumed to be 50 km in the zonal and meridional direction, and use a small smoothing scale of 10 m in the vertical direction, so that each z-level was treated as independent. The smoothness operator suppresses the short-scale adjustments that might be generated by point measurements and maintains a smooth starting field for the model integration.

The background uncertainty covariances for the model initial temperature and salinity controls are computed from the standard deviation (over time) of the model variability (Subsection 2.1). These background error covariances are first multiplied by $\sqrt{\frac{dz_{min}}{dz}}$ (dz_{min} is minimum depth level thickness and dz is vertical z -level thickness) to compensate for the

large raw sensitivities at deep depth levels with large z -level thickness. The background error covariances are further scaled to account for the diagonality assumption of the error covariance, and for the 2D and 3D smoothing scales, by a factor of about 10 for the surface levels (50 km horizontal smoothing scale corresponds to about 5 grid points of $1/12^\circ$ longitude $\times 1/12^\circ$ latitude model grid, and 10 m vertical smoothing scale in the surface consists of four depth levels, together corresponds to a factor of 10), and by a factor of about 5 for the depth levels below. This scaling of the background error covariances is required to allow sufficient adjustments to initial condition controls for fitting the observations. Sensitivity experiments changing this scaling factor by a factor of 2 did not have much effect on the solutions, and the adjustments to initial condition controls are similar to those in the reported state estimates (not shown).

The uncertainties for the temperature and salinity open-ocean boundary controls are set to be the same as for the initial temperature and salinity controls. The uncertainties for the horizontal velocity controls at the open-ocean boundaries are set using velocity normal modes, with an root-mean-square (rms) value of $3 \times 10^{-4} \text{ m s}^{-1}$ for the barotropic mode and $\leq 7 \times 10^{-3} \text{ m s}^{-1}$ for the total of the rest of the modes with maximum energies in the first three baroclinic modes.

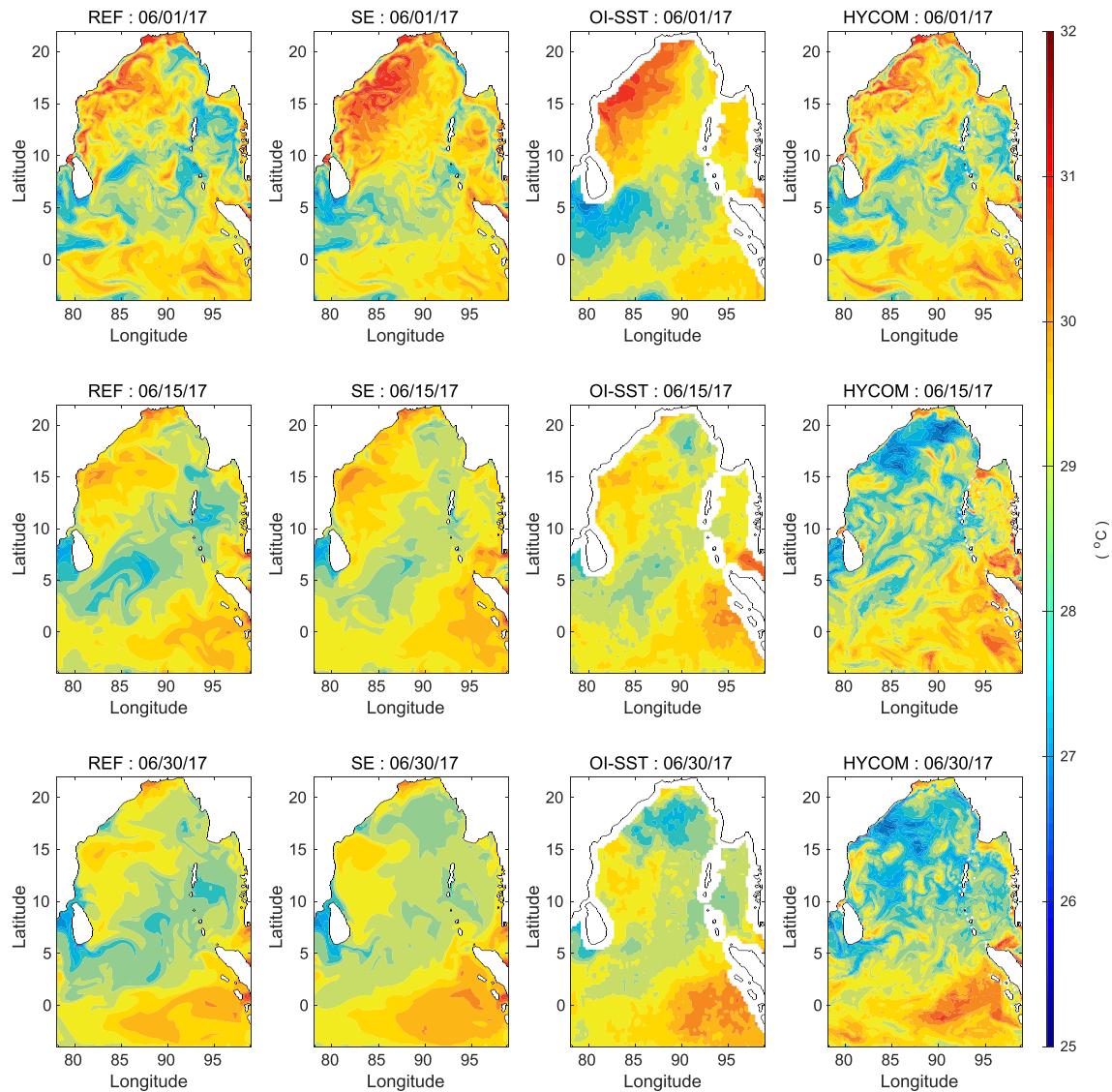


Fig. 5. Hindcast: Sea Surface Temperature snap shots for the day-1 (top panels), day-15 (middle panels), and day-30 (bottom panels) of the hindcast period for REF (first column), SE (second column), OI-SST (third column) and HYCOM/NCODA (last column). The SST values are in °C.

The uncertainties for the atmospheric forcing controls are set to be the standard deviation of National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis-1 (NCEP/NCAR-R1) (Kalnay et al., 1996) fluxes and winds computed over seven years (2004 – 2010), without removing the seasonal cycle.

The prior estimates of uncertainties for observations and controls all include model representational errors, especially near the surface, such as errors due to horizontal and vertical resolution, aliased atmospheric forcing, and model mixed-layer physics.

3.3. State estimation experiment

The satellite-derived along-track SSH, separated into temporal mean and anomalies, and gridded SST data are assimilated using the MITgcm-BoB 4DVAR system over an assimilation window of one-month: June 1 – 30, 2017. The state estimation fits the observations by making adjustments to the control variables and produces an optimized solution using iterative minimization.

The first-guess (reference solution, called as REF from here on) is a simulation initialized using assimilated HYCOM/NCODA 1/12° global daily analysis and are forced using JRA-55 reanalysis surface fluxes and winds, HYCOM/NCODA open-ocean boundary conditions, and

climatological run-off fluxes. The atmospheric forcings and open-ocean boundary conditions are linearly interpolated to each time-step of the model simulation. These initial conditions, open-ocean boundary conditions, and atmospheric forcings for the REF are the “background” meaning first-guess for the state estimation. The model control variables for atmospheric forcing and open-ocean boundary conditions in the state estimation are applied smoothly over 7-day periods. The cost contribution of the controls for the first iteration (background) are zero by definition, and the adjustments to the first-guess controls are penalized in the cost function.

The MITgcm-BoB 4DVAR adjoint model simulation uses increased horizontal and vertical diffusivity and viscosity terms (both second-order and fourth-order terms are increased by a factor of 5 compared to the REF simulation, refer to Table 1), with the KPP mixing parameterization turned off. This is implemented to filter out any growing nonlinear sensitivities at small scales and to extend the duration of system linearity, following Hoteit et al. (2005), Köhl et al. (2007) and Gopalakrishnan et al. (2013b). Using this approach, the large-scale sensitivities are little changed, but increases the damping of small-scale sensitivity structures, including those driven by flow instabilities, resulting in a smoother sensitivity field. This method allows the model to fit observations over longer assimilation periods, and has

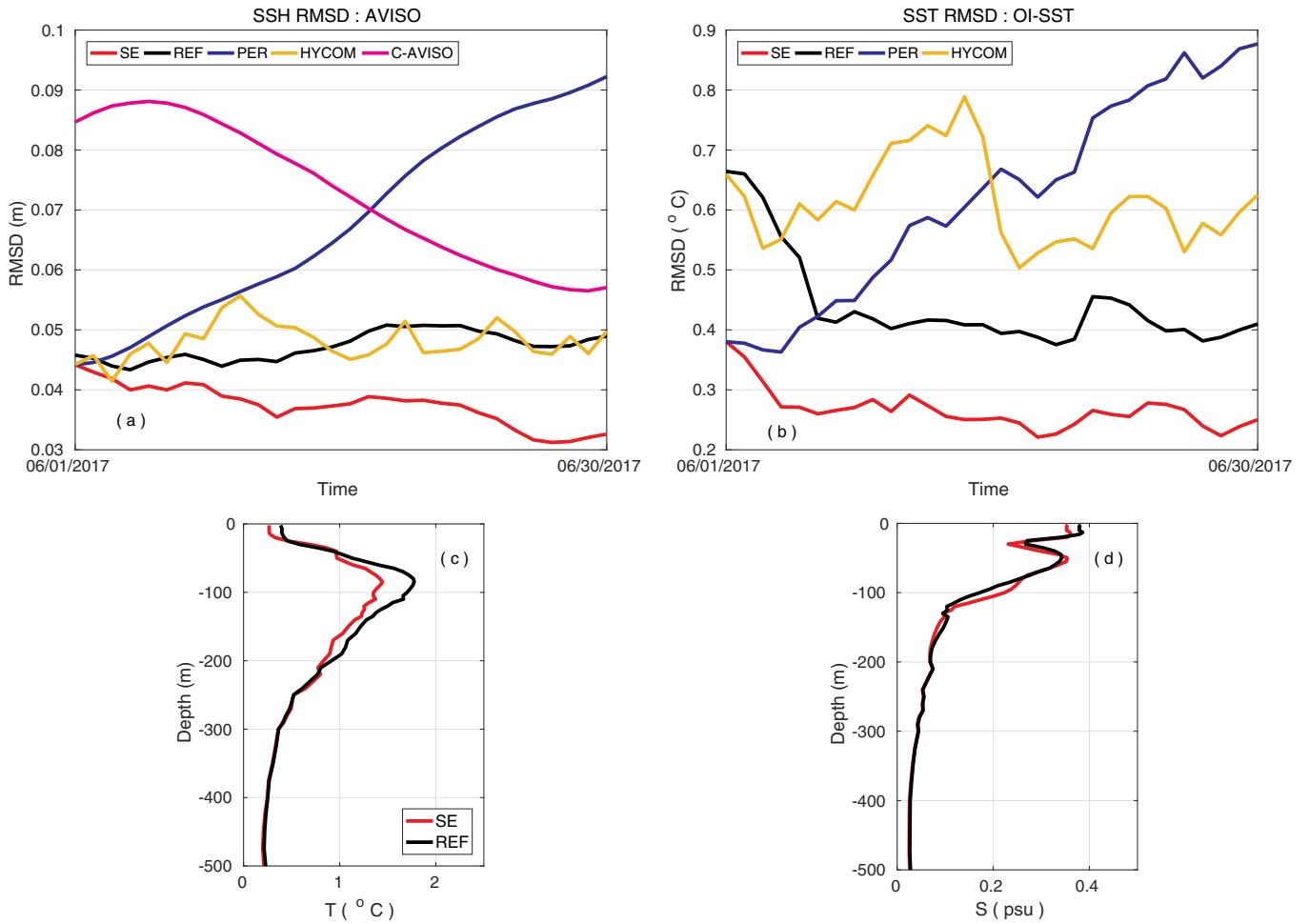


Fig. 6. Hindcast (June 1 – 30, 2017) model-data *rmsd* for Sea Surface Height (computed with respect to AVISO: top left panel) and for Sea Surface Temperature (computed with respect to OI-SST: top right panel). The *rmsd* is averaged over the model domain. The *rmsd* is computed for the optimized state estimate (SE: red), model persistence (PER: blue), reference model simulation (REF, first-guess from iteration 1: black), HYCOM/NCODA daily global analysis (HYCOM: golden), and AVISO SSH climatology (magenta). The SSH *rmsd* values are in m, and SST values are in °C. The *rmsd* comparison of SE and REF with Argo temperature (bottom left panel) and salinity (bottom right panel) data, for the hindcast: June 1 – 30, 2017 period. The *rmsd* values for temperature and salinity are in °C and psu, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

been successfully tested in other regional state estimations (Hoteit et al., 2010; Gopalakrishnan et al., 2019; Zaba et al., 2018).

The iterative optimization took 9 iterations for the total cost function to decrease and converge to a plateau with relatively small slope between the last two total cost functions. The total cost function is reduced by 50% after 9 iterations, and the individual observation costs decreased at roughly the same rate as the total cost function. The cost contribution from adjustments to the model controls, is dominated by the initial temperature and salinity control costs, and atmospheric wind control costs, which are much larger than other atmospheric forcing and open-ocean boundary control costs for this one-month short-term state estimation. The total cost function and the cost contribution from individual observations and dominant model controls for the first, middle, and the last iterations are summarized in Table 2. The optimized solution or “state estimate” (SE) is a dynamically-consistent free model solution over the assimilation period forced using adjusted model controls.

3.4. Forecast experiment

The optimized solutions are cross-validated by forecasting the ocean state for another 30 days (July 1 – 30, 2017) from the end of the assimilation period. The optimized solution (SE) at the end of the assimilation period is used as initial condition for the model forecasts

(F-SE and F-SE-C). The F-SE uses JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings, and F-SE-C uses monthly climatology of JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings. These two forecasts with different atmospheric forcings and open-ocean boundary conditions try to separate the ocean predictability from optimized initial conditions from that due to atmospheric and open-ocean boundary forcings in the BoB. It should be noted that these forecasts are not “true” regional ocean forecasts forced using actual atmosphere and ocean forecasts. F-SE-C gives the performance of a climatological forecast, while F-SE represents a nearly perfect forecast. Additional forecasts include: 1) persistence forecast (F-PER): assuming no time evolution of the F-SE solution (keeping the F-SE initial state fixed over the forecast period), 2) reference forecast (F-REF): forecast initialized from the REF solution at the end of the assimilation period, and 3) HYCOM/NCODA forecast (F-HY): forecast initialized from assimilated HYCOM/NCODA 1/12° global daily analysis for the start date of the forecast using MITgcm-BoB, are also compared with F-SE forecast. The F-REF and F-HY forecasts also uses either (1) JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings, or (2) monthly climatology of JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings (called “F-REF-C” and “F-HY-C”, respectively). All forecast experiments use climatological run-off fluxes and are simulated using the same viscosity and diffusivity

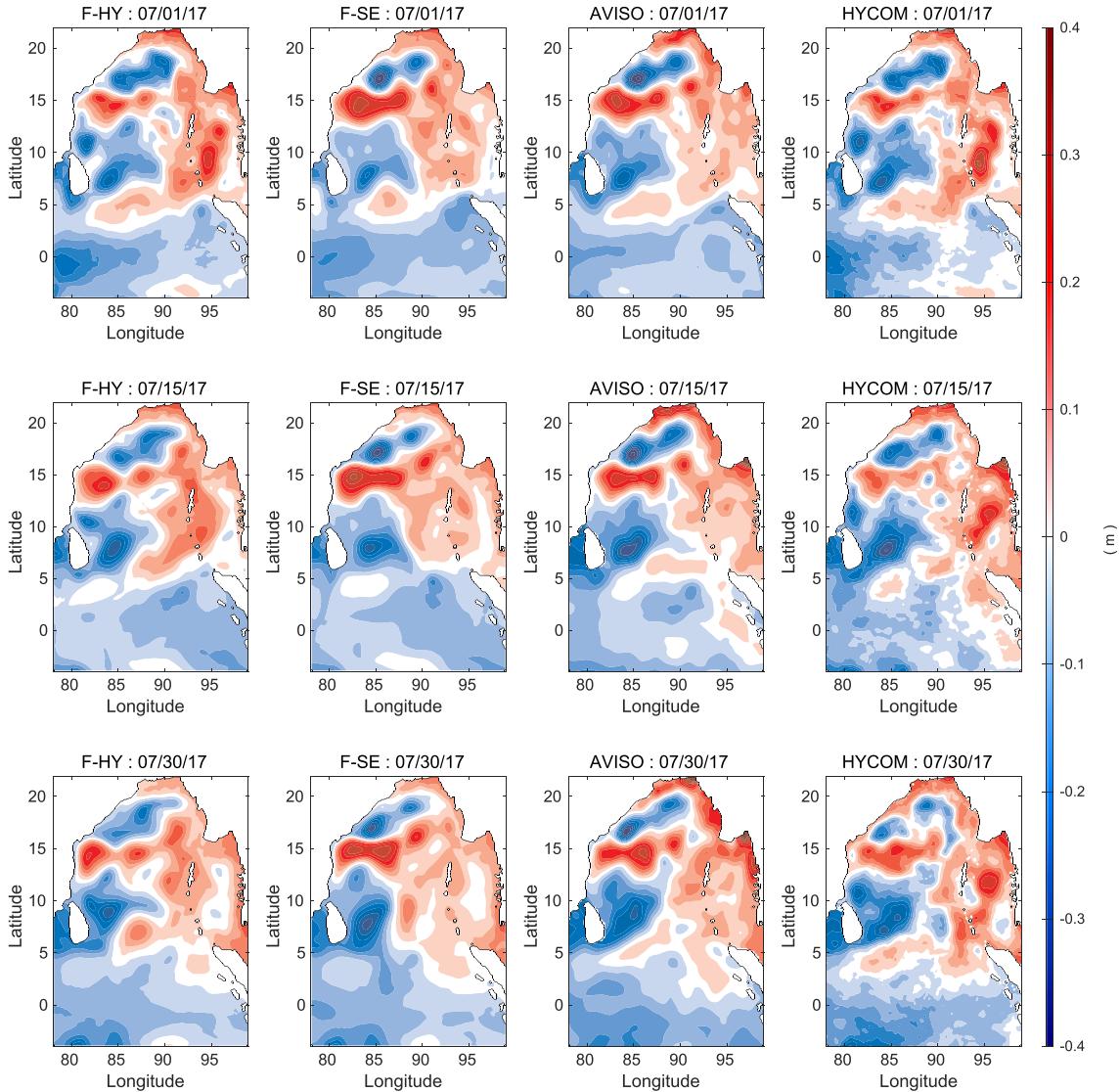


Fig. 7. Forecast: Sea Surface Height snap shots for the day-1 (top panels), day-15 (middle panels), and day-30 (bottom panels) of the forecast period for F-HY (first column), F-SE (second column), AVISO (third column) and HYCOM/NCODA (last column). SSH values are in m. The F-HY is forecast initialized from HYCOM/NCODA solutions using MITgcm-BoB. All the forecasts used JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings, and annual climatological run-off fluxes.

terms as that for the REF. The list of abbreviations used in various model hindcasts and forecasts with their respective initial conditions, open-ocean boundary conditions, and atmospheric forcings is provided in Table 3.

4. Comparison of hindcast and forecast with observations

The performance of MITgcm-BoB state estimates are analyzed by comparing the model hindcasts and forecasts with daily gridded AVISO SSH, daily gridded OI-SST, and Argo temperature and salinity profiles. The Argo temperature and salinity observations are not assimilated, and are used for independent data comparison with the model hindcasts and forecasts.

4.1. Hindcast comparison

The SSH snapshots from the first-guess (REF), optimized solution (SE), AVISO, and HYCOM/NCODA solutions for day-1, day-15, and day-30 of the June 1–30, 2017 state estimation experiment are compared in Fig. 4. The SE SSH for day-1 shows a closer comparison with HYCOM/

NCODA than AVISO. The REF SSH is also very much like that from HYCOM/NCODA due to its initialization. Although the large-scale SSH features of SE are similar to REF for day1, there are small-scale differences, mainly in the eastern and northern bay due to adjustments to the model controls. As time progresses to day-15, SE SSH matches better with AVISO than REF, especially the cyclonic/anticyclonic features in the northern bay and the cyclonic features southeast of Sri Lanka. At the end of the hindcast period (day-30), the SE SSH shows a better comparison with AVISO than REF or HYCOM/NCODA, particularly with respect to the location and intensity of cyclonic/anticyclonic features in the northern bay and southeast of Sri Lanka. The HYCOM/NCODA SSH shows small-scale features, whereas SE, REF and AVISO show smoother SSH fields. This is probably due to the daily assimilation update cycle of the HYCOM/NCODA solution, which allows more structure. The REF and SE are free model solutions forced using first-guess and adjusted controls, respectively, and show smooth fields, which are dynamically-consistent over the one-month hindcast period. The smooth AVISO SSH fields is due to the space-time covariance used in the mapping (Le Traon et al., 1998; Ducet et al., 2000).

The SST fields from SE, REF, OI-SST, and HYCOM/NCODA are

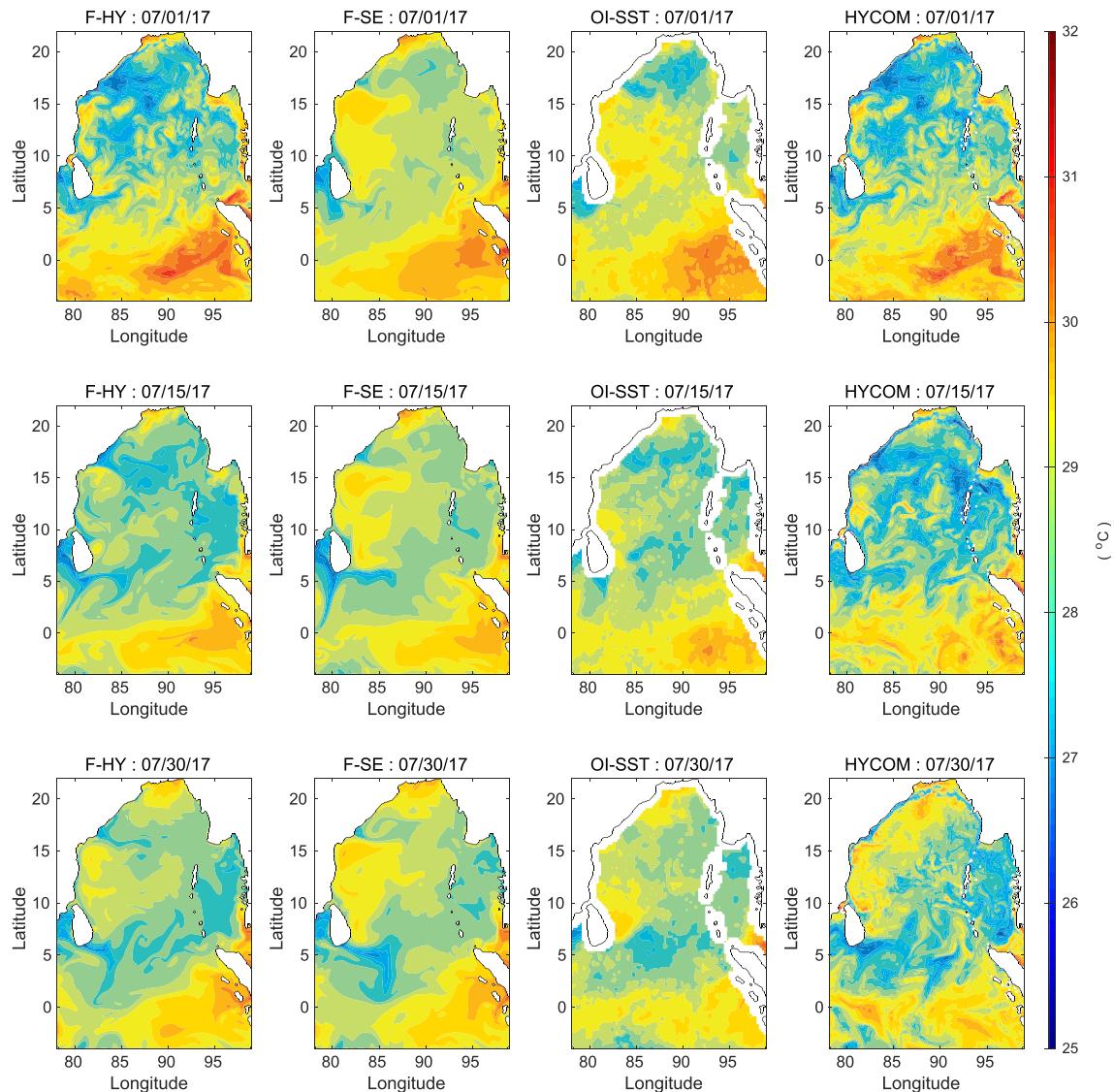


Fig. 8. Forecast: Sea Surface Temperature snap shots for the day-1 (top panels), day-15 (middle panels), and day-30 (bottom panels) of the forecast period for F-HY (first column), F-SE (second column), OI-SST (third column) and HYCOM/NCODA (last column). SST values are in °C. The F-HY is forecast initialized from HYCOM/NCODA solutions using MITgcm-BoB. All the forecasts used JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings, and annual climatological run-off fluxes.

compared for day-1, day-15, and day-30 for the hindcast period in Fig. 5. For day-1, SE SST shows most of the features of OI-SST, especially the warmer temperatures off the Indian coast, but fails to reproduce the cooler temperatures south of Sri Lanka. The REF SST shows a close comparison to HYCOM/NCODA for day-1 as expected. The SE SST reproduces most of the features shown by OI-SST for day15. The REF SST also shows a reasonable comparison to OI-SST for day-15, except showing more cooler temperatures in the center of the bay than OI-SST. The HYCOM/NCODA SST field is cooler (warmer) in the northern (southeastern) bay compared to other SST fields. At the end of the hindcast period, SE SST reproduces most of the OI-SST features than REF, but shows slightly warmer (cooler) temperatures in the northern (southeastern) bay than OI-SST. The HYCOM/NCODA SST shows cooler temperatures in the northwestern bay than SE, REF, and OI-SST fields, and shows more small-scale features than other SST fields.

The model-data root-mean-squared-difference (*rmsd*) for SE, REF, and PER SSH are computed daily with respect to AVISO SSH, and averaged over the whole model domain (Fig. 6a). HYCOM/NCODA daily global analysis SSH is also compared with AVISO to provide another benchmark for the *rmsd* comparison. The SE shows a decrease in SSH

rmsd from 4.5 cm to 3 cm over the hindcast period, and shows the lowest *rmsd* among the solutions. REF closely matches HYCOM/NCODA throughout the hindcast period with an average *rmsd* of ~ 4.8 cm. REF is simulated without any spin-up time, so there is some adjustment during the first day, but it still agrees with the HYCOM/NCODA SSH analysis. This result is important as it demonstrates the compatibility of the MITgcm-BoB model with the HYCOM/NCODA solutions and shows that the physics of the two models are very comparable. It also means that HYCOM/NCODA provides a good initialization (REF) with low SSH *rmsd* for the state estimation. The model persistence (PER), calculated by keeping the SE initial state constant over the hindcast period, shows an increase in *rmsd* from 4.5 cm to almost double, about 9 cm. The AVISO climatology (C-AVISO), which is monthly AVISO SSH climatology for the BoB without removing the seasonality, are also compared with AVISO SSH daily analysis for the hindcast period. The C-AVISO shows a decrease in *rmsd* from 8.5 cm to 6 cm during the hindcast period. This C-AVISO *rmsd* gives an estimate of SSH *rmsd* for the BoB in the absence of any model estimates and uses climatology as a prediction. The increasing PER SSH *rmsd* gives an estimate of the limited model persistence timescales of the BoB SSH evolution. The PER SSH *rmsd*

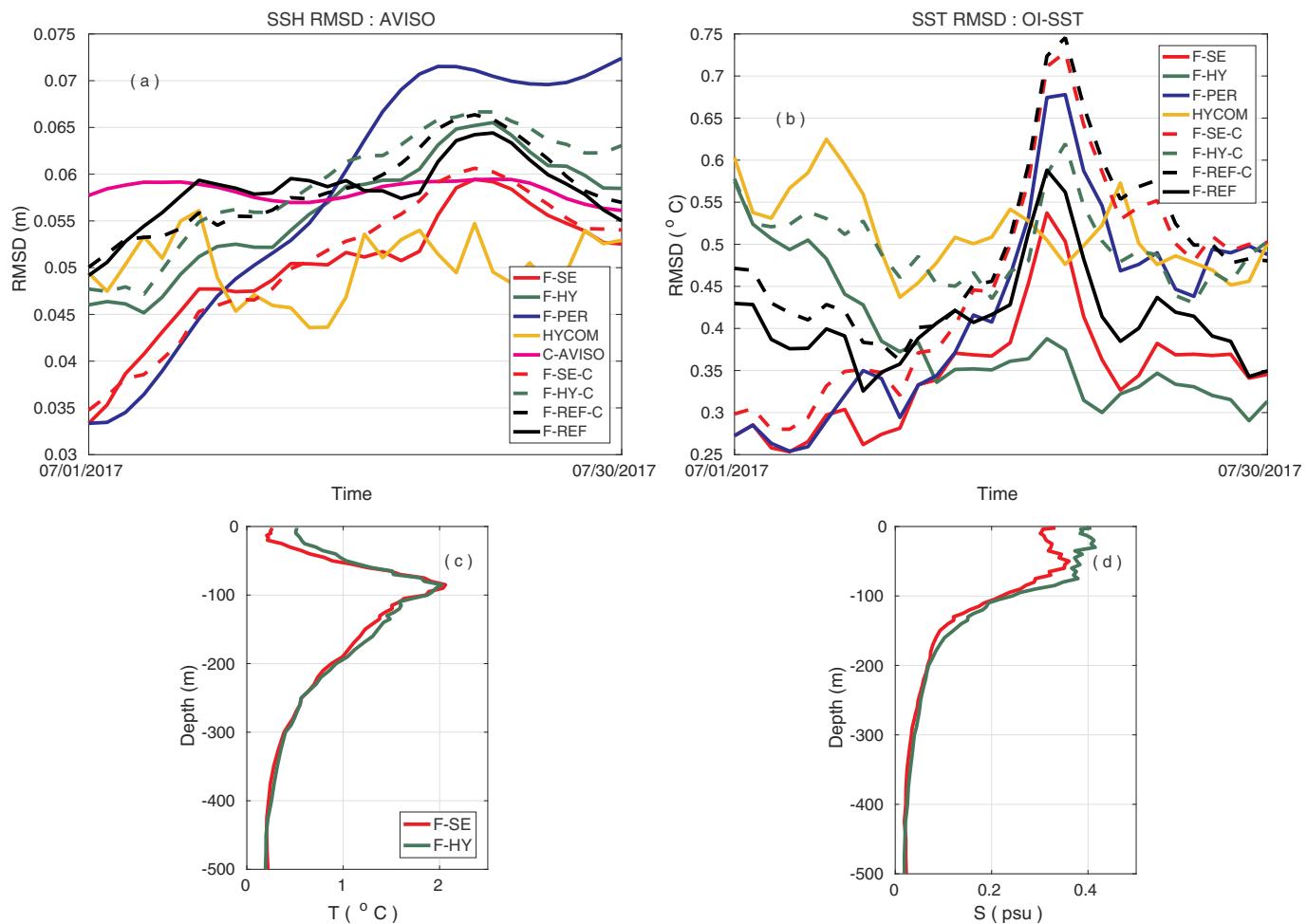


Fig. 9. Forecast (July 1 – 30, 2017) model-data *rmsd* for Sea Surface Height (computed with respect to AVISO: top left panel) and for Sea Surface Temperature (computed with respect to OI-SST: top right panel). The *rmsd* is averaged over the model domain. The *rmsd* is computed for the forecast initialized from SE (F-SE: red), model persistence forecast (F-PER: blue), forecast initialized from REF (F-REF: black), HYCOM/NCODA daily global analysis (HYCOM: golden), and AVISO SSH climatology (magenta). The green curve represents F-HY (forecast initialized from HYCOM/NCODA solutions using MITgcm-BoB). The dashed curves mark model forecasts using monthly climatology of JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings (F-SE-C, F-REF-C, F-HY-C). The SSH *rmsd* values are in m, and SST values are in °C. The *rmsd* comparison of F-SE and F-REF with Argo temperature (bottom left panel) and salinity (bottom right panel) data, for the forecast: July 1 – 30, 2017 period. The *rmsd* values for temperature and salinity are in °C and psu, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

exceeds that of C-AVISO at about two-weeks from the start of the hindcast period, suggesting a SSH persistence timescale of two weeks. We also computed SSH persistence skill from daily gridded AVISO analysis for the BoB region. The AVISO SSH persistence skill: defined as $1 - (\sum(SSH_{(0)} - SSH_{(lag \text{ in weeks})})^2 / \sum SSH_{(0)}^2)$, also drops quickly to about 0.7 within a lag time of two weeks (not shown). The climatology and persistence comparison discussed above are considered as the most widely used standards of reference (base line on which skill is measured) in the field of ocean/atmosphere hindcast and forecast verification (Murphy, 1992).

The model-data *rmsd* for SE, REF, and PER SST are computed daily with respect to OI-SST, and averaged over the whole model domain (Fig. 6b), as with SSH. HYCOM/NCODA SST are also compared with OI-SST for the hindcast period. Similar to SE SSH *rmsd* comparison (Fig. 6a), the SE SST *rmsd* shows a decrease from 0.4 °C to 0.2 °C over the hindcast period, and shows the lowest *rmsd* of all the estimates. The PER SST shows an increase in *rmsd* from 0.4 °C to 0.9 °C, whereas HYCOM/NCODA SST shows an average *rmsd* of 0.65 °C over the hindcast period. The REF SST *rmsd* starts at a higher value, same as HYCOM/NCODA of ~0.65 °C and decreases within a week to 0.4 °C and stays more or less constant over the rest of the hindcast period. Again, this REF solution without any model spin-up demonstrates the MITgcm-BoB compatibility

with HYCOM/NCODA initialization.

The SE and REF are compared with non-assimilated Argo data, and model-data *rmsd* are computed at the space-time locations of the observations and are shown as functions of depth in Fig. 6c,d for temperature and salinity, respectively. During the hindcast period, there are about 750 Argo profiles in the BoB, and the *rmsd* is averaged over all profiles regardless of location. SE shows a slight improvement over REF in the surface layers (0 – 10 m) for both temperature and salinity. SE also shows more improvement for temperature than salinity in the depths between 50 and 200 m, with maximum improvement at 90 m. The assimilation of SST data likely contributes to this temperature improvement over salinity. Even though, the model-data *rmsd* is computed over fewer profiles, which are sparsely located in the BoB, the improvements in SE over REF show that assimilation of satellite-derived SSH and SST not only improves the sea surface state, but also the three-dimensional ocean state.

4.2. Forecast comparison

The SSH forecasts from F-SE and F-HY are compared with AVISO and HYCOM/NCODA daily global analysis for the day-1, day-15, and day-30 of the July 1 – 30, 2017 forecast period in Fig. 7. It should be

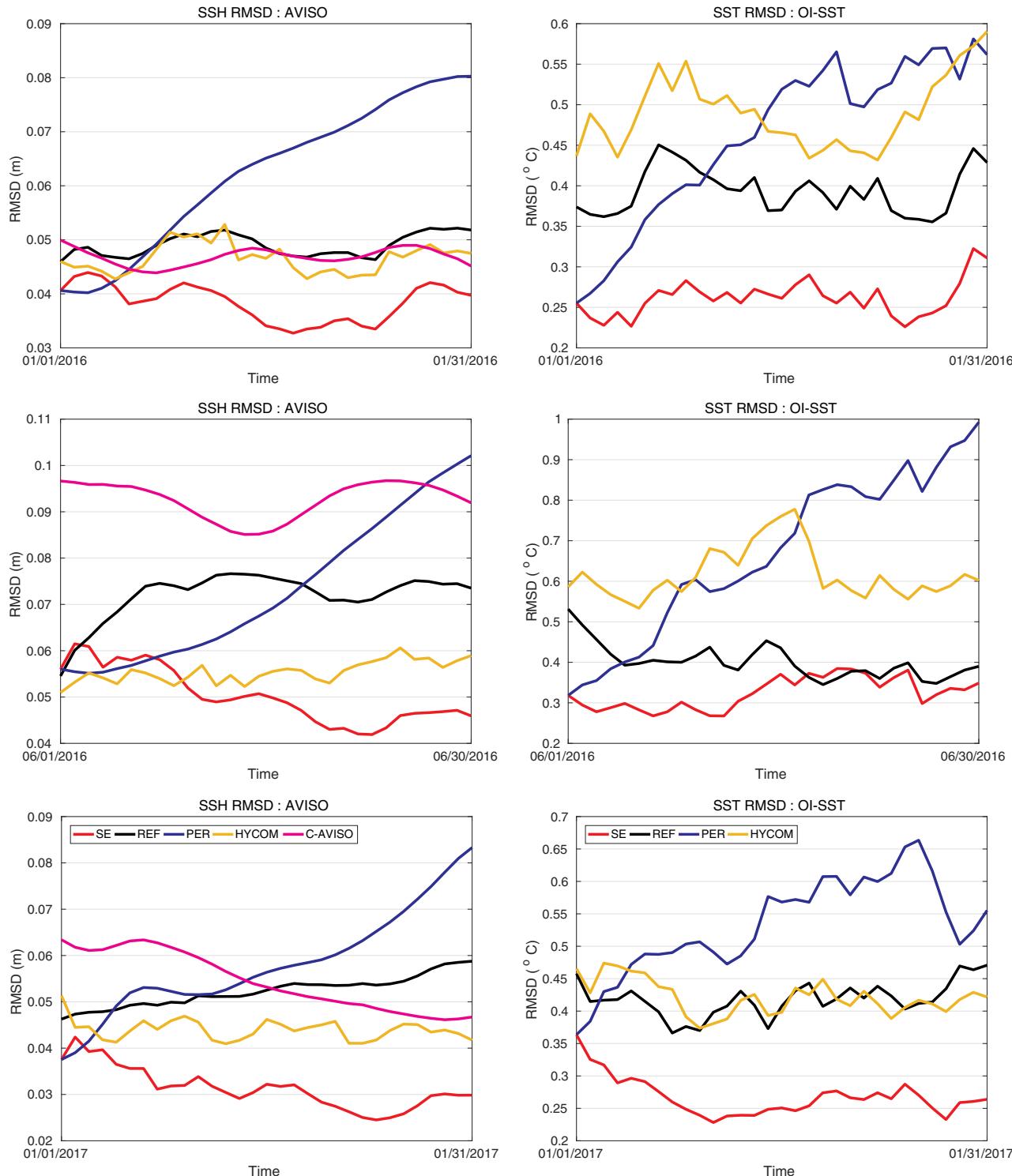


Fig. 10. Hindcast model-data *rmsd* for Sea Surface Height and Sea Surface Temperature computed with respect to AVISO (left column) and OI-SST (right column), respectively for different state estimations experiments. Top panels are for hindcast period: January 1 – 31, 2016, middle panels are for hindcast period: June 1 – 30, 2016, and bottom panels are for hindcast period: January 1 – 31, 2017. The *rmsd* is averaged over the model domain. The description of the curves are the same as Fig. 6. The SSH *rmsd* values are in m and the SST *rmsd* values are in $^{\circ}\text{C}$.

noted that, in this forecast comparison F-HY and F-SE are model solutions without new data assimilation and use realistic forcings, whereas AVISO and HYCOM/NCODA are daily global analyses of the observations. For the first day of the forecast, F-SE SSH shows a close similarity to AVISO, whereas F-HY compares well with HYCOM/NCODA as expected due to its initialization. The F-SE for day-15 is able to reproduce

most of the cyclonic/anticyclonic circulation features in the northern bay and also the cyclonic circulation feature southeast of Sri Lanka in the AVISO analysis. This results in a better comparison of F-SE with AVISO than either HYCOM/NCODA or F-HY for the day-15 forecast. Although F-SE shows a better agreement with AVISO than F-HY at the end of the forecast, HYCOM/NCODA shows better agreement with AVISO than

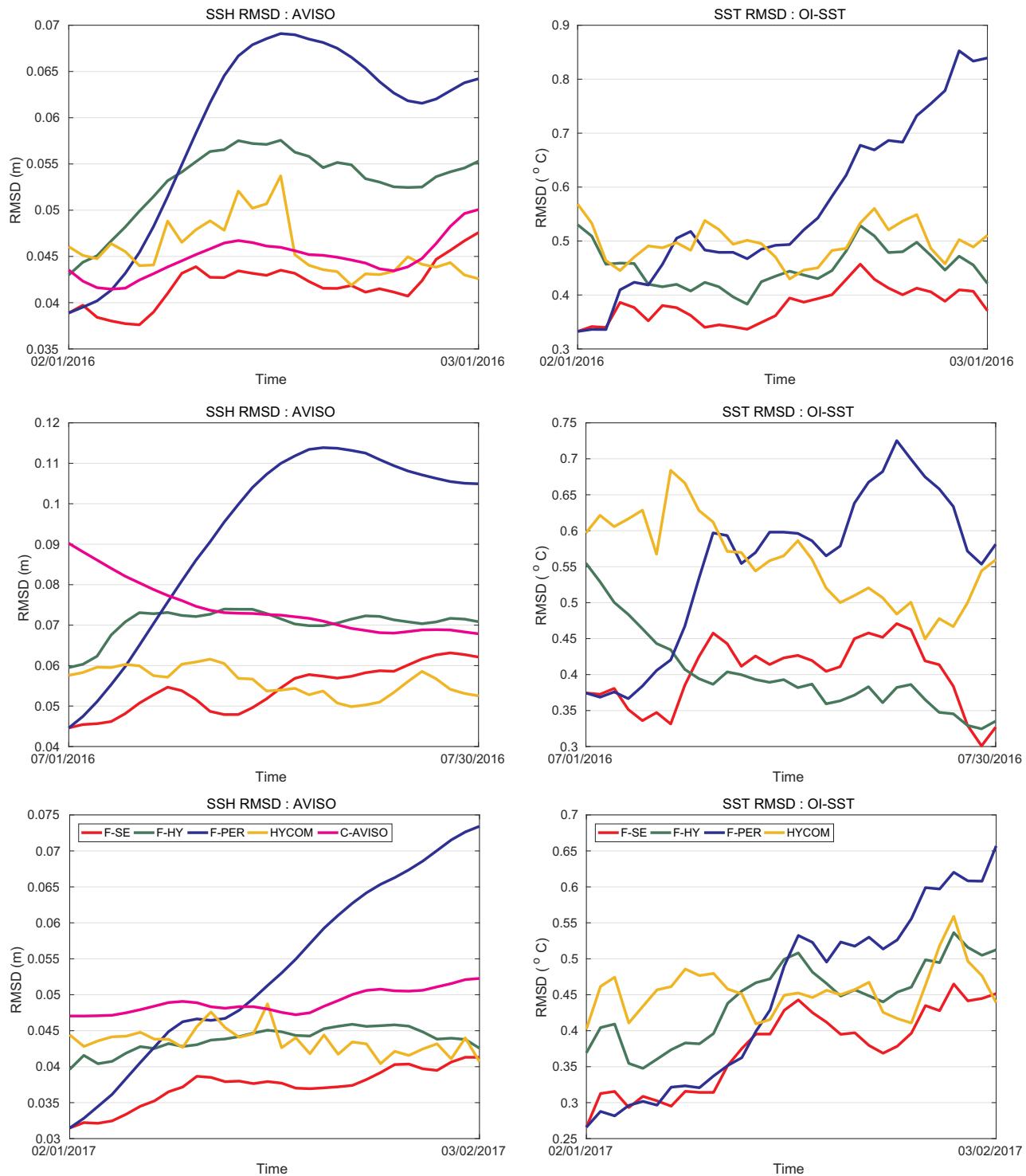


Fig. 11. Forecast model-data $rmsd$ for Sea Surface Height and Sea Surface Temperature computed with respect to AVISO (left column) and OI-SST (right column), respectively for different state estimations experiments. Top panels are for forecast period: February 1 – March 1, 2016, middle panels are for forecast period: July 1 – 30, 2016, and bottom panels are for forecast period: February 1 – March 2, 2017. The $rmsd$ is averaged over the model domain. The description of the curves are the same as Fig. 9. All the forecasts used JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings, and annual climatological run-off fluxes. The SSH $rmsd$ values are in m and the SST $rmsd$ values are in $^{\circ}\text{C}$.

either F-SE or F-HY, especially in the eastern bay. This result is expected because HYCOM/NCODA synthesizes the observations. Similar to the SSH hindcast comparison (Fig. 4), HYCOM/NCODA SSH fields are less smooth compared to F-SE, F-HY, and AVISO.

The SST forecasts from F-SE and F-HY are compared with OI-SST and HYCOM/NCODA daily global analysis for the day-1, day-15, and

day-30 of the July 1 – 30, 2017 forecast period in Fig. 8. For day-1, F-SE shows a reasonable comparison with OI-SST than either F-HY or HYCOM/NCODA, especially the temperatures in the center and southeastern bay, but shows slightly warmer temperatures in the northern and western bay compared to OI-SST. F-HY and HYCOM/NCODA shows similar SST fields for day-1, and are warmer (cooler) than OI-SST in the

southeastern (northwestern) bay. At the middle of the forecast period (day-15), F-HY shows a better comparison with OI-SST than F-SE, while F-SE shows a long-term warmer temperature in the western bay than OI-SST. Although HYCOM/NCODA SST for day-15 compares well with OI-SST in the southeastern bay, it shows generally cooler temperatures in the northern bay compared to OI-SST, and the fields are less smooth compared to other SST fields. F-HY inherits cooler temperature in the northwestern bay (around 17–19° N and 83–90° E) from its HYCOM/NCODA initialization, and the simulated SST using realistic forcings results in a closer comparison with OI-SST than F-SE for day-15 and day-30. While HYCOM/NCODA assimilates real-time observations during the forecast period with a daily assimilation update cycle and shows cooler temperatures in the center and eastern bay than OI-SST. Although F-HY starts with a poor comparison to OI-SST, especially with respect to warmer (cooler) temperatures in the southeastern (northwestern) bay, the model integration using realistic forcings results in a closer comparison with OI-SST than other solutions for the rest of the forecast period.

The model-data *rmsd* for SSH forecasts (F-SE, F-REF, F-HY, F-SE-C, F-REF-C, F-HY-C, and F-PER; keeping the F-SE initial state fixed over the forecast period) are computed with respect to AVISO SSH (Fig. 9a). Again, HYCOM/NCODA daily global analysis SSH and C-AVISO are compared with AVISO for the forecast period. F-SE shows improved forecast with lower *rmsd* (average *rmsd* of 4.5 cm) than data-assimilated HYCOM/NCODA solutions for the first week of the forecast period, and stays close to HYCOM/NCODA for the rest of the forecast period. F-SE also shows improved forecast with lower *rmsd* compared to F-REF, F-HY, and F-PER, throughout the forecast period. F-SE-C, F-REF-C, F-HY-C SSH *rmsd* follows similar trend to that of the F-SE, F-REF, F-HY, and shows slightly higher *rmsd* toward the end of the forecast period. An average *rmsd* of 6 cm is shown by C-AVISO throughout the forecast period.

The model-data SST *rmsd* of model forecasts (F-SE, F-REF, F-HY, F-PER, F-SE-C, F-REF-C, F-HY-C) are computed with respect to OI-SST (Fig. 9b). HYCOM/NCODA daily global analysis SST are also compared with OI-SST for the forecast period. HYCOM/NCODA shows an average *rmsd* of about 0.55 °C, while F-SE shows an average *rmsd* of about 0.4 °C, with a peak during the third week of the forecast. F-HY starts off with a higher *rmsd* of about 0.6 °C, which gradually decreases over the first week of the forecast period and shows the lowest *rmsd* of about 0.3 °C of all solutions for the rest of the forecast period. F-REF *rmsd* is higher than F-SE throughout the forecast period, but lower than HYCOM/NCODA for most of the times, except during the third week of the forecast.

F-SE-C, F-REF-C, F-HY-C SST *rmsd* follows a similar trend to that of F-SE, F-REF, F-HY, but shows larger differences toward the end of the forecast. The differences between the (F-SE, F-REF, F-HY) and (F-SE-C, F-REF-C, F-HY-C) is larger for SST *rmsd* when compared to SSH. This is expected due to the dominance of the atmospheric forcing on the SST evolution. Also, a reanalysis atmospheric forcing is expected to simulate a better SST state than using its monthly climatology. The OI-SST showed a cold front in the northwestern bay during the third week (July 15–20, 2017) of the forecast (not shown). SST *rmsd* peaks during the third week of the forecasts of F-SE, F-SE-C, F-REF, F-REF-C, and F-PER, because they all failed to capture the cold OI-SST front, perhaps because their initial states did not have this cold temperature anomaly. F-HY captures the cold temperature anomaly from its HYCOM/NCODA initialization, and the model integration using realistic forcings produces better SST than other solutions in comparison with OI-SST, especially during the third week of the forecast. At the same time, F-HY-C results in a similar SST *rmsd* peak during the third week as in other forecasts, likely due to its monthly climatological forcings. These spatial gradients of SST, especially in the northwestern BoB, are important in generating deep atmospheric convection over the BoB, which are responsible for the accurate simulation of the low-pressure systems triggering Indian summer monsoon and MISO (Samanta et al.,

2018).

The model forecasts are also compared with Argo temperature and salinity data (Fig. 9c,d). For the forecast comparison, there are about 250 Argo profiles in the BoB, and F-SE shows reasonable improvement over F-HY in the surface layers (upper 50 m) for temperature. For salinity, F-SE shows improved forecasts compared to F-HY in the upper 100 m. Overall, F-SE *rmsd* shows improvement over F-HY, when averaged over fewer, sparsely located profiles in the BoB. The forecast performance might vary with respect to individual profiles, but here we presented a domain-averaged *rmsd* metric for clarity.

5. Summary and conclusions

A regional ocean state estimate and forecast has been produced for the BoB region by implementing MITgcm-4DVAR assimilation system. The MITgcm-BoB state estimates can be used for analysis and prediction of the ocean state and to force coupled ocean-atmosphere modeling and prediction to study the regional air-sea interaction processes at multiple scales. This work is part of the large observational campaign of the ONR's DRI: Monsoon Intra-Seasonal Oscillations in the tropical Indian Ocean and the Bay of Bengal (MISO-BoB) to study the coupled ocean-atmosphere processes in the BoB and the tropical Indian Ocean.

As a first step, we set-up a regional ocean model for the BoB using the MITgcm, and produced nine-year (2009–2017) model solutions using realistic initial conditions, atmospheric forcings, and open-ocean boundary conditions. The solutions reproduced the large-scale features of mean sea surface circulation and variability for both SSH and SST in comparison with AVISO SSH and OI-SST analysis. An adjoint-based 4DVAR assimilation system has been implemented for the BoB region, and produced a one-month ocean state estimate (June 1–30, 2017) and forecast (July 1–30, 2017) by assimilating satellite-derived along-track SSH, separated into time mean and anomalies, and gridded SST.

The state estimates are evaluated by comparing to AVISO SSH, OI-SST, and Argo temperature and salinity observations and by forecasting the ocean state for a month from the end of the hindcast. The forecast is intended to cross-validate the state estimate against independent future observations. Although AVISO analysis uses satellite along-track data, the state estimate fit along-track SSH separated into temporal mean and anomalies, and the comparison of estimates with AVISO absolute SSH can be considered as a somewhat independent validation of the estimate. Whereas comparison of estimates with assimilated OI-SST data implies a dependent validation of the estimate. Argo data are not assimilated and are used for independent validation of hindcasts and forecasts.

SE SSH and SST are able to reproduce most of the surface circulation features as seen in AVISO SSH and OI-SST, and showed lowest *rmsd* when compared to REF, PER, and HYCOM/NCODA solutions. SE also showed significant improvement over PER, for both SSH and SST. SE showed greater improvement over REF for SST than for SSH. The cost descent of individual observations (Table 2) shows that SST cost is the dominant contribution to the total cost, and the state estimation reduces the SST cost by 67% and the total cost by 50%. Although REF is simulated using HYCOM/NCODA initialization without any spin-up time, it provides a good estimate of the initial ocean state, comparable or even better than HYCOM/NCODA daily global analysis. Thus, the total cost reduction of 50% of SE with respect to the REF can be considered significant. This is an important result that highlights the MITgcm-BoB model compatibility to HYCOM/NCODA solutions and validates the underlying model physics. HYCOM/NCODA SST shows larger differences from OI-SST (Figs. 5 and 8) with an average *rmsd* of about 0.55 °C over both the hindcast and forecast period (Figs. 6b and 9b). This results in larger departure of REF SST from the assimilated OI-SST and contributes to higher SST cost. Contrarily, HYCOM/NCODA provides a better initialization for REF with respect to SSH, and the iterative optimization further improves the REF solution for both SSH and SST.

F-SE SSH showed significant improvement over HYCOM/NCODA

daily global analysis for the first week of the forecast period, and are comparable to HYCOM/NCODA for the rest of the forecast period. F-SE SSH also showed significant improvement over F-PER, F-REF, and F-HY. F-SE-C, F-REF-C, and F-HY-C SSH showed only a slight increase in *rmsd* when compared to F-SE, F-REF, and F-HY SSH *rmsd*, at least for this model set-up. F-SE SST showed considerable improvement over HYCOM/NCODA analysis, F-REF, and F-PER. Whereas F-HY SST starts off with high *rmsd* error, which decreases over the first week of the forecast period, and results in a better forecast (lowest *rmsd*) than other forecasts for the rest of the forecast period. This improved performance of F-HY is due to the presence of cold SST front in the northwestern bay from its HYCOM/NCODA initialization and through the reanalysis forcings. F-SE-C, F-REF-C, and F-HY-C SST show higher *rmsd* toward the end of the forecast period. SST is controlled primarily by atmospheric forcing and upper-ocean mixed-layer processes. Due to the dominance of atmospheric forcing on the SST evolution, monthly climatological atmospheric fluxes and winds that deviate from the reanalysis forcings might result in larger *rmsd* with respect to OI-SST.

Comparison of model hindcasts and forecasts with independent Argo temperature and salinity show improvements of three-dimensional ocean state, based on domain-averaged *rmsd* metric against fewer, sparsely located profiling floats in the BoB. SE show reasonable improvement in the surface layers for both temperature and salinity, and in the depths (50 – 200 m) for temperature. F-SE show considerable improvements for salinity in the depths (0 – 100 m), and for temperature in the depths (0 – 50 m), when averaged over the sampled locations. The hindcast/forecast cross-validation with Argo data demonstrates that the adjustments to three-dimensional temperature and salinity fields through assimilation of SSH and SST are in agreement with the observations.

The MITgcm-BoB 4DVAR assimilation system are applied to three other periods to produce state estimates and forecasts of the BoB circulation. The time periods of the three additional hindcast/forecast experiments are as follows: 1) hindcast: January 1 – 31, 2016, forecast: February 1 – March 1, 2016, 2) hindcast: June 1 – 30, 2016, forecast: July 1 – 30, 2016, and 3) hindcast: January 1 – 31, 2017, forecast: February 1 – March 2, 2017. All these experiments followed identical settings as that of the June 1 – 30, 2017 experiment, assimilated the same type of observations of satellite-derived SSH and SST over a period of one month and produced 30-day forecasts from the end of the hindcast period using JRA-55 atmosphere reanalysis and HYCOM/NCODA ocean analysis forcings. The results are analyzed following a similar procedure to that used in the June 1 – 30, 2017 experiment, and the hindcasts and forecasts are cross-validated against AVISO SSH and OI-SST data for SSH and SST, respectively. The model-data *rmsd* curves for hindcast (Fig. 10) and forecast (Fig. 11) show similar performance as that of the June 1 – 30, 2017 experiment, producing improved SSH and SST hindcasts and forecasts. This suggests that the MITgcm-BoB 4DVAR assimilation system is robust in producing skillful hindcasts and forecasts and can be effectively used for coupled ocean-atmosphere modeling and prediction studies in the BoB region.

Overall, the MITgcm-BoB model and assimilation system produced useful hindcasts and forecasts for regional ocean analysis and prediction. The MITgcm-BoB 4DVAR system will be used to assimilate observations from global ocean observing system (such as satellite SSH and SST, Argo) and in situ observations from ASIRI and MISO-BoB observation campaigns in the future work. The optimized states will be used to force coupled ocean-atmosphere models, and high-resolution nested models targeting upper-ocean process studies and seasonal predictions.

Declaration of competing interest

None.

Acknowledgments

We gratefully acknowledge the ECCO consortium, including MIT, JPL, and the University of Hamburg. The MITgcm code used in this study is checkpoint 64Y, and was obtained from <http://mitgcm.org/>. The Ssalto/Duacs altimeter product AVISO is produced and distributed by the Copernicus Marine and Environment Monitoring Service (CMEMS) (<http://marine.copernicus.eu/>). The HYCOM/NCODA 1/12° global analysis were obtained from the HYCOM consortium (<http://hycom.org/dataserver/>). The along-track altimetry data were obtained from the Radar Altimetry Database System (RADS), (<http://rads.tudelft.nl/rads/index.shtml>). The SST data were obtained from the Remote Sensing Systems Inc. (<http://www.remss.com/>). The JRA-55 reanalysis atmospheric forcings were obtained from the https://jra.kishou.go.jp/JRA-55/index_en.html. The MITgcm-BoB 4DVAR state estimates, input files including observations and error fields, are available from the authors upon request. The authors would like to thank Bruce Cornuelle and Matthew Mazloff for helpfull discussions. This work is supported by the United States Office of Naval Research MISO-BOB Departmental Research Initiative. GG, ACS, and AJM are grateful for support by ONR, United States (N00014-17-S-B001). Finally, the authors wish to acknowledge three anonymous reviewers for their constructive remarks.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dsrr.2019.104721>.

References

- Andersen, O., Knudsen, P., 2009. DNSC08 mean sea surface and mean dynamic topography models. *J. Geophys. Res.* 114 (C11), C11001.
- Bessafi, M., Wheeler, M.C., 2006. Modulation of south Indian Ocean tropical cyclones by the Madden-Julian oscillation and convectively coupled equatorial waves. *Mon. Weather Rev.* 134 (2), 638–656.
- Chassignet, E., Hurlburt, H., Smedstad, O., Halliwell, G., Hogan, P., Wallcraft, A., Baraille, R., Bleck, R., 2007. The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system. *J. Mar. Syst.* 65 (1-4), 60–83.
- Cummings, J.A., 2005. Operational multivariate ocean data assimilation. *Q. J. R. Meteorol. Soc.* 131 (613), 3583–3604.
- Ducet, N., Le Traon, P.-Y., Reverdin, G., 2000. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.: Oceans* 105 (C8), 19477–19498.
- Edwards, C.A., Moore, A.M., Hoteit, I., Cornuelle, B.D., 2015. Regional ocean data assimilation. *Annu. Rev. Mar. Sci.* 7, 21–42.
- Fekete, B.M., Vörösmarty, C.J., Grabs, W., 2002. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Glob. Biogeochem. Cycles* 16 (3), 15-1.
- Forget, G., 2010. Mapping ocean observations in a dynamical framework: a 2004–06 ocean atlas. *J. Phys. Oceanogr.* 40 (6), 1201–1221.
- Forget, G., Campin, J.-M., Heimbach, P., Hill, C., Ponte, R., Wunsch, C., 2015. ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev. Discuss. (GMDD)* 8, 3653–3743.
- Fukumori, I., Lee, T., Cheng, B., Menemenlis, D., 2004. The origin, pathway, and destination of Nino-3 water estimated by a simulated passive tracer and its adjoint. *J. Phys. Oceanogr.* 34, 582.
- Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. In: *Advances in Geophysics*, vol. 33. Elsevier, pp. 141–266.
- Giering, R., Kaminski, T., 1998. Recipes for adjoint code construction. *ACM Trans. Math. Software* 24 (4), 437–474.
- Gilbert, J.C., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Program.* 45 (1-3), 407–435.
- Gopalakrishnan, G., Cornuelle, B.D., Hoteit, I., 2013a. Adjoint sensitivity studies of loop current and eddy shedding in the Gulf of Mexico. *J. Geophys. Res.: Oceans* 118 (7), 3315–3335.
- Gopalakrishnan, G., Cornuelle, B.D., Hoteit, I., Rudnick, D.L., Owens, W.B., 2013b. State estimates and forecasts of the loop current in the Gulf of Mexico using the MITgcm and its adjoint. *J. Geophys. Res.: Oceans* 118 (7), 3292–3314.
- Gopalakrishnan, G., Hoteit, I., Cornuelle, B.D., Rudnick, D.L., 2019. Comparison of 4DVAR and EnKF state estimates and forecasts in the Gulf of Mexico. *Q. J. R. Meteorol. Soc.* 145, 1354–1376.
- Goswami, B., Ajayamohan, R., Xavier, P.K., Sengupta, D., 2003. Clustering of synoptic activity by Indian summer monsoon intraseasonal oscillations. *Geophys. Res. Lett.* 30 (8).
- Heimbach, P., Hill, C., Giering, R., 2002. Automatic generation of efficient adjoint code for a parallel Navier-Stokes solver. In: Dongarra, J.J., Sloot, P.M.A., Tan, C.J.K.

- (Eds.), Lecture Notes in Computer Science (LNCS), vol. 230. Springer-Verlag, pp. 1019–1028. Part II.
- Hendon, H.H., Liebmann, B., 1990. The intraseasonal (30–50 day) oscillation of the Australian summer monsoon. *J. Atmos. Sci.* 47 (24), 2909–2924.
- Hoteit, I., Cornuelle, B.D., Heimbach, P., 2010. An eddy-permitting, dynamically consistent adjoint-based assimilation system for the Tropical Pacific: hindcast experiments in 2000. *J. Geophys. Res.* 115 (C3), C03001.
- Hoteit, I., Cornuelle, B., Kim, S.Y., Forget, G., Köhl, A., Terrill, E., 2009. Assessing 4D-VAR for dynamical mapping of coastal high-frequency radar in San Diego. *Dyn. Atmos. Oceans* 48 (1–3), 175–197.
- Hoteit, I., Cornuelle, B.D., Köhl, A., Stammer, D., 2005. Treating strong adjoint sensitivities in tropical eddy-permitting variational data assimilation. *Q. J. R. Meteorol. Soc.* 131 (613), 3659–3682.
- Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B.D., Köhl, A., Heimbach, P., 2013. A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. *Dyn. Atmos. Oceans* 63, 1–23.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* 77 (3), 437–472.
- Kim, J.-H., Ho, C.-H., Kim, H.-S., Sui, C.-H., Park, S.K., 2008. Systematic variation of summertime tropical cyclone activity in the western North Pacific in relation to the Madden-Julian oscillation. *J. Clim.* 21 (6), 1171–1191.
- Köhl, A., Stammer, D., Cornuelle, B.D., 2007. Interannual to decadal changes in the ECCO global synthesis. *J. Phys. Oceanogr.* 37 (2), 313–337.
- Large, W., McWilliams, J., Doney, S., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* 32 (4), 363–403.
- Large, W.G., Pond, S., 1981. Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.* 11 (3), 324–336.
- Lau, K.-M., Chan, P.H., 1988. Intraseasonal and interannual variations of tropical convection: a possible link between the 40–50 day oscillation and ENSO? *J. Atmos. Sci.* 45 (3), 506–521.
- Le Dimet, F., Talagrand, O., 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* 38 (2), 97–110.
- Le Traon, P., Nadal, F., Dufet, N., 1998. An improved mapping method of multi-satellite altimeter data. *J. Atmos. Ocean. Technol.* 15 (2), 522–534.
- Lucas, A.J., Shroyer, E.L., Wijesekera, H.W., Fernando, H.J.S., D'Asaro, E., Ravichandran, M., Jinadasa, S.U.P., MacKinnon, J.A., Nash, J.D., Sharma, R., Centurioni, L., 2014. Mixing to monsoons: air-Sea Interactions in the Bay of Bengal. *Eos Trans. Am. Geophys. Union* 95 (30), 269–270.
- Madden, R.A., 1986. Seasonal variations of the 40–50 day oscillation in the tropics. *J. Atmos. Sci.* 43 (24), 3138–3158.
- Marshall, J., Adcroft, A., Hill, C., Perelman, L., Heisey, C., 1997. A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.* 102 (C3), 5753–5766.
- Mazloff, M.R., Heimbach, P., Wunsch, C., 2010. An eddy-permitting southern ocean state estimate. *J. Phys. Oceanogr.* 40, 880–899.
- Mazloff, M.R., Gille, S.T., Cornuelle, B.D., 2014. Improving the geoid: combining altimetry and mean dynamic topography in the California coastal ocean. *Geophys. Res. Lett.* 41 (24), 8944–8952.
- Menemenlis, D., Fukumori, I., Lee, T., 2005. Using green's functions to calibrate an Ocean general circulation model. *Mon. Weather Rev.* 133 (5), 1224–1240.
- Murphy, A.H., 1992. Climatology, persistence, and their linear combination as standards of reference in skill scores. *Weather Forecast.* 7 (4), 692–698.
- Neena, J., Goswami, B., 2010. Extension of potential predictability of Indian summer monsoon dry and wet spells in recent decades. *Q. J. R. Meteorol. Soc.* 136 (648), 583–592. A journal of the atmospheric sciences, applied meteorology and physical oceanography.
- Oh, J.-H., Kim, B.-M., Kim, K.-Y., Song, H.-J., Lim, G.-H., 2013. The impact of the diurnal cycle on the MJO over the Maritime Continent: a modeling study assimilating TRMM rain rate into global analysis. *Clim. Dyn.* 40 (3–4), 893–911.
- Pavlis, N.K., Holmes, S.A., Kenyon, S.C., Factor, J.K., 2012. The development and evaluation of the Earth gravitational model 2008 (EGM2008). *J. Geophys. Res.: Solid Earth* 178–2012 (B4), 117.
- Roudy, P.E., 2008. Analysis of convectively coupled Kelvin waves in the Indian Ocean MJO. *J. Atmos. Sci.* 65 (4), 1342–1359.
- Samanta, D., Hameed, S.N., Jin, D., Thilakan, V., Ganai, M., Rao, S.A., Deshpande, M., 2018. Impact of a narrow coastal Bay of Bengal sea surface temperature front on an Indian summer monsoon simulation. *Sci. Rep.* 8 (1), 17694.
- Scharroo, R., Leuliette, E., Lillibridge, J., Byrne, D., Naeije, M., Mitchum, G., 2013. RADIS: consistent multi-mission products. In: *20 Years of Progress in Radar Altimetry*, vol. 710.
- Seo, H., Subramanian, A.C., Miller, A.J., Cavanaugh, N.R., 2014. Coupled impacts of the diurnal cycle of sea surface temperature on the Madden-Julian oscillation. *J. Clim.* 27 (22), 8422–8443.
- Sikka, D., Gadgil, S., 1980. On the maximum cloud zone and the ITCZ over Indian longitudes during the southwest monsoon. *Mon. Weather Rev.* 108 (11), 1840–1853.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C., Marshall, J., 2002. Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res. Ocean.* 107 (C9), 3118.
- Takayabu, Y.N., Iguchi, T., Kachi, M., Shibata, A., Kanzawa, H., 1999. Abrupt termination of the 1997–98 El Niño in response to a Madden-Julian oscillation. *Nature* 402 (6759), 279.
- Tian, B., Waliser, D.E., Fetzer, E.J., 2006. Modulation of the diurnal cycle of tropical deep convective clouds by the MJO. *Geophys. Res. Lett.* 33 (20).
- Verdy, A., Cornuelle, B.D., Mazloff, M.R., Rudnick, D.L., 2017. Estimation of the tropical pacific ocean state 2010–13. *J. Atmos. Ocean. Technol.* 34 (7), 1501–1517.
- Vinayachandran, P., Yamagata, T., 1998. Monsoon response of the sea around Sri Lanka: generation of thermal domes and anticyclonic vortices. *J. Phys. Oceanogr.* 28 (10), 1946–1960.
- Webster, P.J., Magana, V.O., Palmer, T., Shukla, J., Tomas, R., Yanai, M., Yasunari, T., 1998. Monsoons: processes, predictability, and the prospects for prediction. *J. Geophys. Res.: Oceans* 103 (C7), 14451–14510.
- Wijesekera, H.W., Shroyer, E., Tandon, A., Ravichandran, M., Sengupta, D., Jinadasa, S., U.P., Fernando, H.J., Agrawal, N., Arulanathan, K., Bhat, G.S., Baumgartner, M., 2016. ASIRI: an ocean-atmosphere initiative for Bay of Bengal. *Bull. Am. Meteorol. Soc.* 97 (10), 1859–1884.
- Wunsch, C., 1996. The Ocean Circulation Inverse Problem. Cambridge University Press.
- Wunsch, C., Heimbach, P., 2013. Dynamically and kinematically consistent global ocean circulation and ice state estimates. In: *International Geophysics*, vol. 103. Elsevier, pp. 553–579.
- Zaba, K.D., Rudnick, D.L., Cornuelle, B.D., Gopalakrishnan, G., Mazloff, M.R., 2018. Annual and interannual variability in the California current system: comparison of an ocean state estimate with a network of underwater gliders. *J. Phys. Oceanogr.* 48 (12), 2965–2988.