RESEARCH ARTICLE
10.1029/2020MS002149

Key Points:
- The next-generation SPEAR seasonal prediction system uses the recently developed coupled general circulation model at GFDL
- The updated ocean data assimilation system for the MOM6 ocean model produces ocean state estimation and initial conditions for predictions
- The new coupled model and ocean tendency adjustment (OTA) reduce model forecast drift and improve seasonal prediction skill

Supporting Information:
- Supporting Information S1

Correspondence to:
F. Lu,
feiyu.lu@noaa.gov

GFDL's SPEAR Seasonal Prediction System: Initialization and Ocean Tendency Adjustment (OTA) for Coupled Model Predictions

Feiyu Lu[1,2] , Matthew J. Harrison[2] , Anthony Rosati[2,3] , Thomas L. Delworth[2] , Xiaosong Yang[2,3] , William F. Cooke[2,3] , Liwei Jia[2,3] , Colleen McHugh[2,4], Nathaniel C. Johnson[2] , Mitchell Bushuk[2,3] , Yongfei Zhang[1,2], and Alistair Adcroft[1,2]

[1]Department of Geosciences, Princeton University, Princeton, NJ, USA, [2]Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, NJ, USA, [3]University Corporation for Atmospheric Research, Boulder, CO, USA, [4]SAIC, Science Applications International Corporation, Reston, VA, USA

Abstract The next-generation seasonal prediction system is built as part of the Seamless System for Prediction and EArth System Research (SPEAR) at the Geophysical Fluid Dynamics Laboratory (GFDL) of the National Oceanic and Atmospheric Administration (NOAA). SPEAR is an effort to develop a seamless system for prediction and research across time scales. The ensemble-based ocean data assimilation (ODA) system is updated for Modular Ocean Model Version 6 (MOM6), the ocean component of SPEAR. Ocean initial conditions for seasonal predictions, as well as an ocean state estimation, are produced by the MOM6 ODA system in coupled SPEAR models. Initial conditions of the atmosphere, land, and sea ice components for seasonal predictions are constructed through additional nudging experiments in the same coupled SPEAR models. A bias correction scheme called ocean tendency adjustment (OTA) is applied to coupled model seasonal predictions to reduce model drift. OTA applies the climatological temperature and salinity increments obtained from ODA as three-dimensional tendency terms to the MOM6 ocean component of the coupled SPEAR models. Based on preliminary retrospective seasonal forecasts, we demonstrate that OTA reduces model drift—especially sea surface temperature (SST) forecast drift—in coupled model predictions and improves seasonal prediction skill for applications such as El Niño–Southern Oscillation (ENSO).

Plain Language Summary Dynamic seasonal prediction systems employ global climate models to predict climate variations on monthly to seasonal time scales. A new state-of-the-art seasonal prediction system has been developed at the Geophysical Fluid Dynamics Laboratory (GFDL) of the National Oceanic and Atmospheric Administration (NOAA). The prediction models are based on GFDL's new component models participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6). The new seasonal prediction system includes new ways to apply observational information to initialize the model simulations, including an updated data assimilation system for the Modular Ocean Model Version 6 (MOM6). Bias correction is applied to the coupled dynamic models to reduce prediction bias, namely, the gap between model-simulated and real-world climatology. Preliminary seasonal prediction experiments demonstrate reduced errors in the predicted climate state globally, as well as improved prediction skill for applications such as El Niño–Southern Oscillation (ENSO).

1. Introduction

Seasonal predictions with coupled dynamic models have become an essential activity across operational meteorology and climate centers worldwide over the past decade. Some latest examples include the SEAS5 from European Centre for Medium-Range Weather Forecasts (Johnson et al., 2019), the Climate Forecast System Version 2 (CFSv2) from National Centers for Environmental Prediction (NCEP) (Saha et al., 2014), and the Global Seasonal forecast system Version 5 (GloSea5) from the U.K. Met Office (MacLachlan et al., 2015). Real-time experimental seasonal predictions have been conducted at the Geophysical Fluid Dynamics Laboratory (GFDL) since 2015 as part of the North American Multi-Model Ensemble (NMME; Kirtman et al., 2014). The two currently used coupled climate models for these predictions are the GFDL Coupled Model Version 2.1 (CM2.1; Delworth et al., 2006) and the Forecast-oriented

Low Ocean Resolution model (FLOR; Vecchi et al., 2014). FLOR is derived from CM2.5 (Delworth et al., 2012) but uses the lower-resolution ocean and sea ice components based on CM2.1. The CM2.1 and FLOR seasonal predictions are initialized from the Ensemble Coupled Data Assimilation (ECDA) system (Zhang et al., 2007). Developed with CM2.1, ECDA provides a consistent and continuous estimate of the coupled model state and its uncertainty, as well as initialization for numerical climate prediction (Vecchi et al., 2014; Zhang et al., 2009; Zhang & Rosati, 2010). Seasonal predictions with CM2.1 and FLOR were conducted on an experimental basis from approximately 2012 until 2015, and since then as part of the real-time NMME. The participation in NMME seasonal prediction is part of a seamless seasonal to centennial climate prediction and projection effort at GFDL, as the same models are also used for decadal prediction since 2011 (Yang et al., 2013) and centennial scale climate change projections (Held et al., 2010; Knutson et al., 2006; Liu & Curry, 2010; Merlis et al., 2014).

A new seasonal-to-decadal modeling and prediction system called SPEAR (Seamless System for Prediction and EArth System Research) has been developed at GFDL (Delworth et al., 2020). The SPEAR model incorporates a multitude of substantial component model development at GFDL since CM2.1 and FLOR, including the adoption of the new Modular Ocean Model Version 6 (MOM6) ocean and SIS2 sea ice code (Adcroft et al., 2019), an updated dynamical core in the atmosphere (Harris & Lin, 2013), improved atmospheric physics (Zhao et al., 2018b), and a substantially updated land model (Zhao et al., 2018a). The SPEAR model shares many commonalities (Delworth et al., 2020) with the recently developed GFDL CM4 (Held et al., 2019) and ESM 4.1 (Dunne et al., 2020) models, but with configuration and physical parameterization choices made toward physical climate prediction and projection on seasonal to multidecadal time scales.

Model bias presents several issues for the state estimation and prediction of the ocean and the climate system in general. For state estimation, the existence of systematic model bias could negatively impact the performance of the assimilation algorithm and also cause discontinuity in the estimated state when the observation network changes. The latter occurs when the introduction of new observations corrects some model bias that otherwise has persisted and results in a spurious change in the climate mean state and/or variability (Balmaseda et al., 2007; Vidard et al., 2007). For prediction, model bias, particularly the coupled model bias in a dynamic prediction system, directly leads to initialization shock, which occurs when the forecast model adjusts to the observation-constrained initial conditions. Model drift is another persistent issue throughout the history of dynamic seasonal prediction. Once prediction starts, the model tends to drift back to its own time-mean climate, and this drift emerges at different time scales for different climate processes, many of which are relevant for seasonal prediction. For example, the equatorial sea surface temperature (SST) and wind will revert back to its own time-mean climate in a few weeks to months, while the adjustment of the deep ocean could take years to decades (Balmaseda, 2017; Balmaseda & Anderson, 2009; Doblas-Reyes et al., 2011; Kim et al., 2012).

There have been many approaches to reducing the impact of model bias in the context of ocean state estimation and seasonal predictions. For state estimation, online and adaptive bias correction schemes have been developed for ocean reanalysis products (Balmaseda et al., 2007; Chepurin et al., 2010; Dee, 2005; Dee & da Silva, 1998; Zuo et al., 2019). The common idea behind these bias correction methods is to estimate model bias from the model-observation misfits in prerun data assimilation experiments and apply the estimated systematic bias, including both time-averaged and time-varying components, back to the model for subsequent analysis or reanalysis experiments. The estimated model bias could include both time-averaged and time-varying components (Chepurin et al., 2010; Dee, 2005), and the online bias correction could apply either on the model state variables directly or in a multivariate fashion through additional treatment (Balmaseda et al., 2007). Online bias correction methods for dynamic predictions have been explored for the atmosphere component in both numerical weather prediction (NWP) (Danforth et al., 2007; Yang et al., 2008) and coupled subseasonal-to-seasonal (S2S) prediction (Y. Chang et al., 2019). In both cases, the online bias correction successfully reduces systematic model drift but has limited impact on prediction skill. In the former NWP case, the authors argue that forecast bias has less impact on prediction skill in short-term weather prediction, while in the latter S2S case, the authors argue that the short atmospheric predictability limits the impact of reduced bias on prediction skill. However, similar bias correction methods have not been attempted for the ocean component in dynamic seasonal or decadal predictions. For dynamic seasonal predictions, the common practice across most operational products is the a posteriori removal of the lead time and initialization time dependent model drift based on a large set of retrospective forecasts

(reforecasts for short) or hindcasts (Johnson et al., 2019; Kirtman et al., 2014; MacLachlan et al., 2015; Rosati et al., 1997; Stockdale, 1997; Xue et al., 2013). However, this linear bias removal could be insufficient if the errors in initial conditions are nonstationary or if the model drift has nonlinear relationship with the anomalous seasonal climate signals (Balmaseda & Anderson, 2009; Kumar et al., 2012; Tietsche et al., 2020; Xue et al., 2013). Anomaly initialization and flux adjustment (FA) are two other strategies to deal with the model bias and its resulting initialization shock in predictions (Magnusson et al., 2013; Smith et al., 2013). FA has been previously used in one version of the GFDL FLOR models to reduce model bias (FLOR_FA; Vecchi et al., 2014). Specifically, FA makes climatological adjustment to the momentum, enthalpy, and freshwater fluxes from the atmosphere to the ocean to reduce the model's climatological bias in SST and surface wind stress.

The multimodel ensemble approach has been successful in mitigating the impact of bias from individual prediction models, demonstrated by NMME (Kirtman et al., 2014) and the EUROSIP (EUROpean Seasonal to Interannual Prediction) multimodel ensemble system (Palmer et al., 2004; Stockdale et al., 2008). These large multimodel ensemble projects are intended to increase the forecast ensemble size for better probabilistic representation of the forecast uncertainty, while from the ensemble-mean perspective, to reduce forecast error and increase forecast skill by averaging out individual model biases to some extent. However, model bias and forecast error in these multimodel ensembles are not completely uncorrelated, especially considering the limited number of independent component models. Therefore, it is important to continue pursuing the reduction of model bias and forecast error through model development, data assimilation, and other bias correction methods.

In this paper, we introduce a bias correction scheme called ocean tendency adjustment (OTA) for GFDL's new SPEAR prediction system. OTA is a simplified and ensemble filter-based adaptation of previous bias correction schemes (Balmaseda et al., 2007; Chepurin et al., 2010). However, developed in a coupled model framework, OTA is innovative for its application in coupled model dynamic predictions. OTA is made possible by the unprecedented coverage of Argo program (Argo, 2019). Argo, with its continuous near-global coverage and contemporaneous temperature and salinity measurements, could enable a modern ocean data assimilation (ODA) system to constrain the climatology and variability of the global ocean state down to 2 km, or even deeper with statistical and dynamic impact vertically.

OTA is part of SPEAR_ECDA that encapsulates all the data assimilation and initialization efforts for seamless predictions with SPEAR across time scales. SPEAR_ECDA inherits the name from the previous ECDA system (Zhang et al., 2007) and is designed to accompany the new component models and to expand prediction capabilities in SPEAR. This paper will focus on the new coupled initialization scheme for SPEAR seasonal prediction, which includes an updated ODA system based on the MOM6 ocean model, the OTA bias correction scheme, and nudging for atmosphere, land, and sea ice initialization.

In section 2, we describe the SPEAR system with the focus on MOM6 ODA, coupled model initialization, and OTA bias correction method. Section 3 presents the SPEAR_ODA ocean state estimation from SPEAR_ECDA, and section 4 evaluates the SPEAR seasonal prediction skills from preliminary retrospective forecast experiments.

2. SPEAR

This section includes a brief introduction of the SPEAR model and detailed description of SPEAR_ECDA, which currently includes the ODA system for MOM6, the OTA technique, and the initialization of the atmosphere, land, and sea ice components from nudging experiments.

2.1. SPEAR Models and SPEAR_ECDA

The SPEAR model is a newly developed coupled ocean-atmosphere-land-sea ice model at GFDL of the National Oceanic and Atmospheric Administration (NOAA) and serves as the physical foundation for seamless predictions and projections from seasonal to centennial time scales. Comprehensive details of the SPEAR model can be found in Delworth et al. (2020).

The ocean and sea ice components are constructed from the publicly released MOM6 and SIS2 code base (https://github.com/NOAA-GFDL/MOM6; https://github.com/NOAA-GFDL/SIS2). The ocean and sea ice components both use a nominal horizontal resolution of about 1° with meridional refinements to 1/3° in

the Tropics. More specifically from the equator to the poles, the meridional grid spacing is 40 km in the deep tropics (10°S to 10°N) to better represent tropical variability, gradually increases to a maximum of 105 km around 20°S and 20°N, and then gradually decreases to 46 km at 65° latitude in both hemispheres. The zonal grid spacing reduces monotonically from 110 km at the equator to 46 km at 65° latitude. A tripolar grid is used with poles over northern Eurasia, northern Canada, and Antarctica. The ocean model uses a hybrid vertical coordinate with 75 layers, with layer thickness as fine as 2 m near the surface and 30 layers in the top 100 m. The hybrid vertical coordinate is a function of depth in the upper ocean, transitioning to isopycnal layers in the interior. Overall, the physics used in the ocean and sea ice components is similar to OM4, the ocean component of CM4 (Adcroft et al., 2019). The changes in the SPEAR ocean configuration from CM4 are limited to horizontal resolution, lateral viscosity, and submesoscale parameters.

The atmospheric and land components are identical to those in the GFDL AM4.0/LM4.0 global atmosphere and land model (Zhao et al., 2018a, 2018b) with the only exception being the near-infrared isotropic reflectance parameter for cold snow over glacial surfaces. The effect of the parameter change is discussed in Delworth et al. (2020).

Two versions of SPEAR—SPEAR_LO and SPEAR_MED—that differ in their horizontal resolution for the atmosphere and land components are currently used. The atmosphere/land resolution is approximately 100 km for SPEAR_LO and 50 km for SPEAR_MED. The 100 km resolution of SPEAR_LO is identical to that in AM4.0/LM4.0 (Zhao et al., 2018a, 2018b), as well as the GFDL CM4.0 coupled climate model (Held et al., 2019). The physics in the 50 km SPEAR_MED are identical to those in the 100 km SPEAR_LO. SPEAR_MED's 50 km atmosphere and land resolution is geared toward seasonal prediction applications such as regional hydroclimate and tropical storms. The atmospheric component uses 33 vertical levels. Further details of the atmosphere and land components can be found in Zhao et al. (2018a, 2018b). Both SPEAR_LO and SPEAR_MED are coupled to the same 1° MOM6 ocean model as described previously. The simulation characteristics of SPEAR_LO and SPEAR_MED, including time-mean climate, internal variability, and responses to idealized and realistic radiative forcing changes, are presented in Delworth et al. (2020).

SPEAR_ECDA, the comprehensive data assimilation and initialization system within SPEAR, is designed for seamless predictions across model configurations and time scales and to accommodate ongoing model development. The scope of SPEAR_ECDA will include data assimilation capabilities for multiple component models, state estimations of the past climate, initialization schemes for coupled predictions and projections from seasonal-to-decadal time scales, and relevant techniques beneficial to the aforementioned objectives. Currently SPEAR_ECDA includes the ODA system in MOM6 as its backbone (section 2.2), incorporates the OTA bias correction scheme (section 2.5), and initialization of atmosphere, land, and sea ice via nudging (section 2.4). SPEAR_ECDA gets its name from the previous ECDA system at GFDL (Zhang et al., 2007). SPEAR_ECDA also includes coupled data assimilation capability, but it is beyond the scope of this paper and will be explored in future studies.

2.2. ODA System for MOM6

The ODA system at GFDL is fully updated from ECDA, in conjunction with the development of the MOM6 ocean model. Following MOM6 as the forecast model, its ODA system also uses the Flexible Modeling System (FMS; Balaji, 2012) libraries, which provides access to parallelism, data communication, I/O libraries, time management, and interpolation among other common modeling functions. The MOM6 ODA system consists of three parts, namely, the data assimilation (DA) interface to MOM6, the handling of ocean observations, and the assimilation filter computation.

The new DA interface to MOM6 is designed to take advantage of recent developments in both MOM6 and FMS and could facilitate applications more than the current ODA effort at GFDL. The DA interface to MOM6 has been included in MOM6 public release as part of the open development (https://github.com/NOAA-GFDL/MOM6). The DA interface in MOM6 provides a direct way to incorporate observational data into the model and enables fully online DA applications to simultaneously run within MOM6 simulations. The concurrent execution of ensembles is supported in MOM6, which provides the foundation for efficient ensemble-based DA implementation through the DA interface. The ensemble execution of MOM6 is fully generalized to allow for different configurations across the ensemble forecast models, such as physics,

parameters, and even the vertical coordinate. Although this multimodel capability within the FMS and MOM6 codebase is not currently utilized, it is planned for future studies. The functions of object-oriented MOM6 DA interface include setting up DA workflow and data structure, processing observations, redistributing model state variables between forward model and DA domains, remapping between different vertical coordinate systems, and applying ODA/OTA increments. The object-oriented features of the DA interface refer to its data and code structure, which follows the same moderate object-oriented MOM6 codebase (https://github.com/NOAA-GFDL/MOM6). Different from the Object-Oriented Prediction System (OOPS) approach used by Joint Effort for Data assimilation Integration (JEDI) project or the European Centre for Medium-range Weather Forecast (ECMWF), the MOM6 DA interface embeds the DA code within the MOM6 model, rather than embedding the models in a DA framework. Therefore, it does not have the capability to easily incorporate various models like OOPS, but it is optimized for efficiency within MOM6 and FMS and can still support multiple DA methods like OOPS.

The handling of the ocean observations and the assimilation filter computation in the new MOM6 ODA system follow a similar framework as those in the previous ECDA system (Zhang et al., 2007), however, there have been numerous improvements, mostly centered around the efficiency and robustness of DA code, as well as the use of climatological information during DA.

The handling of observational data is fully parallelized in the new MOM6 ODA system, that is, each DA computing core only stores localized observations within its boundaries. This significantly reduces memory usage and speeds up DA initialization time. The observations are also stored in a new object-oriented and linked data structure such that the observations can be easily searched, and the DA parameters can be independently configured for each observation at runtime.

MOM6 ODA keeps the two-step Ensemble Adjustment Kalman Filter (EAKF) from ECDA (Zhang et al., 2007). EAKF is an ensemble-based sequential assimilation scheme that facilitates parallel implementation of online filter computation. When each observation is introduced into the filter, we obtain its ensemble background through bilinear interpolation of the ensemble model forecast at the assimilation time. The first step of EAKF is to calculate the increments from this observation and its ensemble background through the EAKF square root filter (Anderson, 2003), and the second step is to apply the increments to all target model grid points within the localization radius through the covariance between the observation ensemble background and the ensemble model forecast at the target grid point. One major improvement of filter computation in MOM6 ODA from ECDA is the implementation of a geospatial K-d Tree package (https://github.com/JCSDA/geoKdTree), which enables efficient lookup of the closest model grid points to each observation. This implementation significantly improves searching accuracy and efficiency during the second step (increment application through covariance) of EAKF. The K-d Tree package comes from a collaboration with the JEDI project at the Joint Center for Satellite Data Assimilation (JCSDA). Other improvements include online detection of outlier observations, potential to in situ temperature conversion using the international thermodynamic equation of seawater—2010 (TEOS-10, IOC, SCOR, & IAPSO, 2010), and the capability to output observation minus forecast/analysis statistics in the observation space.

Compared to ECDA, MOM6 ODA also removes two ancillary DA features that involve climatological information. First, MOM6 ODA removes climatological background covariances from EAKF. In MOM6 ODA, the background variances and covariances for EAKF are calculated solely from the ensemble model forecasts. The sole dependence of the background covariances on ensemble model forecasts is made possible by the improvement in model simulation and increased ensemble size compared to the previous system. Second, climatological restoring of salinity and the deep ocean temperature is also removed from MOM6 ODA. There are limited subsurface in situ salinity data prior to the introduction of Argo, mainly from moorings and Conductivity Temperature Depth (CTD) observations. However, these salinity data outside of the Argo network are not currently assimilated due to their limited temporal and spatial coverage. The ocean salinity is now only constrained by Argo salinity data, while the deep ocean below 2,000 m is constrained by extending the increments from the deepest level of Argo float profiles all the way down to the bottom of the ocean via vertical covariance projection. Although the typical time scale of ocean dynamics increases with depth, we assume that the observations at 2,000 m would not project too much excessive high-frequency noise into the deep and abyssal ocean. This method currently serves as a stopgap between previous climatological restoring and future increasing deployment of deep Argo floats (Jayne

Table 1
Summary of Assimilation Experiments Discussed in This Paper

Assimilation experiments	ECDA	SPEAR_ODA	SPEAR_LO_nudged (SPEAR_MED_nudged)
Model	CM2.1 (Delworth et al., 2006)	SPEAR_LO (Delworth et al., 2020)	SPEAR_LO (SPEAR_MED) (Delworth et al., 2020)
Ensemble size	12	30	15 (5)
Time period considered	1990 to present (full analysis starts in 1961)	1990 to present	1990 to present
Atmosphere assimilation	EAKF, 6-hourly	None	Nudging, 6-hourly
Atmosphere observations	6-hourly temperature and wind from CFSR	None	6-hourly temperature, wind and humidity from CFSR
Ocean assimilation	EAKF, daily	EAKF, daily	Nudging, daily
Ocean observations	Daily OISST v2[a] and GTSPP profiles (XBT, Argo, CDT, moorings)	Daily OISST v2, Argo, XBT and Tropical moorings[b] (see section 2.3)	Daily OISST v2
OTA	None	Applied through full experiment length	None
Climatological restoring	Temperature and salinity to WOA05[c] below 1,000 m with 1-year time scale	None	None

[a]Weekly OISST v2 (Reynolds et al., 2002) is used before 2012   [b]Only temperature data from the tropical moorings are assimilated   [c]WOA05 (World Ocean atlas 2005) data references (Antonov et al., 2006; Locarini et al., 2006)

et al., 2017). MOM6′s improved simulation of the deep ocean, owing to the new hybrid coordinate as well as other physics and dynamics advances (Adcroft et al., 2019), is also key for the removal of deep ocean restoring.

2.3. Configuration for Ocean State Estimation

A new ocean state estimation using MOM6 ODA within SPEAR_LO is produced for the time period starting 1990 to 2019. This ocean state estimation will be continued every month to facilitate real-time SPEAR seasonal prediction in the future. Ocean initial conditions throughout this new ocean analysis are used for the reforecasts described in sections 2.4 and 4 and will be continued in real time each month for future experimental operational seasonal forecasts with SPEAR. This ocean state estimation will be called SPEAR_ODA. A summary of SPEAR_ODA, along with ECDA ocean state estimation and the SPEAR_nudged experiments described in section 2.4, is shown in Table 1. During the production of the ocean state estimation, the ocean component is constrained by MOM6 ODA, the sea ice component indirectly constrained by SST assimilation, and the atmosphere and land components freely evolving given the ocean and sea ice boundary conditions. The radiative forcing, aerosol forcing, and land surface properties for the estimation period are the same as the SPEAR large-ensemble historical and Shared Socioeconomic Pathway 5–8.5 (SSP5–8.5) projection simulations (Delworth et al., 2020). SPEAR_ODA in its current form is not considered a comprehensive ocean reanalysis product like ECDA (Zhang et al., 2007), ECMWF ORAS5 (Zuo et al., 2018) or SODA3 (Carton et al., 2018). Additional efforts to explore coupled data assimilation approach and minimize atmosphere forcing biases in SPEAR_ECDA are ongoing at GFDL.

Thirty ensemble members are used in this ocean state estimation, an increase from 12 in the previous system (Zhang & Rosati, 2010). Although experiments in the previous system did not show significant improvement from 12 to 24 members, the larger 30-member ensemble is chosen given the improved computational capability and efficiency, as well as the need for a larger prediction ensemble. The initial conditions on 1 January 1990 are obtained by combining the atmosphere/land initial conditions from the SPEAR large ensemble historical simulations (Delworth et al., 2020) on 1 January 1990, and an analyzed ensemble of ocean/sea ice initial conditions during the ARGO era, or 1 January 2009 to be specific. The assimilation is performed daily at 0000 UTC, and this high assimilation frequency is made possible by the online implementation of the MOM6 ODA system. The assimilated observations of SPEAR_ODA include daily Optimum Interpolation Sea Surface Temperature (OISST) v2 from NOAA (Banzon et al., 2016; Reynolds et al., 2007) subsampled to 1° regular grid, daily Global Tropical Moored Buoy Array temperature data from NOAA/Pacific Marine Environmental Laboratory (PMEL) (https://www.pmel.noaa.gov/gtmba/), XBT (eXpendable BathyThermograph) data from the Global Temperature and Salinity Profile Programme (GTSPP) (Sun et al., 2010), and Argo temperature and salinity data (Argo, 2019). Currently, observations from altimeters, tide gauges, drifters, etc. are not yet assimilated. The Argo data were

collected and made freely available by the International Argo Program and the national programs that contribute to it (http://www.Argo.ucsd.edu, http://Argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. More details of XBT data from GTSPP, Argo data, and quality control procedures at GFDL are presented in Chang et al. (2013).

The analysis in SPEAR_ODA is performed on the same horizontal ocean grid as the SPEAR_LO MOM6 ocean component but on a z vertical coordinate instead of the hybrid coordinate used by SPEAR_LO. Both the analysis z coordinate and the MOM6 hybrid coordinate have 75 levels, and their resolution is the same in the upper ocean where the hybrid coordinate is still depth-based. The remapping between the two vertical coordinate systems is handled by the MOM6 ODA interface, where model first-guess column temperature and salinity are regridded from the hybrid vertical grid, which evolves with the model state within each ensemble member, to a common geopotential coordinate system using the MOM6 Arbitrary Lagrangian Eulerian algorithm.

The assimilation parameters for SPEAR_ODA are configured as follows. The observation-error standard deviation at the ocean surface is set at 0.5°C for temperature and 0.05 psu for salinity globally. Both temperature and salinity observation-error standard deviations decrease in depth with an e-folding distance of 2 km. These standard deviations are set manually by taking into considerations the forecast ensemble spread, typical error statistics from the assimilated observations, and experience from prior system development. An observation thinning technique is used for subsurface ocean observations such that no more than three data points are used within each model layer. Data windows are set individually for each observation type, including ±4 days for Argo, ±2 days for XBT, and ±2 h for buoys and SST. The time window for the daily buoy and SST data is chosen such that each daily DA cycle only uses data in that corresponding day. The longer time windows for Argo and XBT result in the same observations being used in multiple daily DA cycles; however, a temporal weighting function based on Gaspari and Cohn (1999) is applied such that the weight for observations at the edges of the time window reduces to 0. The data windows for Argo and XBT observations, separate from and longer than 1-day assimilation cycle of SPEAR_ODA, resembles a sequential Kalman smoother in practice (Evensen & van Leeuwen, 2000). As the name suggests, this implementation is used to smooth out the large corrections to the model state variables from new observations. The use of ensemble covariance provides spatial localization and anisotropy by shaping the remote projection of data assimilation increments through ocean dynamics. Additionally, to limit spurious long-distance ensemble covariance due to the ensemble sampling error, horizontal covariance localization is used for all observations to facilitate the sequential and parallel implementation of ODA using the same Gaspari and Cohn (1999) weighting function. The impact distance is 400 km at the equator and decreases in proportion to the cosine of observation latitude. The decrease of the impact distance stops poleward of 60° in both hemispheres, where the minimum impact distance is always 200 km. Vertical covariance localization is limited to ±6 model levels for subsurface observations and down to about 100 m for SST data, while the SST vertical impact is reduced to only 10 m if the ocean surface is determined to be sea ice covered in either the observation or model forecast. The reason for reduced vertical impact under sea ice is that such SST observations are set at freezing temperatures and have much more limited correlation with subsurface ocean temperature compared to the ice-free ocean mixed layer. The vertical covariance localization is extended to the bottom of the ocean from the bottom level of Argo floats as mentioned in section 2.2. There is an additional treatment to the SST assimilation related to sea ice. Using the accompanying sea ice concentration data from daily OISST v2, the observed times and locations where the sea ice concentration is greater than 25% are deemed sea ice covered, and the corresponding SST observations are replaced by the seawater freezing temperatures based on the model forecasted salinity at those observed times and locations. The newly released daily OISST v2.1 (Banzon et al., 2020) has several improvements upon the v2 used in this study, one of which is the correction of Arctic SST biases. However, v2.1 was not available at the time of this study, and the impact of its change from v2 for SPEAR_ODA will be investigated in the future. This treatment of SST data based on sea ice observations serves as an indirect constraint on the sea ice state estimation, which is an interim approach while direct sea ice data assimilation is being actively developed at GFDL.

2.4. Initialization of Seasonal Prediction

SPEAR_ODA provides the ocean initial conditions for both the SPEAR_LO and SPEAR_MED models, which have the same ocean component. Meanwhile, the initial conditions for the atmosphere, land,

and sea ice components are obtained from separate SPEAR_nudged experiments, as summarized in Table 1. In the SPEAR nudged experiments, atmosphere and SST are nudged toward the NOAA/NCEP Climate Forecast System Reanalysis (CFSR; Saha et al., 2010) and daily OISST v2 (Banzon et al., 2016), respectively. The nudged atmosphere variables include three-dimensional temperature, wind, and humidity. The atmospheric nudging is performed with a 6-hourly e-folding time scale for temperature and wind and a daily time scale for humidity. The nudging experiments are performed in both SPEAR_LO and SPEAR_MED models, and the resulting analyses are called SPEAR_LO_nudged and SPEAR_MED_nudged, respectively. The name SPEAR_nudged is used where it applies to both experiments. In order to achieve the best possible balance among the component models from these separate experiments, the SST observations for SPEAR_nudged undergo the same freezing point adjustment based on salinity and sea ice coverage as those for SPEAR_ODA, such that the nudged atmosphere component has the SST lower boundary as if it comes from SPEAR_ODA. The threshold to determine sea ice coverage and freezing point adjustment for the nudging experiment is set at 30% instead of 25% for SPEAR_ODA. This difference is a byproduct of the developmental nature of the current study. An additional nudging experiment with a 25% threshold does not show any significant change in the sea ice analysis.

The SPEAR_nudged experiments are motivated by the following reasons. First, accurate atmosphere and land initializations have been shown to improve seasonal predictions, especially on the shorter subseasonal side (Jia et al., 2016, 2017; Yang et al., 2018; Zhang et al., 2019), and are routinely implemented in seasonal prediction initializations for operational systems such as those from ECMWF (Johnson et al., 2019) and NOAA (Saha et al., 2014). Second, the analysis and prediction of sea ice, especially summer Arctic sea ice, is another major reason, and sea ice prediction is a crucial component for the new SPEAR seasonal prediction system. Although the sea ice extents from both SPEAR_ODA and SPEAR_nudged have good agreement with observations in terms of their climatology, trends, and interannual variability (see Figures 5 and 6), the sea ice thickness from the SPEAR_ODA ocean state estimation lacks the interannual variability as estimated by ice-ocean reanalysis products (e.g., Zhang & Rothrock, 2003). This is due to the lack of an atmospheric constraint in SPEAR_ODA. Without sufficient in situ ocean observations in the polar regions, the atmospheric constraint can provide both thermodynamic and dynamic controls on sea ice thickness via surface air temperatures and winds. The SPEAR_nudged experiment benefits from these air temperature and wind constraints over ice-covered grid points, thus providing an improved representation of interannual sea ice thickness variations, which have been shown as the crucial source of predictability for Arctic summer sea ice extent (Bonan et al., 2019; Bushuk et al., 2017).

This combination of initial conditions from SPEAR_ODA and SPEAR_nudged will be the primary initialization scheme for the upcoming SPEAR operational seasonal prediction. We have tested this scheme with retrospective seasonal forecasts, or reforecasts for short, in both the SPEAR_MED and SPEAR_LO models. These reforecast experiments are named RF_MED and RF_LO_3, respectively. Both RF_MED and RF_LO_3 use the OTA bias correction technique, which will be described later in section 2.5. Two additional initialization schemes, RF_LO_2 and RF_LO_1, are also tested in the SPEAR_LO model. Specifically, RF_LO_2 is the same as RF_LO_3 except that it uses the initial conditions of all model components from SPEAR_ODA, while RF_LO_1 further removes the use of OTA from RF_LO_2. These additional schemes are designed to test the impact of OTA in forecast mode (between RF_LO_1 and RF_LO_2) and the impact of adding observationally constrained atmosphere and land initial conditions (between RF_LO_2 and RF_LO_3). All reforecast configurations are summarized in Table 2, along with GFDL's current CM2.1 and FLOR seasonal forecasts.

Reforecast experiments are initialized four times per year on the 1st day of January, April, July, and October for the time period of 1995 to 2018. These 1-year reforecasts are done the same way as if they were actual real-time forecasts at their initialization times; therefore, they can be fairly compared to the real-time CM2.1 and FLOR forecasts, and "forecast" and "reforecast" are sometimes used interchangeably. For each configuration, there are 96 reforecast initialization times in total (four per year for 24 years). Fifteen ensemble members, out of the 30 members from SPEAR_ODA, are used for rapid testing and more fair comparison with CM2.1 and FLOR seasonal predictions. The results from the reforecast experiments will be evaluated in section 4.

Table 2
Summary of Seasonal Reforecast Configurations

Reforecast configuration	RF_MED	RF_LO_3	RF_LO_2	RF_LO_1	CM2.1	FLOR
Model	SPEAR_MED (50 km atmos, 1° ocean)	SPEAR_LO (100 km atmos, 1° ocean)			CM2.1 (2° atmosphere, 1° ocean)	FLOR (50 km atmosphere, 1° ocean)
Frequency and time period	4 times (1 January, 1 April, 1 July, 1 October) per year for 1995–2018					
Ensemble size	15				10	12
Forecast length	1 year					
Atmosphere and land IC	SPEAR_MED_nudged (5 independent members, each used 3 times)	SPEAR_LO_nudged (15 independent members)	SPEAR_ODA		ECDA	SST-forced atmosphere-land-only simulations
Sea ice IC	SPEAR_MED_nudged (5 independent members, each used 3 times)	SPEAR_LO_nudged (15 independent members)	SPEAR_ODA		ECDA	
Ocean IC	SPEAR_ODA*				ECDA	
OTA	Yes	Yes	Yes	No	No	No

*RF_LO_2 and RF_LO_1 use different developmental versions of SPEAR_ODA than RF_MED and RF_LO_3

Given the high computational cost of such reforecasts, it is usually not possible to use perfectly designed control experiments to test every change in the initialization system. Therefore, lessons from building previous prediction systems and conclusions from relevant prediction and predictability studies are all considered when building a new prediction system. For example, the similarity in the simulations of SPEAR_LO and SPEAR_MED models provides the opportunity to use much less resources to test the initialization schemes in SPEAR_LO before implementing in SPEAR_MED, hence the multiple reforecast configurations for SPEAR_LO. Another example is that the ocean initial conditions for RF_LO_1 and RF_LO_2 reforecasts are slightly different from those for RF_LO_3 and RF_MED. The changes include the correction of a tripolar model grid issue for OTA implementation in the high-latitude Arctic regions, and the addition of the sea-water freezing point adjustment for SST data mentioned in section 2.3. These differences are primarily limited to the Arctic regions, and do not significantly interfere with the results presented in this paper. The SPEAR sea ice prediction results will be reported in future studies.

2.5. OTA

OTA is the bias correction scheme for the MOM6 ocean component in the coupled SPEAR models. Conceptually similar to some existing online bias correction methods (Balmaseda et al., 2007; Chepurin et al., 2010), OTA estimates the annual cycle of temperature and salinity biases from ODA experiments during the time period with good Argo coverage. More specifically, an ODA experiment is run for the period of 2003–2018 (ODA-Argo hereafter). ODA-Argo has the same setup as the SPEAR_ODA ocean state estimation (section 2.3) except for two differences: ODA_Argo assimilates only Argo (both temperature and salinity) and daily OISST v2 data but not XBT and moorings; ODA_Argo uses 15 ensemble members to facilitate fast testing instead of 30 for SPEAR_ODA. The initial conditions of ODA_Argo is generated in the same fashion as SPEAR_ODA, namely, combining the same analyzed ocean/sea ice initial conditions on 1 January 2009 as SPEAR_ODA and the atmosphere/land initial conditions on 1 January 2003 from the SPEAR_LO ensemble historical simulations. From ODA_Argo, we calculate the OTA increments as the annual-cycle climatology of the three-dimensional ocean temperature and salinity increments during 2007–2018. The first 4 years of ODA-Argo is skipped due to model spin-up and Argo deployment ramp-up.

The OTA increments account for the systematic bias from the atmospheric forcing, the ocean model, and the ODA system. They are the necessary short-term corrections to the ocean state variables to offset their fast error growth, which is on the time scale of the data assimilation cycle. Similar to SPEAR_ODA, ODA_Argo is also performed in the fully coupled SPEAR_LO model where the atmosphere component is only constrained by analyzed SST. This means that the OTA increments can be applied in the forecast model for coupled predictions. Regardless of atmospheric initialization, the atmosphere component in coupled dynamic predictions will revert back to the model climatology in a few weeks at most. Since the OTA

increments are calculated from the ocean error growth under a free atmosphere forced only by analyzed SST, they could mitigate the ocean model drift under such atmospheric forcing beyond the time scale of synoptic predictability. We note that the atmosphere is forced by forecasted SST in the prediction, however, due to the application of OTA, the forecasted SST maintains its climatology from the analysis experiments, therefore both the atmosphere and SST would exert continuously consistent forcing onto each other.

Contemporaneous temperature and salinity observations are critical to maintain the density profile in the ocean, and the only continuous in situ subsurface observations with nearly global coverage for the OTA increments are from the Argo network (Argo, 2019) since early 2000s. Apart from the Argo network, the feasibility of OTA also comes from improvement in the coupled model and computing resources. The improved MOM6 ocean model reduces the misfits to observations, which leads to reduced OTA increments and in turn, better ability to absorb the OTA increments by the model. More computing resources enable larger ensembles of the improved model to calculate the increments more accurately.

Although conceptually similar, the main difference of OTA from established bias correction methods (Balmaseda et al., 2007; Chepurin et al., 2010) is that OTA is formulated in a fully coupled model with a free atmosphere, and therefore can be applied to coupled dynamic predictions. OTA is also simplified compared to some existing bias correction methods in two aspects. First, OTA applies directly on the ocean state variables globally, instead of converting temperature and salinity increments to pressure perturbations in the tropics (Balmaseda et al., 2007, 2013); second, OTA only has an annual cycle component, and lacks the additional short-term flow-dependent component in the estimated bias (Balmaseda et al., 2007; Chepurin et al., 2010). Similar bias correction approaches for the atmosphere component in dynamic predictions (Chang et al., 2019; Danforth et al., 2007; Yang et al., 2008) can successfully reduce systematic model drift, but has limited impact on prediction skill due to less impact of bias on prediction skill in weather prediction and short atmospheric predictability in general. However, when OTA is applied to the ocean component in a coupled seasonal prediction, the bias reduction lasts through the length of the prediction, and there is enough time for the reduced ocean drift to impact climate variability at seasonal time scale. These benefits of OTA to SPEAR seasonal predictions will be presented in section 4.

FA is another bias correction method used in previous GFDL models such as FLOR (Vecchi et al., 2014). The primary difference between OTA and FA is that OTA applies three-dimensional increments based mainly on in situ subsurface ocean observations, while FA applies two-dimensional adjustments on the surface fluxes at the top of the ocean based on reanalysis or observations at the marine surface. As a result, the effect of OTA is instantaneous across the whole ocean column, while FA applies mostly to the surface of the ocean and relies on the model to respond to the FAs.

When used iteratively in an ODA run, the OTA increments would provide the necessary climatological increments while the additional ODA increments could further improve variability in the iterative state estimation. We rerun ODA_Argo with the addition of OTA increments from ODA_Argo itself, and call this iterative experiment OTA_Argo. By design, the increments in OTA Argo would be smaller than their counterparts in ODA_Argo, which is shown by the temperature increments at a few select model diagnostic depths (15, 86, 192, 462, and 985 m) in Figure 1. The scales in Figure 1 decrease with depth to account for smaller increments at depth. Close to the ocean surface at 15 m (Figure 1a), ODA_Argo assimilates global SST data, so the OTA increments have large basin-scale patterns, such as those in the tropical oceans, subtropical gyres and north Atlantic. There are also smaller and more "noisy" spatial structures, especially in regions with strong eddy kinetic energy around boundary currents and in the Southern Ocean. We should note that these OTA increments are counter to the short-term error growth on the time scale of the data assimilation cycle. Therefore, they will be different from the model's climatological bias from control or historical simulations (Delworth et al., 2020). Also, when scaled to degrees per year, the OTA increments are large with some locations reaching a few degrees per year or more. With depth, most of the large-scale increment patterns disappear, while the remaining increments are concentrated in the regions with strong eddy kinetic energy (Figures 1g and 1i), such as the western boundary current regions (Kuroshio and Gulf Stream), where the strong gradients cause small spatial shifts to manifest as large local biases. There are also relatively large increments close to the equator in the mixed layer and around the thermocline (Figures 1c–1h). This is partly caused by considerable coupled model drift in the equatorial oceans. It is also attributed to large dynamic adjustments to ODA temperature and salinity increments due to the
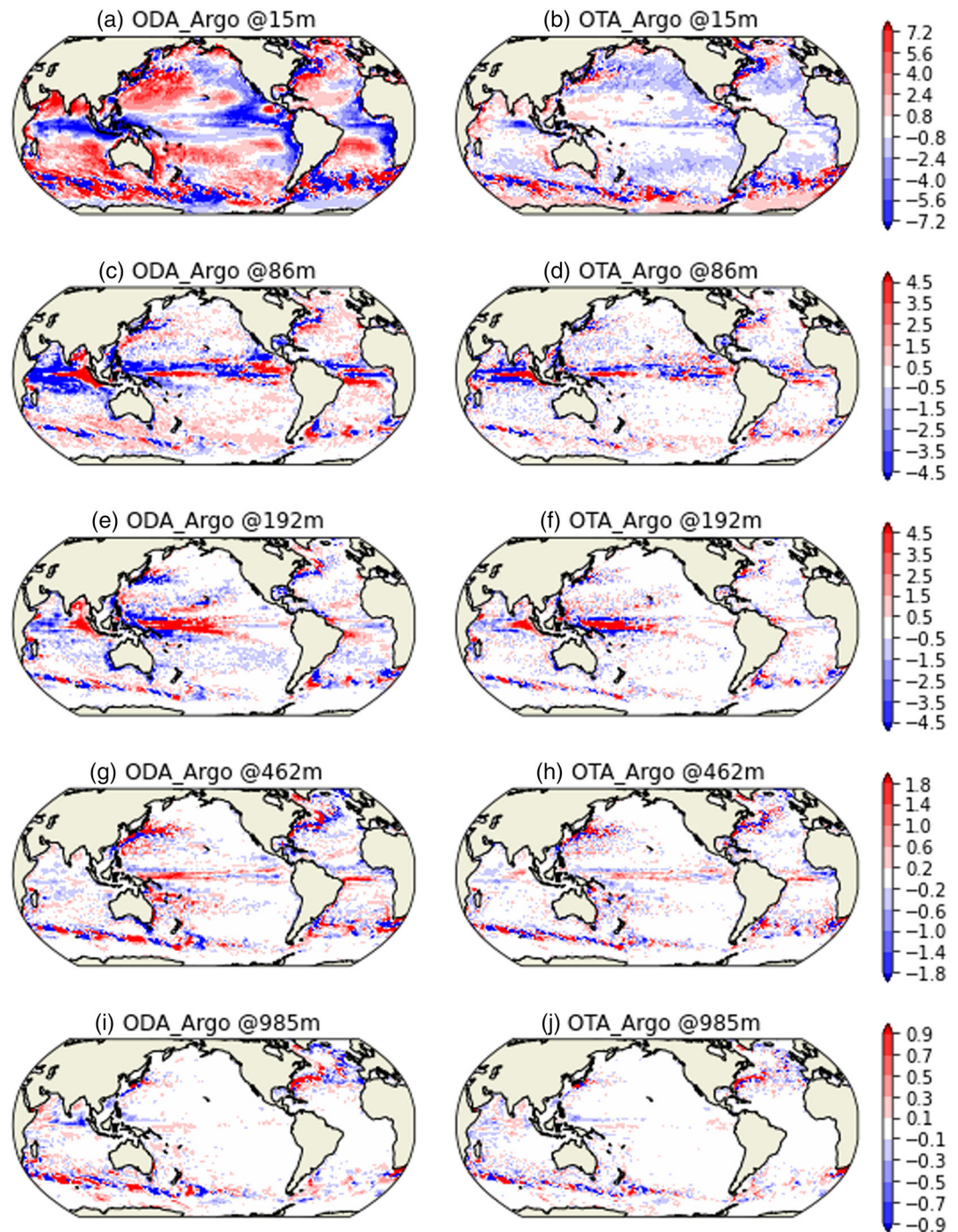
Figure 1. Time-averaged (2007–2018) ODA_Argo (left) and OTA_Argo (right) temperature increments (units: °C/year) at select model depths of 15, 86, 192, 462, and 985 m from top to bottom.

ageostrophic nature of ocean dynamics near the equator. After OTA increments are applied in the iterative OTA_Argo, the remaining increments are shown in the right column of Figure 1. At 15 m (Figure 1b), the magnitudes of the increments are greatly reduced globally, particularly in the tropical and mid-latitude oceans. The remaining large increments are mostly limited to the Southern Ocean, or along the Kuroshio and Gulf Stream, where the climatological OTA increments are less capable to reduce model drift. The contrast between Figures 1a and 1b applies to most of the top 50-meter ocean globally. The OTA_Argo increments are also consistently smaller than the OTA increments throughout the depths in Figure 1.

The OTA increments are subject to the sampling period and assimilated observations. They will change if a different sampling period or different observational dataset is chosen, as also found from the bias correction in ORAS5 (Zuo et al., 2019). Although the OTA increments are calculated from the time period of 2007–2018, they are applied to the full SPEAR_ODA analysis period starting in 1990, as well as the retrospective forecasts starting in 1992. It is inevitable that the OTA increments will include information that pertains to their sampling period, but we think such information does not dominate the OTA increments because the increments in every assimilation cycle only account for the fast error growth (daily for the ocean in our case), which may not be heavily influenced by underlying interannual or decadal variability. Besides, we calculate the OTA increments as a monthly annual cycle from daily increments over 12 years, which should average out interannual variability such as El Niño–Southern Oscillation (ENSO). Overall, we believe the benefits of OTA to SPEAR_ODA and reforecasts prior to Argo would outweigh such concerns, considering that model bias is often comparable to or even larger than the range of observed interannual to decadal climate variability. For example, the climatological forecast biases of the Nino3.4 SST index in both GFDL's current seasonal prediction models (CM2.1 and FLOR) could exceed 2°C over a 12-month period (Figure 15), which resembles the magnitude of strong ENSO events.

In addition to sampling period and assimilated observations, the OTA increments are also subject to the atmosphere forcing, namely, the forcing from the free atmosphere component in our case. SPEAR_MED and SPEAR_LO models share the same ocean component, so their seasonal forecast experiments both use the same ocean initial conditions from SPEAR_ODA. However, due to their different atmospheric resolutions, ODA_Argo experiments are performed in both SPEAR_LO and SPEAR_MED to generate the OTA increments for forecast experiments in each model, respectively. The similar simulation characteristics between SPEAR_LO and SPEAR_MED are again confirmed by the similar increments from their respective ODA_Argo experiment (not shown). The increments shown in Figure 1 are from the SPEAR_LO model.

OTA is currently only performed for a single cycle in SPEAR_ECDA, meaning that the increments from an ODA_Argo is used for all further assimilation or forecast experiments. However, the OTA increments from ODA_Argo will not correct all the systematic model drift, due to various reasons such as un-optimized filter, model error growth on different time scales, limited observational coverage and so on. In some cases, the OTA increments may be much smaller than the actual model drift, such as in high-latitude or coastal regions without consistent Argo coverage. We have tested performing more iterations of the OTA process (e.g., using increments from ODA_Argo and OTA_Argo together for additional iterative experiments) and found limited impact on the quality of the ocean state estimation when and where sufficient observations are available (e.g., top 2,000 m in the Argo era) since ODA is the dominant constraint on the ocean state. However, more iterations of OTA might improve the ocean state estimation prior to and deeper than ARGO, as well as forecast experiments. The more iterative approach will be explored in future studies.

3. Evaluation of Ocean State Estimation

The SPEAR_ODA ocean state estimation (1990–2019) described in section 2.3 is evaluated in this section. Figure 2 shows the global-mean ensemble spread (i.e., standard deviation) of ocean temperature and salinity profiles over the estimation time period, which represent the overall time-varying observational constraint on the ocean temperature and salinity through ODA. Except for surface and near surface temperature which is constrained by continuous SST observations, the ensemble spreads of both ocean temperature and salinity gradually increase from 1990 due to the overall lack of observations, until the introduction of the Argo floats in the early 2000s. The relatively small initial ensemble spread in the 1990s of SPEAR_ODA, despite very limited subsurface in situ observations and no observational constraint on atmosphere forcing, is the result of initializing SPEAR_ODA from an analyzed ensemble with present-day climatology and very small spread due to the ARGO network.

The annual-mean climatological SST bias is shown in Figure 3 for the SPEAR_ODA (top) and ECDA (bottom) ocean state estimations. The bias and errors are based on the difference from the monthly NOAA Optimum Interpolation (OI) SST V2 (Reynolds et al., 2002). Although they belong to the same family of products, the monthly OISST v2 used here for validation is a different product from the daily OISST v2 assimilated in SPEAR_ODA (Reynolds et al., 2007). SPEAR_ODA has smaller MAE, and a
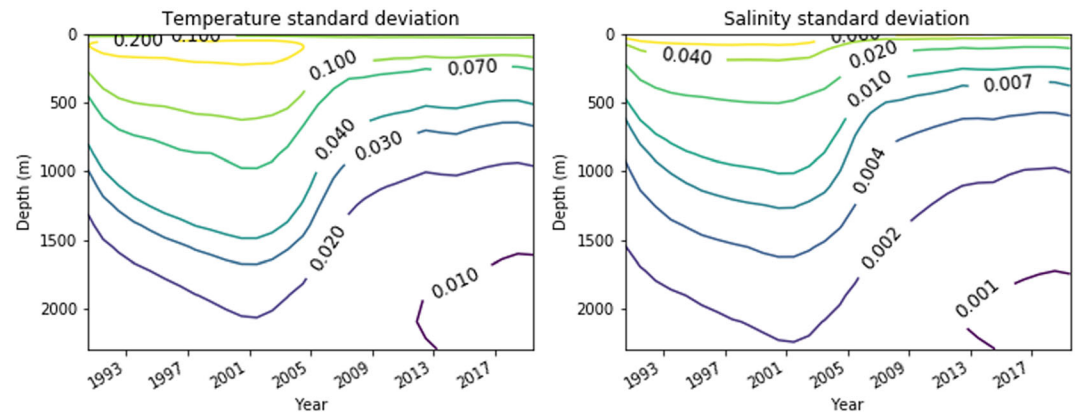
Figure 2. Global-mean ensemble spread profiles of temperature (left, units: °C) and salinity (right, units psu) for the SPEAR_ODA ocean state estimation.

cool SST bias with smaller absolute magnitude. The colder SST is directly related to the colder observations in the daily OISST (Banzon et al., 2016) compared to the monthly OISST (Reynolds et al., 2002) used in ECDA.

The accuracy of SST variability is also improved in SPEAR_ODA compared to ECDA. Figure 4 shows the area-averaged RMSE (root-mean-square error) of monthly SST for three different latitude bands (60°S to 30°S, 30°S to 30°N, and 30°N to 60°N) in SPEAR_ODA and ECDA. The largest improvement comes from the Southern Hemisphere (SH) mid-latitudes (average RMSE from 0.781 to 0.437), which agrees with the large reduction in bias (Figure 3). Tropical SST also exhibits significant reduction in RMSE (0.498 to 0.337), while Northern Hemisphere mid-latitudes have the smallest, but still positive RMSE reduction

Figure 3. Annual-mean climatological SST bias (units: °C) in SPEAR_ODA (top) and ECDA (bottom) for 1990–2019 compared to monthly OISST v2. Area-averaged bias and mean absolute error (MAE) are indicated above each panel. All values have the units of °C.
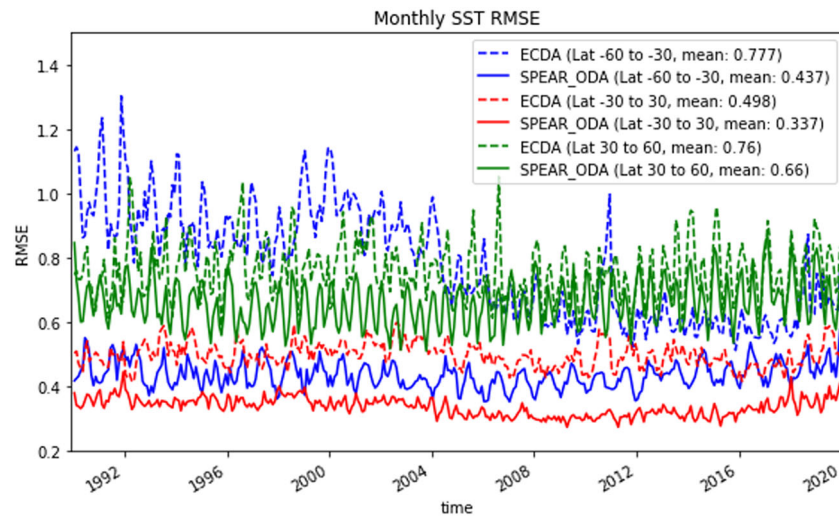
Figure 4. Area-averaged monthly RMSE (units: °C, compared to monthly OISST v2) of SST for three different latitude bands (blue for 60°S to 30°S, red for 30°S to 30°N, and green for 30°N to 60°N) in SPEAR_ODA (solid) and ECDA (dashed).

(0.759 to 0.660). We note that ECDA subsamples the high resolution (0.25°) daily OISST data depending on latitude, while SPEAR_ODA subsamples the data to 1° resolution globally before assimilation. The apparent annual cycle of mid- to high-latitude SST RMSE in Figure 4 is caused by the seasonality in the models' SST simulation bias.

In Figure 5 and Figure 6 we consider the representation of Arctic and Antarctic sea ice in SPEAR_ODA, SPEAR_LO_nudged, and ECDA. SPEAR_LO_nudged is included here since it is providing the sea ice, in addition to atmosphere and land, initial conditions for SPEAR seasonal forecasts. SPEAR_MED_nudged is not included because it is almost identical to SPEAR_LO_nudged, given that both experiments are tightly nudged to the same CFSR and daily OISST v2 data. For both SPEAR_ODA and SPEAR_LO_nudged, sea ice is mainly constrained by SST assimilation and nudging, respectively. They differ in that SPEAR_LO_nudged additionally has observationally constrained atmosphere temperature and winds. Both SPEAR_ODA and SPEAR_LO_nudged accurately capture the observed seasonal cycle of Arctic sea ice extent (SIE), with RMSE values of 0.17×10^6 and 0.13×10^6 km^2, respectively (Figure 5a). For comparison, the median RMSE for Arctic SIE from models participating in the Coupled Model Intercomparison Project phase 5 (CMIP5) is 1.45×10^6 km^2 (Shu et al., 2015). SPEAR_ODA and SPEAR_LO_nudged have larger biases for Antarctic SIE of 1.83×10^6 and 1.67×10^6 km^2, respectively, owing primarily to overly extensive sea ice in austral winter and spring (Figure 5b and Figure 6). These biases are roughly half of the CMIP5 median RMSE of 3.42×10^6 km^2 (Shu et al., 2015). Compared with ECDA, SPEAR_ODA and SPEAR_LO_nudged have notably improved Antarctic SIE biases in austral summer, owing to improvements in the Weddell, Amundsen, Bellingshausen, and Ross Seas (Figure 6).

SPEAR_ODA and SPEAR_LO_nudged also accurately represent the observed interannual variability in September and March Arctic SIE (Figures 5c and 5e). SPEAR_ODA and SPEAR_LO_nudged both correlate strongly with observed September SIE (0.96 and 0.97, respectively) as well as linearly detrended September SIE anomalies (0.92 and 0.94, respectively). These correlations improve upon ECDA, which has full and detrended correlations of 0.93 and 0.76, respectively. The correlations with observed March SIE (0.79 and 0.81, respectively) and detrended March SIE (0.65 and 0.67, respectively) are lower than the September SIE correlations and are slightly lower than ECDA (Figure 5e). SPEAR_ODA and SPEAR_LO_nudged generally improve upon the ECDA representation of Antarctic SIE interannual variability. Both SPEAR runs have notably higher correlations and substantially reduced biases compared to ECDA for March Antarctic SIE (Figure 5d). Conversely, the SPEAR runs have larger biases for September Antarctic SIE, but have higher correlations with the observed data than ECDA (Figure 5f). Despite their improved interannual variability, both SPEAR_ODA and SPEAR_LO_nudged display a decadal-scale September sea ice anomaly in the 1990s
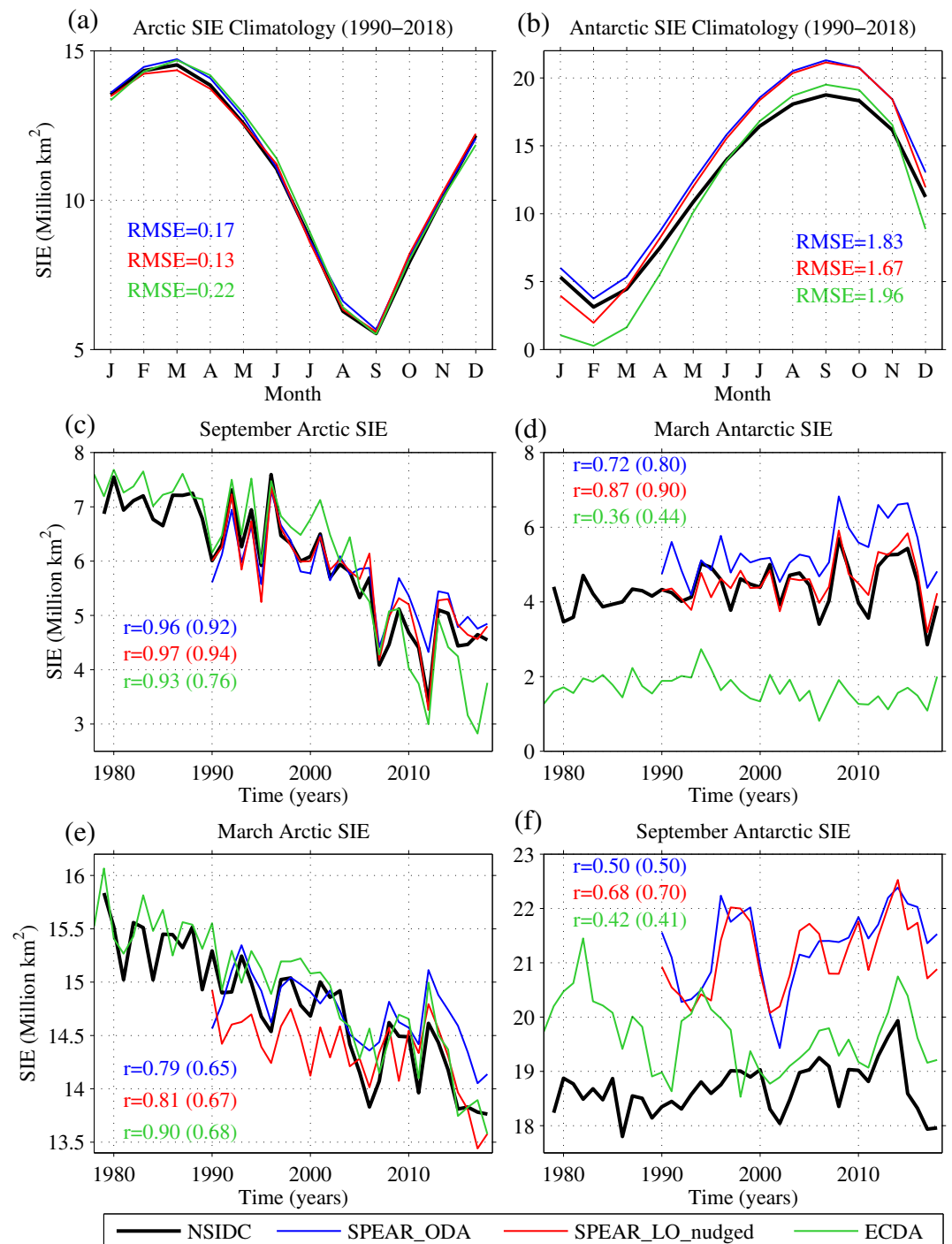
Figure 5. Arctic (a) and Antarctic (b) climatological sea ice extent (SIE) annual cycle from NSIDC (National Sea Ice Data Center) observational data (black), SPEAR ODA ocean state estimation (blue), SPEAR_LO_nudged (red) and ECDA (green). Interannual SIE variability for September Arctic (c), March Antarctic (d), March Arctic (e), and September Antarctic (f). Correlations between observed and simulated SIE are indicated in Panels c–f, with values in parentheses indicating correlations after linearly detrending the time series.

that is not in phase with the observed record. This sea ice anomaly is associated with adjustments in Southern Ocean convective activity that occur over this time period. These convective events occur in the Ross Sea in SPEAR_ODA and the Weddell Sea in SPEAR_LO_nudged and are reflected as winter sea ice concentration biases in these regions (Figure 6). Southern Ocean convection is quite sensitive to the
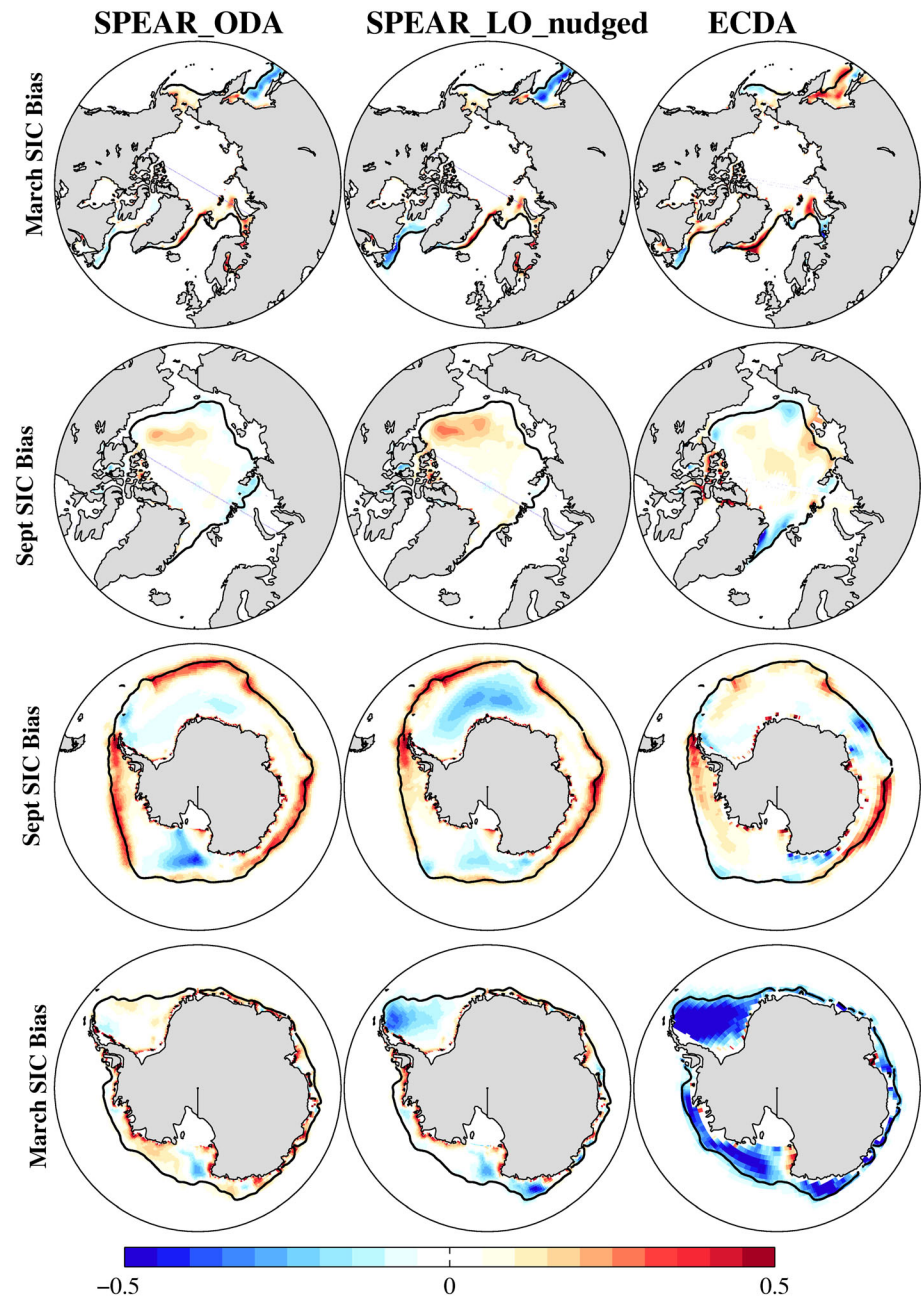
Figure 6. Climatological sea ice concentration (SIC) bias in (left to right) SPEAR_ODA, SPEAR_LO_nudged, and ECDA for (top to bottom) March Arctic, September Arctic, September Antarctic, and March Antarctic regions.

surface energy balance in the SPEAR models (Delworth et al., 2020), consistent with the adjustment that occurs after the constraints are applied in the SPEAR_ODA and SPEAR_LO_nudged runs. Overall, the representation of Arctic and Antarctic sea ice is improved relative to ECDA, and the impact of these improvements on seasonal sea ice predictions will be explored in future work.

Next, we look at the subsurface ocean temperature estimation of SPEAR_ODA compared with ECDA. The spatial differences of climatological ocean temperature (1990–2019) from EN4 objective analysis (Good et al., 2013) for SPEAR_ODA and ECDA at a few depths (100, 270, 540, and 800 m) are shown in Figure 7. The climatological differences are widely reduced for the top 1,000 m throughout most of the analysis period, resulting from improvements in both the MOM6 ocean model and new ODA system.
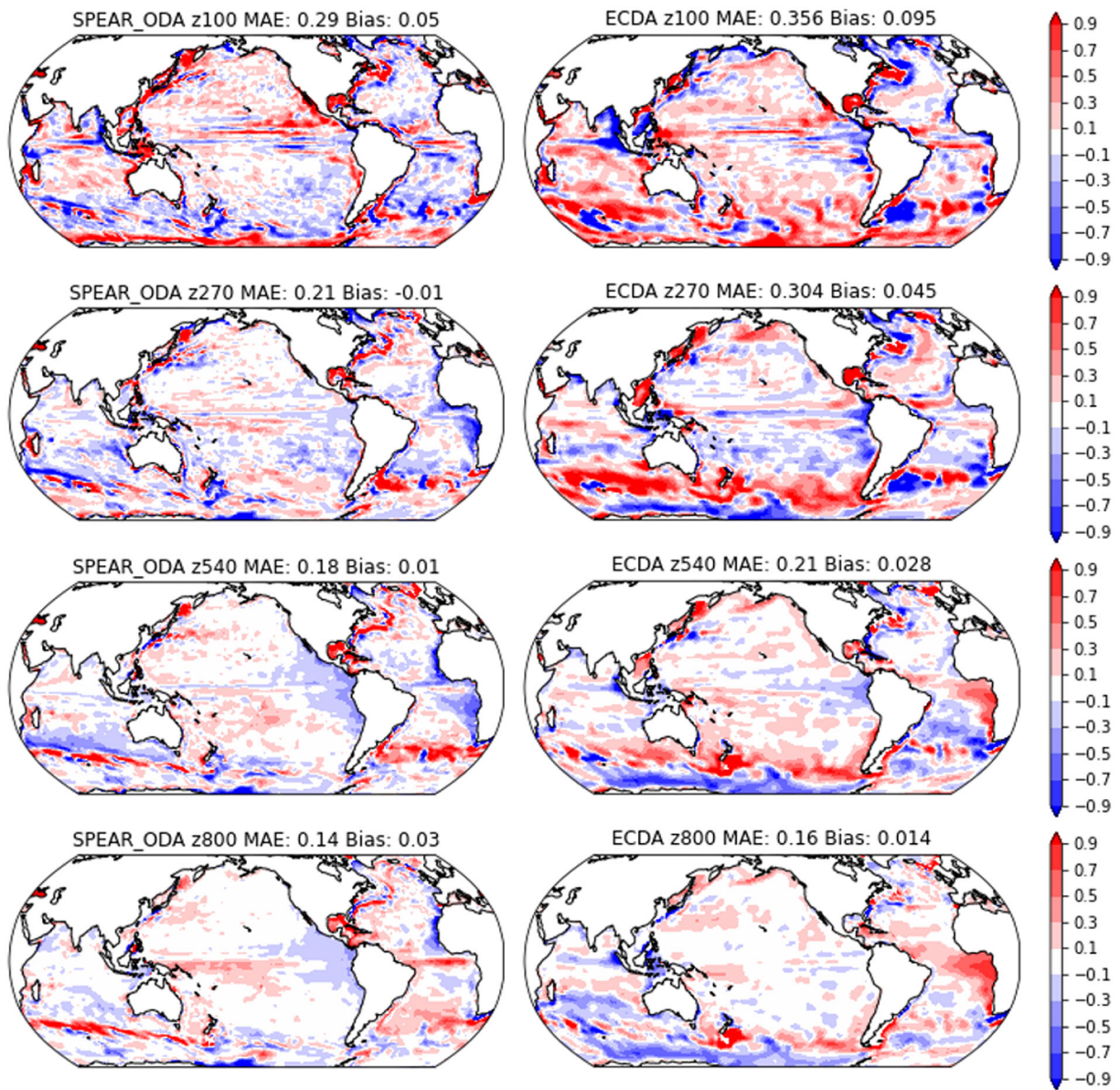
Figure 7. Climatological (1990–2019) annual-mean ocean temperature differences from EN4 objective analysis for SPEAR_ODA (left) and ECDA (right) ocean analyses. Select depths include 100, 270, 540, and 800 m (top to bottom). All values have the units of °C.

The differences in the annual-mean MAE time series between SPEAR_ODA and ECDA ocean analyses are shown in Figure 8. SPEAR_ODA has comparable or smaller errors in the upper 500 m prior to ARGO and upper 1,000 m after ARGO was widely deployed. SPEAR_ODA improves upon ECDA in the upper 500 m of the ocean throughout most of the analysis period, which is most relevant for seasonal prediction purposes. Regarding the larger MAE below 500 m prior to Argo, EN4 objective analysis restores its subsurface ocean to climatology in the absence of any observations (Good et al., 2013), and ECDA also restores the deep ocean to WOA05 climatology (Table 1), which would lead to small deviations from EN4. Meanwhile, SPEAR_ODA depends on OTA and the MOM6 model itself to simulate a realistic deep ocean prior to Ago given the observational constraint above. The smaller MAE in ECDA due to climatological restoring may not necessarily result in better ocean state estimation or prediction from an ocean dynamics perspective, since the large-scale circulation of the model could be distorted in favor of staying close to the climatological estimation, resulting in greater model adjustment in predictions. While in SPEAR_ODA, the deep ocean has more
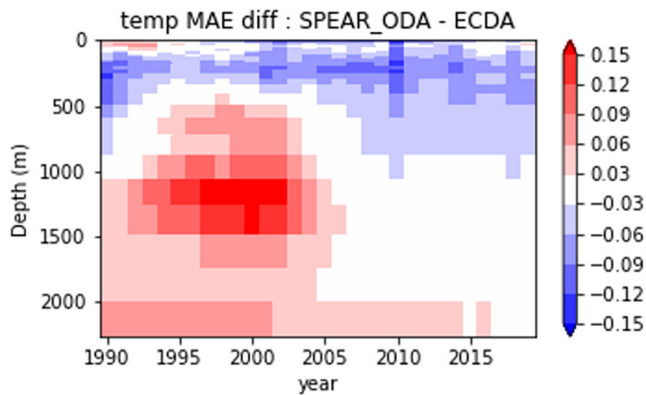
Figure 8. Annual global-mean ocean temperature MAE (relative to EN4 objective analysis) differences between SPEAR_ODA and ECDA ocean state estimation as a function of time and depth. All values have units of °C.

freedom to evolve according to the forcing from above and maintain a circulation structure in balance with the model dynamics.

4. Preliminary Evaluation of Seasonal Reforecasts

We will evaluate the SPEAR seasonal reforecasts in this section. Current NMME forecasts at GFDL based on CM2.1 and FLOR models are compared to the new SPEAR reforecasts. The CM2.1 and FLOR forecasts include the same initialization times (four per year for 1995–2018) as the SPEAR reforecasts. All components of the CM2.1 forecasts are initialized from the weakly coupled ECDA system (S. Zhang et al., 2007; Zhang & Rosati, 2010), while FLOR forecasts are initialized by combining the ocean/sea ice initial conditions from ECDA and atmosphere/land initial conditions from an SST-forced atmosphere-land-only simulations (Vecchi et al., 2014). FLOR_B01, one of the two similar FLOR model versions (A06 and B01), is chosen for comparison here. CM2.1 and FLOR forecasts include 10 and 12 members, respectively, which are less than the 15 members here in SPEAR_LO and SPEAR_MED. However, this small difference in ensemble size should not significantly affect our assessment based on ensemble-mean statistics. More details on the CM2.1 and FLOR seasonal predictions can be found in previous literature (Barnston et al., 2019; Kirtman et al., 2014; Vecchi et al., 2014). In this section, the forecast systems are compared both with and without a posteriori bias correction. Specifically, when model forecast biases are compared (Figures 9–13 and 15), the total forecasts are considered without a posteriori bias correction; when ENSO anomaly prediction skills are compared (Figures 14 and 16), the lead time- and initialization time-dependent biases are removed from all the forecast systems.

The purpose of this paper is primarily to document the initialization scheme for the SPEAR coupled seasonal predictions and the implementation of OTA as a novel application of bias reduction to coupled dynamic prediction. There are many aspects of seasonal prediction skill that could be evaluated, including but not limited to ENSO, sea ice, atmospheric teleconnections, atmospheric rivers, temperature and precipitation outlooks over land, storm tracks, hurricanes and many more. In this section we will focus on the effect of OTA on reducing prediction bias and its potential implications for seasonal prediction applications. Given the dominant role of ENSO variability in seasonal predictability, there will be an evaluation of ENSO prediction skill in the new SPEAR system based on available reforecasts. Other applications of the new SPEAR seasonal prediction system will be reported in future papers.

The lead time- and initialization month-dependent RMSE of global SST forecasts averaged over 1995–2018 is shown in Figure 9. This RMSE is calculated from the differences between each 12-month model forecast and corresponding monthly OISST v2 (Reynolds et al., 2002), then averaged over 1995–2018 for each initialization month. Therefore, it includes the deviations from both observed climatology and variability and is an overall quantification of the SST forecast biases. SPEAR always has smaller RMSE than FLOR and CM2.1, and within SPEAR forecasts, OTA further reduces the forecast RMSE. For each initialization month, all SPEAR reforecasts in Figure 9 start from the same ocean initial conditions, which have the RMSE values close to the SPEAR_ODA ocean state estimation, while the FLOR and CM2.1 forecasts have higher initial RMSE because of the higher RMSE in ECDA ocean analysis (Figure 4). Among the SPEAR reforecasts, RF_LO_1's RMSE of the predicted monthly SST grows more rapidly in the first few months once the reforecasts start, and plateau over the latter half of the reforecasts. For the 3 reforecast configurations with OTA (RF_LO_2/3 and RF_MED), the growth of SST RMSE over the 1-year reforecasts is reduced by 50% or more from RF_LO_1. The 3 reforecast configurations with OTA also have almost identical SST error growth. This resemblance indicates that the addition of the initialized atmosphere/land/sea ice from nudging does not have a significant impact on the large-scale climatology of the ocean forecasts. This resemblance also confirms that the SPEAR_LO and SPEAR_MED models have very similar simulation characteristics (Delworth et al., 2020). Both CM2.1 and FLOR have larger RMSEs of global SST forecasts than even RF_LO_1 without OTA, confirming the improved simulation characteristics in the new SPEAR models. The improvement in the ocean initial conditions can partially explain the reduction in SST forecast bias,
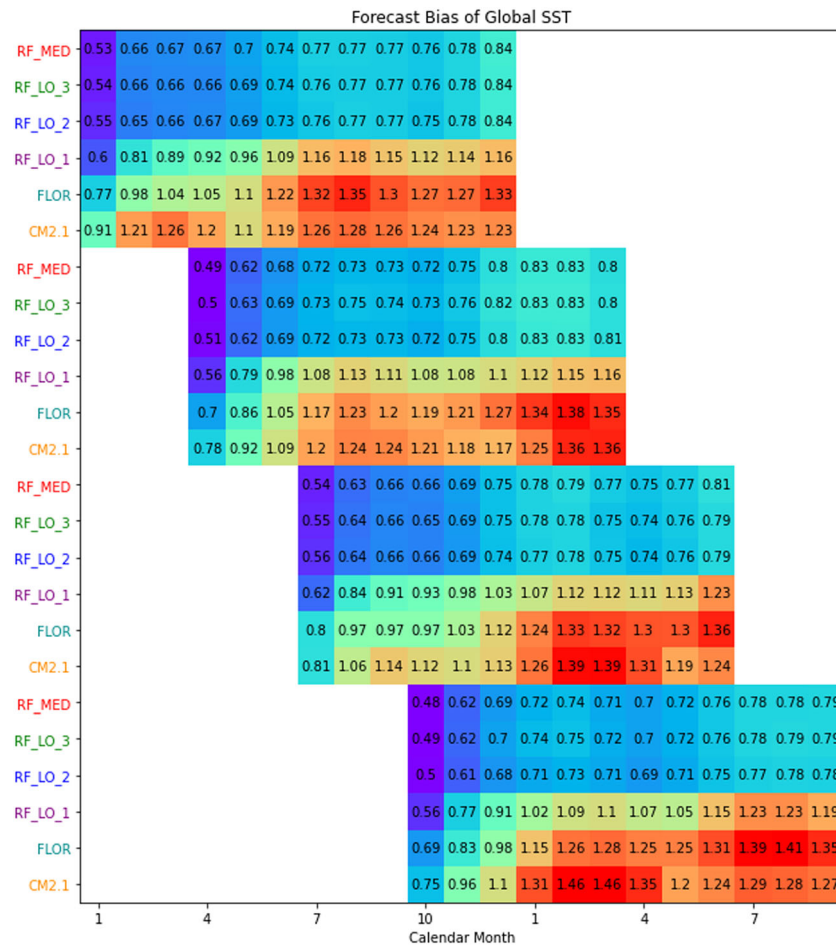
Figure 9. Area-weighted RMSE (units: °C) of global (60°S to 60°N) ensemble-mean monthly SST forecast averaged over 1995–2018, compared to monthly OISST v2. Prediction lead times increase to the right from each of the four initialization months. Forecasts from FLOR and CM2.1 are also included. Colors and texts represent the same values in this and all following heatmaps.

since the eventual RMSE increases over the 12-month forecast experiments are similar for FLOR and RF_LO_1 (around 0.5–0.7). However, the RMSE increases in the reforecasts with OTA are only about 0.3, much smaller than the other cases. CM2.1 has a notable seasonal cycle bias in the SST forecast, peaking from January to March. Therefore, although the final RMSE increase may not be larger than FLOR or RF_LO_1, its maximum RMSE over the forecast period is among the largest for all initialization months. The comparison between the current (CM2.1 and FLOR) and new (RF_LO_3 and RF_MED targeted for NMME) seasonal predictions highlights the reduction in overall SST prediction bias in the new SPEAR seasonal forecast system.

One example of the reduced SST forecast bias is shown in Figure 10, which has the spatial maps of the ensemble-mean climatological SST prediction bias for the fifth month after the models are initialized on 1 April. More specifically, Figure 10 shows the differences between the ensemble-mean August SST averaged from all 1 April initialized forecasts and the observed August SST climatology. Among SPEAR forecasts, RF_LO_1 has the largest SST prediction bias, including cold bias in most of the Southern Hemisphere and central equatorial Pacific, warm bias over the eastern parts of the Pacific and Atlantic basins, biases due to shifted boundary currents, and biases of both signs in the Southern Ocean. Most of the large-scale SST prediction bias is eliminated in the 3 reforecasts with OTA (RF_MED, RF_LO_3 and RF_LO_2), all of which have very similar bias patterns and magnitudes. The remaining bias in eddy-rich regions, that is, along boundary currents and in the Southern Ocean, is much smaller than RF_LO_1, with the only exception being the cold bias in the central equatorial Pacific. The persistent cold bias in the central equatorial
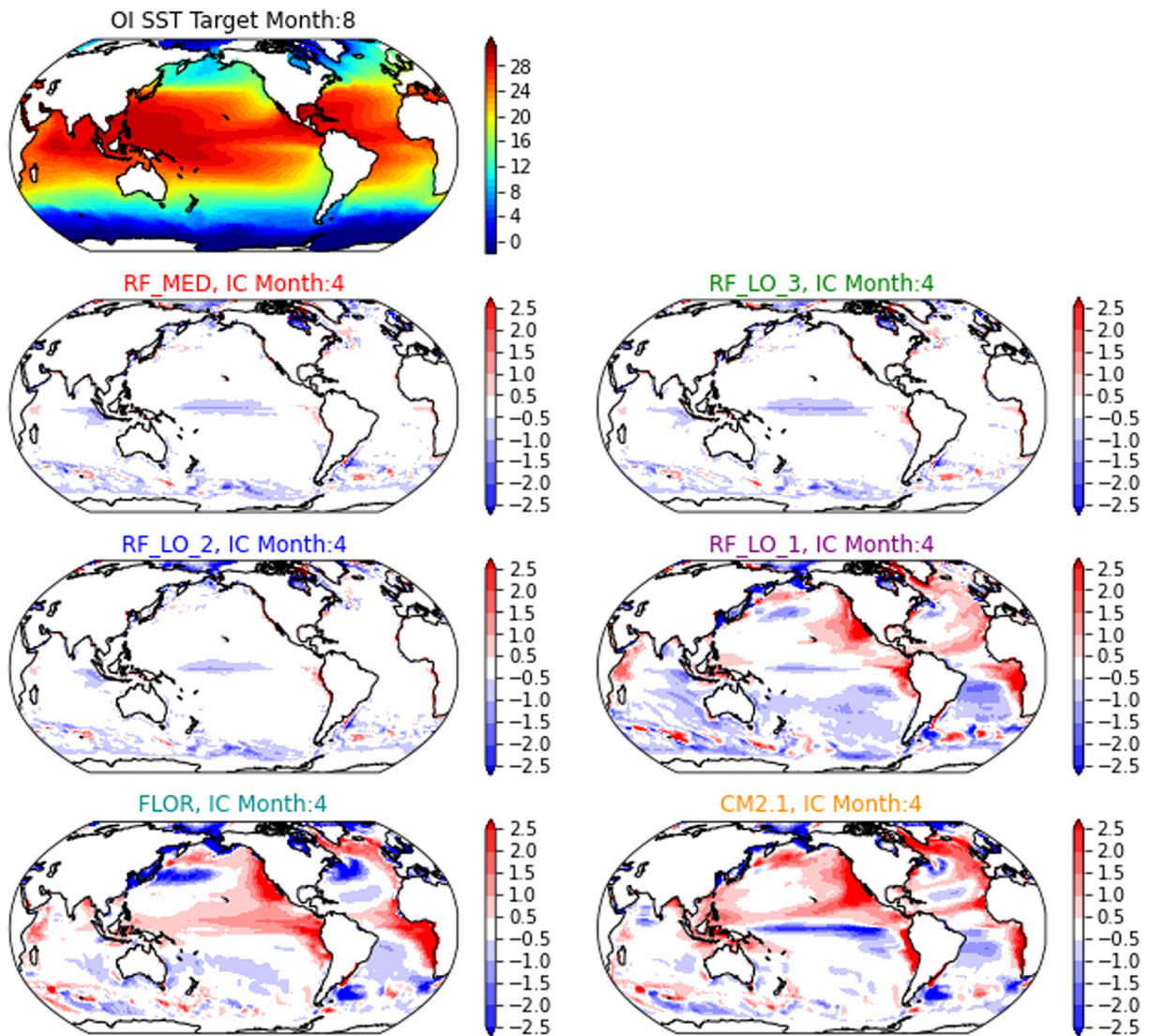
Figure 10. Climatological August SST (units: °C) from monthly OISSTv2 (top) and predicted ensemble-mean August SST bias (units: °C) from all 1 April initialized reforecasts for each configuration.

Pacific during the boreal summer is consistent regardless the initialization months, and it is partly related to positive easterly wind bias in the atmospheric component over the same region, which could not be fully eliminated by OTA due to the strong ocean-atmosphere interaction in the tropics. It could also be related to OTA increments elsewhere, for example the off-equator negative OTA increments in the central Pacific (Figure 1). For the August SST climatological forecast initialized from 1 April, CM2.1 and FLOR both have comparable or larger biases than even RF_LO_1. Significant bias reduction by OTA as shown in Figure 10 occurs for all initialization months and lead times, which can be expected based on Figure 9. The cold SST prediction bias in the central equatorial Pacific after OTA remains a concern for the SPEAR prediction system, with potential impact on local convective activity and ocean-atmosphere coupling. Currently, this cold bias is corrected diagnostically by removing the lead time- and initialization month-dependent forecast climatology in the same fashion as most current seasonal prediction systems including CM2.1 and FLOR. We are also exploring additional methods to further lower the residual SST prediction bias after OTA, including atmospheric FAs to reduce the forcing bias.

The reduction of SST prediction bias by the improved SPEAR models as well as OTA could benefit various seasonal prediction applications. Reduced SST prediction bias has been shown to improve seasonal tropical

cyclone prediction in GFDL's FLOR models (Vecchi et al., 2014), and less biased SST prediction could also serve as a good foundation for Madden-Julian Oscillation (MJO), ENSO teleconnections (Bayr et al., 2019), atmospheric rivers, seasonal temperature and precipitation outlooks, and storm track predictions. One example is the reduced forecast RMSE of surface air temperature over continental United States in SPEAR forecasts compared to FLOR and CM2.1 (supporting information Figure S1). The RMSE in Figure S1 is calculated the same way as in Figure 9, but for surface air temperature over continental United States against the NOAA GHCN_CAMS land temperature analysis (Fan & van den Dool, 2008). Over a majority of the initialization months and lead times, the SPEAR forecasts have smaller RMSE than FLOR or CM2.1. There is more variation among the SPEAR forecasts than in Figure 9 because the land model is neither directly initialized via data assimilation nor directly impacted by OTA. The SPEAR forecasts also have two resolutions and are both interpolated to the same grid as the analysis data. In-depth analysis on such seasonal prediction applications is beyond the scope of this paper and will be reported in future studies.

One important advancement of the new OTA method is its ability to reduce prediction bias for the subsurface ocean. Equatorial Pacific thermocline variability is a key component of ENSO variability and predictability, and model drift toward a biased climatology could be detrimental to ENSO prediction. Figures 11 and 12 show the prediction drift of upper ocean temperature averaged in the Niño3 (5°S to 5°N, 150°W to 90°W) and Niño4 (5°S to 5°N, 160°E to 150°W) regions for 1 January initiated reforecasts, respectively. The drift is calculated as the deviations from the corresponding climatology in the SPEAR_ODA analysis for SPEAR forecasts, or ECDA analysis for CM2.1 and FLOR forecasts. In the Eastern Pacific, the cold drift in RF_LO_1 (Figure 11e) shows up early in the reforecasts, and reappears in July and quickly intensifies to as much as 2°C in October around the depth of 100 m, indicating the tendency for a shallower thermocline than in the ocean analysis. This cold drift is much reduced in the reforecasts with OTA (Figures 11b–11d), which have much weaker cold deviations over the latter half of the reforecasts. Similarly, in the Western Pacific, RF_LO_1 tends to shoal its thermocline too much as soon as the reforecasts are initialized and persists the shallow thermocline throughout the reforecasts (Figure 12e). Such cold drift is again reduced by OTA, resulting in smaller cold biases over the latter half of the reforecasts (Figures 12b–12d). The equatorial Pacific subsurface temperature drift in either CM2.1 or FLOR is larger than that in even RF_LO_1. For the Niño3 region, FLOR develops and persists large warm biases (Figure 11g) while CM2.1 does the opposite (Figure 11h). For the Niño4 region, FLOR forecast has a warm surface bias along with a shallow thermocline (Figure 12g), while CM2.1 forecast has a cold bias mostly in the upper 100 m (Figure 12h).

The reduction in subsurface ocean drift with OTA is not limited to the equatorial Pacific. The ocean temperature drift averaged over the North Pacific (30°N- 50°N, 160°E-150°W) is shown in Figure 13. RF_LO_1 starts with smaller drift than in the equatorial Pacific (Figure 13e), and OTA further eliminates such drift almost completely throughout the forecast period in RF_LO_2 (Figure 13d), RF_LO_3 (Figure 13c), and RF_MED (Figure 13b). FLOR (Figure 13g) and CM2.1 (Figure 13h) again have larger subsurface temperature drift than RF_LO_1, indicating the improvement in the simulation quality of the SPEAR models even before the use of OTA. The drift of ocean temperature forecasts across all latitudes is reduced by OTA, as shown by the lead time-dependent RMSE of the 1 July climatological forecasts of 0–300 m average ocean temperature (Figure S2). RF_LO_1 has the largest drift between 15°S and 15°N, around 40°N, and between 35°S and 60°S. With OTA, forecast drift of 0-300 m ocean temperature starts later in the forecast period and grows to a lesser extent.

Next, we show the ENSO prediction skills in terms of ensemble-mean Niño3.4 SST indices (average SST in 5°S to 5°N, 170°W to 120°W). Figure 14 shows the ACC (anomaly correlation coefficient) and RMSE for the predicted ensemble-mean Niño3.4 SST anomalies evaluated with respect to monthly OISST v2 (Reynolds et al., 2002) anomalies. Figure 14 includes all initialization months and years of each configuration. The forecast anomalies of each configuration are calculated by the a posteriori removal of the corresponding lead time- and initialization month-dependent forecast climatology. In this case, the RMSE does not include the forecast bias of each system as in Figure 9; instead, this RMSE is another indicator of the anomaly prediction skill. It usually agrees with the ACC skills as shown in Figure 14, but emphasizes more on "false alarm" predictions. RF_MED has overall the highest ACC at all lead times except for the first month (CM2.1 and FLOR are highest) and last month (RF_LO_3 is highest). The first-month ACC is very high across the board, at 0.97 for the SPEAR reforecasts and 0.98 for CM2.1/
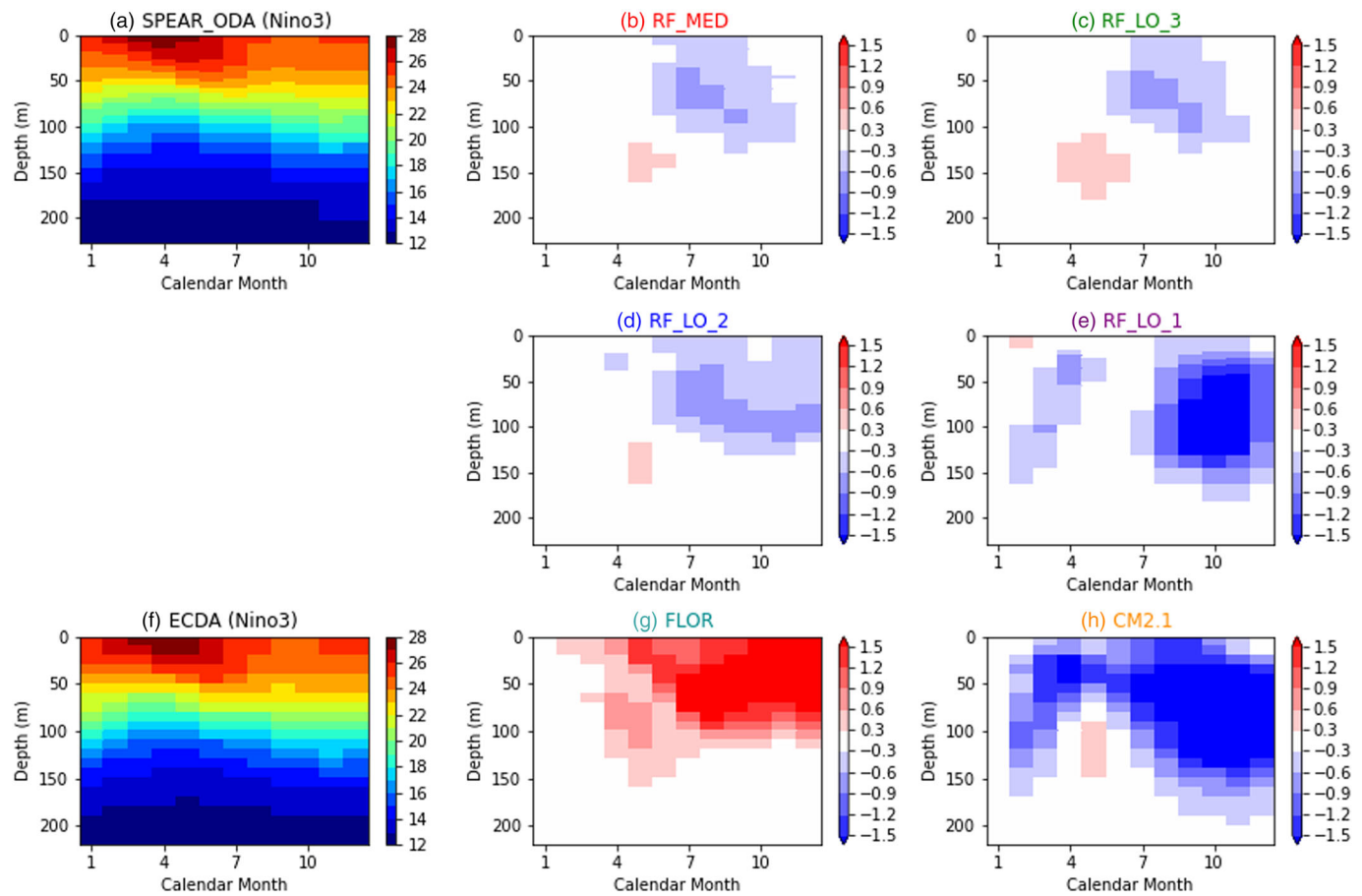
Figure 11. Climatological annual cycle of ocean temperature profile averaged in the Niño3 region (5°S to 5°N, 150°W to 90°W) from the (a) SPEAR_ODA and (f) ECDA, respectively, and the average prediction drift from all 1 January initiated forecasts for each reforecast configuration. SPEAR forecast drifts are relative to SPEAR_ODA analysis, while CM2.1 and FLOR forecast drifts are relative to ECDA analysis. All values have the units of °C.

FLOR, and this small deficit quickly reverses as the forecast continues. The ACC of RF_LO_3 (0.57) at the 11-month lead is slightly higher than RF_MED (0.54), and it is caused by smaller ACC at the end of July and October reforecasts (Figure S5), when the simulations have passed the ENSO spring prediction barrier for the next calendar year, and the ENSO predictability is heavily affected by highly uncertain and unpredictable variability in the atmospheric forcing. In terms of RMSE for the ensemble-mean Niño3.4 SST anomalies, RF_MED is again among the smallest except for the first 2 months and last month, with RF_LO_3 having the lowest RMSE in the last month. The slightly larger RMSE initially again quickly reverses, and the smaller last-month RMSE for RF_LO_3 coincides with the larger ACC as mentioned earlier. The ENSO prediction skills in terms of ensemble-mean Niño3 and Niño4 SST anomalies (Figures S3 and S4, respectively) show similar results as those based on Niño3.4 anomalies (Figure 14). RF_MED also has the largest ACC and smallest RMSE except for a few lead times where RF_LO_3 or CM2.1 are comparable or slightly better.

When the ensemble-mean Niño3.4 SST anomaly reforecasts are individually analyzed for each of the four individual initialization months (Figure S5), RF_MED still holds the largest ACC and smallest RMSE at most lead times across all initialization months, although there are more cases than in Figure 14 where other forecasts, primarily RF_LO_3 and CM2.1, would outperform RF_MED. The increased variation in the relative quality among the forecast configurations is caused by the smaller sample size (24 forecast experiments for each month instead of 96 for all 4) and more influence from each model's climatological annual cycle.

Figure 15 shows the forecast bias for the total ensemble-mean Niño3.4 SST index by initialization month, which does not remove the lead time-dependent forecast climatology. All SPEAR reforecasts with OTA
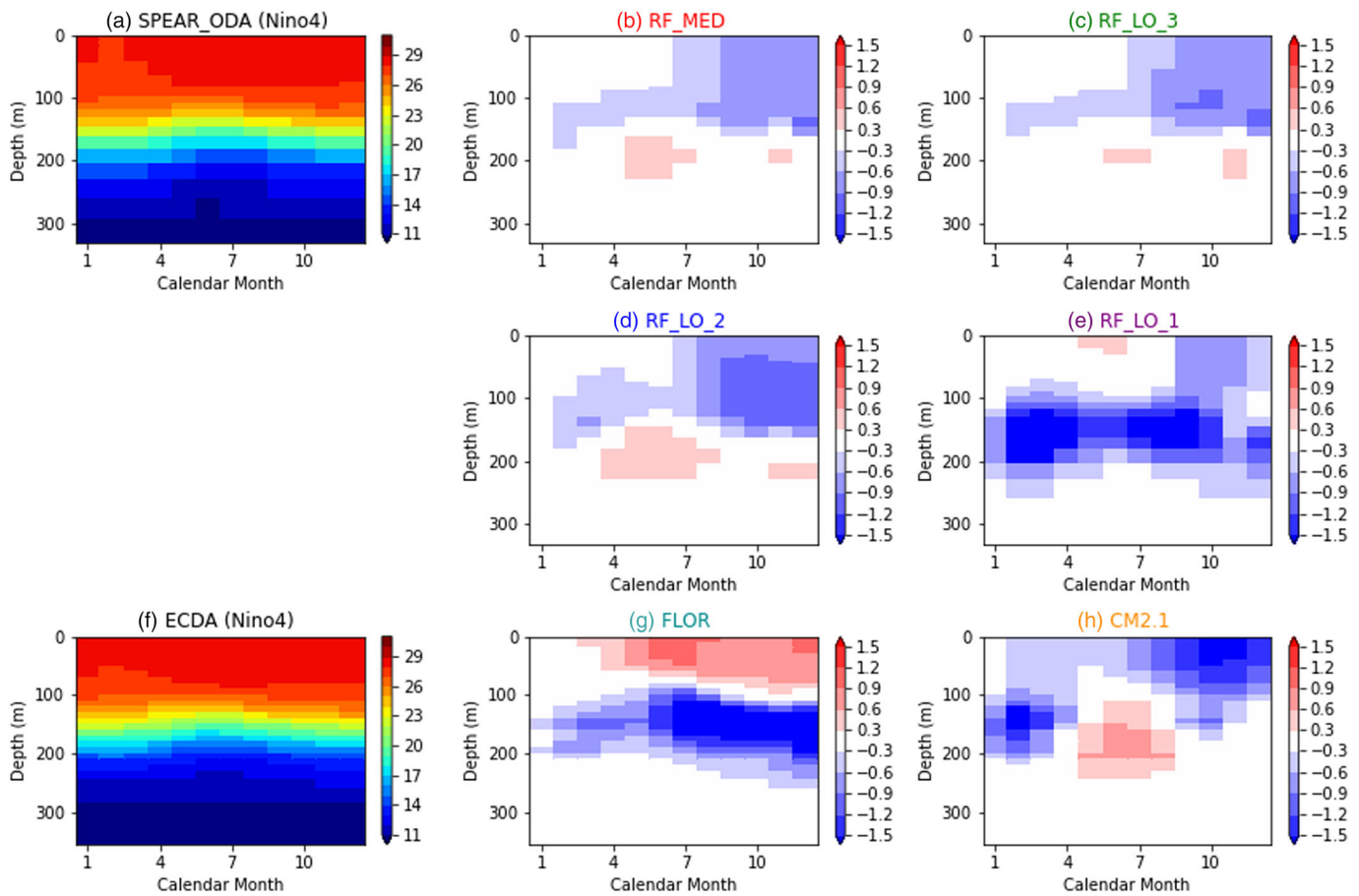
Figure 12. Similar to Figure 11, but for the Niño4 region (5°S to 5°N, 160°E to 150°W).

have ensemble-mean Niño3.4 SST prediction biases that increase slower, and do not exceed a magnitude of 1°C, except for the final two months of RF_LO_3 October reforecasts. Without OTA, RF_LO_1 has an annual cycle of bias that shows positive values around boreal spring and negative values around boreal fall. The forecast biases in CM2.1 and FLOR both have larger magnitude than any SPEAR reforecasts, with CM2.1 having peak negative values exceeding 2°C in boreal fall, and FLOR having peak positive values exceeding 2°C in boreal winter. For FLOR and CM2.1, although their Niño3.4 SST forecast biases are large compared to SPEAR reforecasts and sometimes even exceed the magnitude of ENSO anomaly signals, the impact of such large biases on their ENSO SST anomaly prediction skill in terms of ACC is somewhat limited. One reason, in this current study, may be that the specific bias structure of the models involved does not interfere strongly with the nonlinear aspect of ENSO dynamics (Timmermann et al., 2018). Another possible and more general explanation may be that the ENSO phenomenon, when viewed as a whole in terms of Niño3.4 SST index, behaves like linear dynamics more often than not. This view has been demonstrated by successful ENSO prediction using a Linear Inverse Model (LIM) approach (Newman et al., 2009). Detailed analysis of the relationship between forecast bias and anomaly prediction skill (e.g., Ding et al., 2020) is beyond the scope of this paper. The newly proposed Tropical Pacific Observing System 2020 Project (TPOS2020, Kessler et al., 2019) may bring new opportunities to ODA and tropical climate prediction with its revamped buoy network and proposed doubling of Argo sampling within 10°S to 10°N. Beyond just ensemble-mean ENSO SST anomaly prediction skill, the demonstrated reduction in prediction bias of both SST and subsurface temperature could benefit seasonal prediction by improving the background conditions for atmospheric teleconnections (Alexander et al., 2002; Bayr et al., 2019) and mid-latitude ocean temperature reemergence (Deser et al., 2003). This topic will be further explored in future studies of various seasonal prediction applications.
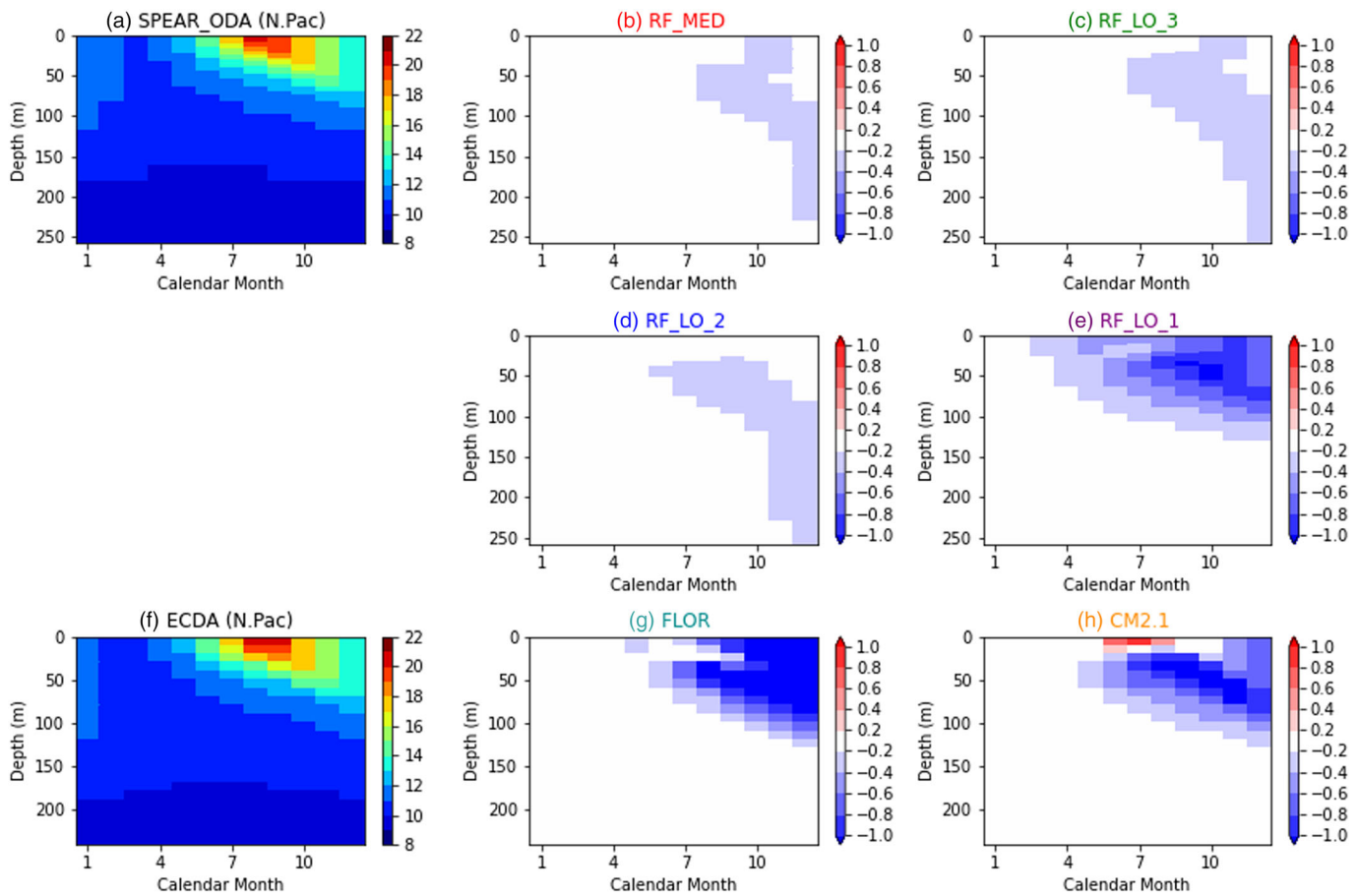
Figure 13. Similar to Figure 11, but for the North Pacific region of 30°N to 50°N, 160°E to 150°W.

Forecast skill of the two strongest El Niño events in the last 30 years is examined by ensemble-mean Niño3.4 SST anomaly forecasts for the time periods of 1996–2000 and 2014–2018 in Figure 16. The full forecast ensembles during the same two periods are shown in Figures S6 and S7. The initialization months include January, April and July. The ENSO prediction skill in dynamic forecast systems is usually low when the forecast is initialized on or before April. This spring prediction barrier is demonstrated by the sharp drop in ACC at 4 month lead in the January forecasts and at 1–2 month lead in the April forecasts across all systems (Figure S5). First we look at the case of the 1997–1998 El Niño. In the January 1997 ensemble-mean forecasts (Figure 16a), RF_LO_3 and RF_MED predict stronger El Niño by the end of 1997 than CM2.1 or FLOR. The observed Niño3.4 anomalies by the end of 1997 also sit within the RF_MED and RF_LO_3 ensembles (Figure S6a), instead of being an outlier for CM2.1 and FLOR ensembles (Figure S6g). All ensemble-mean forecasts, except FLOR, predict the strong El Niño from April 1997 (Figure 16b), and RF_MED, RF_LO_3 and CM2.1 are the closest to the observed Niño3.4 anomaly magnitude. By July 1997 past the spring prediction barrier, all the ensemble-mean forecasts predict the strong El Niño event (Figure 16c). Moving on to the 2015–2016 El Niño event, none of the January 2015 ensemble-mean forecasts predict the subsequent El Niño condition (Figure 16d). Among all ensemble members, only a few predict strong El Niño anomalies by the end of 2015 (Figures S7a, S7d, and S7g), which include 1 member of RF_MED, 1 of RF_LO_3, 2 of RF_LO_1 and 1 of FLOR. From April 2015 (Figure 16e), the ensemble-mean Niño3.4 anomaly forecasts are in the range of 1.0–1.5°C across the forecasts, with CM2.1 and FLOR stronger than the SPEAR reforecasts. There are also multiple ensemble members in each configuration predicting Niño3.4 anomalies above 1.5°C or 2°C. However, the observed Niño3.4 anomalies throughout 2015 are still the outlier across all forecast ensembles (Figures S7b, S7e, and S7h). The weak magnitude of the 2015–2016 El Niño predicted from April 2015 also occurs in many operational seasonal prediction systems, highlighting the challenge from the
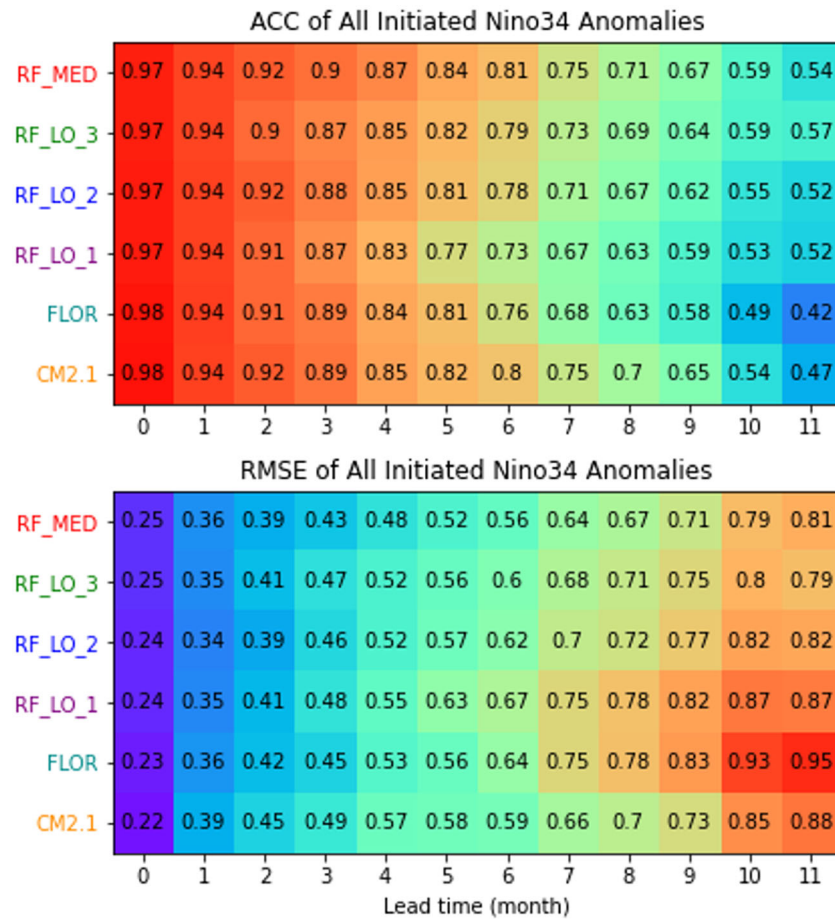
Figure 14. Prediction skills of Niño3.4 SST in terms of (top) ACC and (bottom) RMSE. ACC and RMSE are calculated based on forecast anomalies and observed anomalies, which are obtained by removing the lead time- and initialization month-dependent forecast climatology and observed climatology, respectively.

ENSO spring prediction barrier (L'Heureux et al., 2017). Meanwhile, the wide ensemble spreads for all the forecasts in January and April 2015 demonstrate the need for large ensembles and a probabilistic view on ENSO seasonal prediction. Back to April 2014, RF_MED and CM2.1 are the only two ensemble-mean forecasts that do not predict Niño3.4 anomalies above 1.0°C throughout 2014, avoiding the so-called "failed El Niño prediction" (McPhaden, 2015). For the July 2015 forecasts (Figure 16f), El Niño conditions are predicted by all ensemble-mean forecasts. However, RF_MED, RF_LO_3 and RF_LO_2 forecast ensembles best encompass the observed Niño3.4 anomalies (Figures S7c and S7f), resulting in the most accurate ensemble-mean magnitude (Figure 16f). While the observed Niño3.4 anomalies stay mostly as an outlier for RF_LO_1, FLOR and CM2.1 July 2015 forecast ensembles (Figures S7f and S7i). Similar advantages of the RF_MED, RF_LO_3 and RF_LO_2 forecast ensembles are also found in the cases of April 1997 (Figures S6b, S6e, and S6h) and July 1997 (Figures S6c, S6f, and S6i). We also note that the SPEAR ensemble-mean forecasts with OTA generally track the evolution and magnitude of the subsequent La Niña events better than CM2.1 or FLOR across all cases shown in Figure 16, which should be expected given their overall larger ACC and smaller RMSE across all forecasts (Figure 14). Another notable feature is FLOR's colder anomaly forecasts for both El Niño (Figures 16b and 16f) and La Niña (Figures 16a and 16d) events. These cold anomaly forecasts could be directly related to the warm forecast bias in FLOR (Figure 15). As discussed for Figure 14 and Figure 15, the 1997–1998 and 2015–2016 El Niño forecasts suggest that CM2.1's forecast bias has less negative impact on its ensemble-mean anomaly prediction skill than FLOR. Specific case and mechanism studies will be reported in future papers.
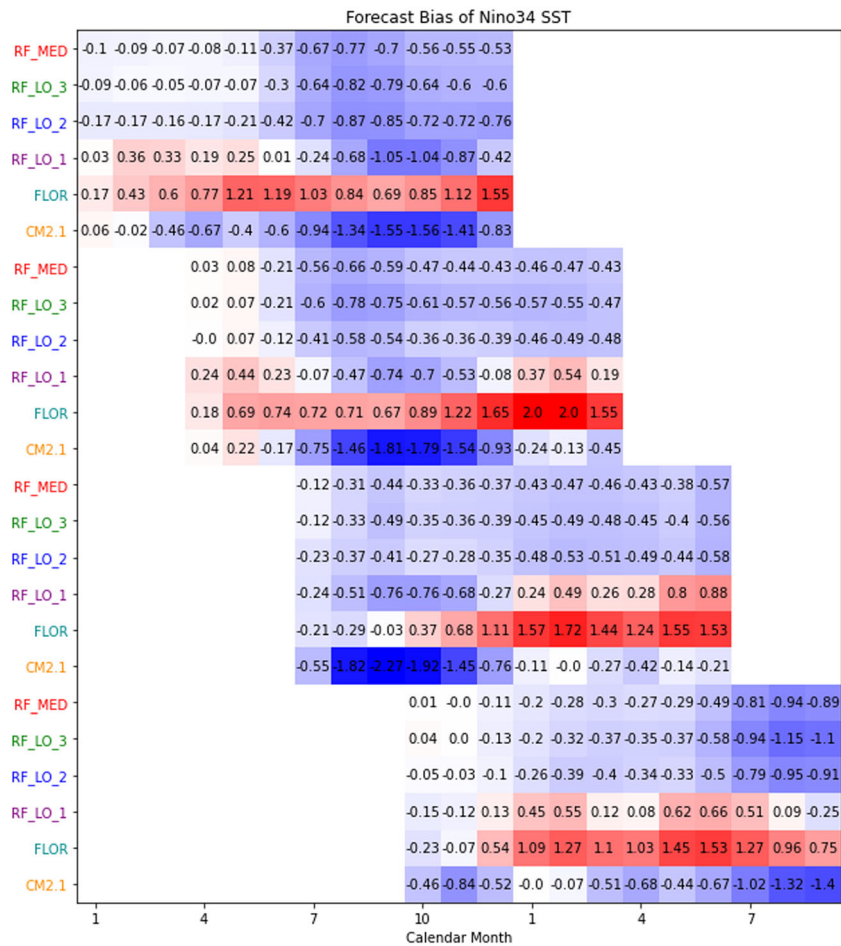
Figure 15. Average forecast bias (units: °C) of total Niño3.4 SST index, shown for each initialization month individually.

Overall, the new RF_MED forecast configuration targeted for NMME has the best ensemble-mean ENSO prediction skills in terms of bias, ACC and RMSE across multiple ENSO SST indices, followed by RF_LO_3 and CM2.1, then RF_LO_2, RF_LO_1 and FLOR in order. The similarity between RF_MED and RF_LO_3 is expected and also desired since the SPEAR_MED and SPEAR_LO models share everything except for the atmospheric resolution, and the two reforecast configurations also have the same initialization scheme except that the nudging experiments are performed with each model, respectively. CM2.1's high ENSO anomaly prediction skills, especially in comparison to FLOR, may be explained by two primary reasons. One is that CM2.1 has demonstrated exceptional skill of simulating ENSO variability, and has been widely used for ENSO studies both at GFDL and across the climate community (e.g., Atwood et al., 2016; Capotondi et al., 2015; Chen et al., 2017; Karamperidou et al., 2014; Ogata et al., 2013; Wittenberg, 2009; Wittenberg et al., 2006, 2014). On the other hand, FLOR is initially designed to focus on applications such as hurricane and hydroclimate to take advantage of its high atmospheric resolution. The other reason is that ECDA is initially developed for CM2.1, but later adapted to also provide the initialization for FLOR. This leads to two key differences between the initializations of CM2.1 and FLOR forecasts. First, the ECDA ocean initial conditions have a slight mismatch with the FLOR ocean model (Vecchi et al., 2014), which could lead to initialization shock in the forecasts. Second, CM2.1 forecasts have an initialized atmosphere component from the coupled data assimilation in ECDA, while FLOR forecasts use atmosphere-only experiments with prescribed SST to generate atmosphere initial conditions, which are not constrained by any atmospheric observations. The improvement from RF_LO_2 to RF_LO_3 also shows that initialized atmosphere and land components could benefit seasonal forecasts throughout the prediction period.
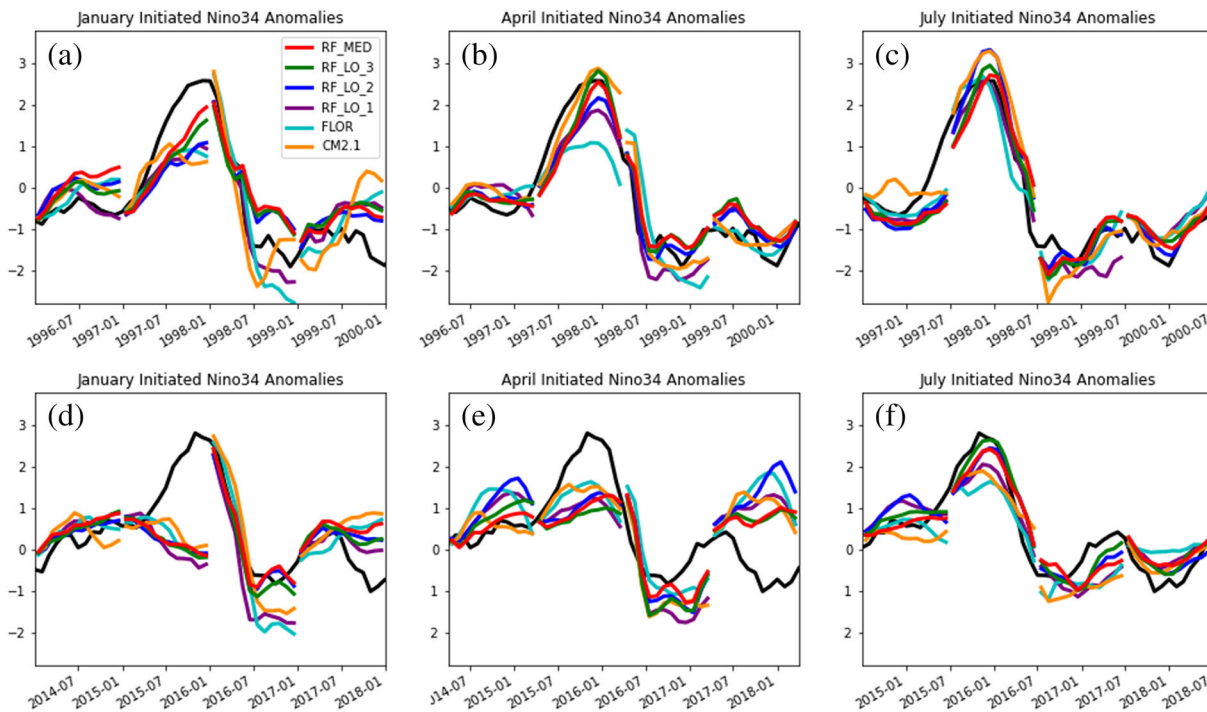
Figure 16. Ensemble-mean forecasts of the Niño3.4 SST anomalies (units: °C) during 1996–2000 (a–c) and 2014–2018 (d–f). Black line is the observed Niño3.4 anomalies from monthly OISST v2. The initialization months are January, April, and July (left to right).

5. Summary and Outlook

Following the successful development of the next-generation SPEAR prediction models (Delworth et al., 2020), here we document the development of the updated ODA system for MOM6 and coupled initialization scheme for the next-generation SPEAR seasonal predictions at GFDL, as well as the novel application of a bias correction method called OTA to coupled model predictions. These efforts are part of SPEAR_ECDA, which includes all the data assimilation and initialization development to facilitate the goals of seamless seasonal-to-decadal prediction and projection within SPEAR.

The updated ensemble-based ODA system for the MOM6 ocean model makes use of advances in modeling system infrastructure and MOM6 development at GFDL. It inherits the EAKF algorithm from GFDL's previous ECDA system but has improved flexibility and efficiency to enable rapid research and development on subjects like bias correction methods and seasonal prediction initialization. A new ocean state estimation (SPEAR_ODA) for the time period of 1990 to present is produced with the MOM6 ODA system in the coupled SPEAR_LO model. Going forward, SPEAR_ODA will be updated in real time at the start of each month to provide ocean initial conditions for seasonal predictions in both the SPEAR_MED and SPEAR_LO models, which have the same MOM6 ocean component. The atmosphere, land, and sea ice initial conditions for SPEAR_MED and SPEAR_LO seasonal predictions are obtained from separate SPEAR_nudged experiments where atmospheric variables are nudged to CFSR data and SST is nudged to the same SST dataset as assimilated by SPEAR_ODA. SPEAR_nudged experiments are performed in both SPEAR_MED and SPEAR_LO coupled models to provide initial conditions for seasonal predictions with each model, respectively.

A bias correction method, named OTA, is developed as part of SPEAR_ECDA. Similar to some existing bias correction methods for ODA, OTA estimates the climatological temperature and salinity increments from an Argo-era data assimilation experiment and re-apply these increments to the ocean model to reduce forecast bias. However, the innovation of OTA is its application to coupled model seasonal predictions, which is feasible because the OTA increments are estimated by ODA experiments in coupled models with a freely evolving atmosphere component, therefore matching the atmospheric forcing in coupled model predictions.

Preliminary seasonal reforecast experiments demonstrate that the forecast configurations from the SPEAR system targeted for NMME produce overall superior seasonal forecasts compared to GFDL's current NMME seasonal forecasts with CM2.1 and FLOR models. Additional reforecast configurations are used to test the effect of OTA and SPEAR_nudged initial conditions. The improvements in the SPEAR models and the application of OTA both reduce the seasonal forecast bias in the ocean, including both SST and subsurface temperature. Among the SPEAR seasonal prediction configurations, the reduction in forecast bias due to OTA is also accompanied with increase in the anomaly prediction skill, such as the ACC and RMSE of ensemble-mean ENSO SST anomaly indices. Although the ENSO anomaly prediction skill is not significantly improved in the new SPEAR system compared to previous CM2.1 predictions, we are actively investigating the impact of reduced SST forecast bias on other seasonal prediction applications such as ENSO teleconnections, atmospheric rivers, and extratropical storms.

The RF_MED reforecast configuration using the SPEAR_MED model will be the target forecast configuration for NMME as the successor to the current CM2.1 and FLOR seasonal forecasts. The transition is currently scheduled to start by the end of 2020.

Data Availability Statement

CM2.1 and FLOR seasonal forecast data are available through the NMME project (https://www.cpc.ncep.noaa.gov/products/NMME/). Production SPEAR seasonal reforecast and real-time forecast data will be publicly available through NMME once the transition is finished.

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL Global Ocean and Sea Ice Model OM4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*, 3167–3211. https://doi.org/10.1029/2019MS001726

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N. C., & Scott, J. D. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*, *15*(16), 2205–2231. https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2

Anderson, J. L. (2003). A local least squares framework for ensemble filtering. *Monthly Weather Review*, *131*(4), 634–642.

Antonov, J. I., Locarini, R. A., Boyer, T. P., Mishonov, A. V., & Garcia, H. E. (2006). *World Ocean Atlas 2005, Volume 2: Salinity*. In S. Levitus (Ed.). Washington, DC: NOAA Atlas NESDIS 62.

Argo. (2019). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182

Atwood, A. R., Battisti, D. S., Wittenberg, A. T., Roberts, W. H. G., & Vimont, D. J. (2016). Characterizing unforced multi-decadal variability of ENSO: A case study with the GFDL CM2. 1 coupled GCM. *Climate Dynamics*, *49*(7–8), 2845–2862. https://doi.org/10.1007/s00382-016-3477-9

Balaji, V. (2012). The Flexible Modeling System. In *Earth system modelling* (Vol. 3, pp. 33–41). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-23360-9_5

Balmaseda, M. (2017). Data assimilation for initialization of seasonal forecasts. *Journal of Marine Research*, *75*(3), 331–359. https://doi.org/10.1357/002224017821836806

Balmaseda, M., & Anderson, D. (2009). Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, *36*, L01701. https://doi.org/10.1029/2008GL035561

Balmaseda, M., Dee, D., Vidard, A., & Anderson, D. L. T. (2007). A multivariate treatment of bias for sequential data assimilation: Application to the tropical oceans. *Quarterly Journal of the Royal Meteorological Society*, *133*(622), 167–179. https://doi.org/10.1002/qj.12

Balmaseda, M., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. https://doi.org/10.1002/qj.2063

Banzon, V., Smith, T. M., Chin, T. M., Liu, C., & Hankins, W. (2016). A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth System Science Data*, *8*(1), 165–176. https://doi.org/10.5194/essd-8-165-2016

Banzon, V., Smith, T. M., Steele, M., Huang, B., & Zhang, H.-M. (2020). Improved estimation of proxy sea surface temperature in the Arctic. *Journal of Atmospheric and Oceanic Technology*, *37*(2), 341–349. https://doi.org/10.1175/JTECH-D-19-0177.1

Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2019). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, *53*(12), 7215–7234. https://doi.org/10.1007/s00382-017-3603-3

Bayr, T., Domeisen, D. I. V., & Wengel, C. (2019). The effect of the equatorial Pacific cold SST bias on simulated ENSO teleconnections to the North Pacific and California. *Climate Dynamics*, *53*(7–8), 3771–3789. https://doi.org/10.1007/s00382-019-04746-9

Bonan, D. B., Bushuk, M., & Winton, M. (2019). A spring barrier for regional predictions of summer Arctic sea ice. *Geophysical Research Letters*, *46*, 5937–5947. https://doi.org/10.1029/2019GL082947

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., & Yang, X. (2017). Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophysical Research Letters*, *44*, 4953–4964. https://doi.org/10.1002/2017GL073155

Capotondi, A., Wittenberg, A. T., Newman, M., di Lorenzo, E., Yu, J. Y., Braconnot, P., et al. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, *96*(6), 921–938. https://doi.org/10.1175/BAMS-D-13-00117.1

Carton, J. A., Chepurin, G. A., & Chen, L. (2018). SODA3: A new ocean climate reanalysis. *Journal of Climate*, *31*(17), 6967–6983. https://doi.org/10.1175/JCLI-D-18-0149.1

Chang, Y., Schubert, S. D., Koster, R. D., Molod, A. M., & Wang, H. (2019). Tendency bias correction in coupled and uncoupled global climate models with a focus on impacts over North America. *Journal of Climate*, *32*(2), 639–661. https://doi.org/10.1175/JCLI-D-18-0598.1

Chang, Y.-S., Zhang, S., Rosati, A., Delworth, T. L., & Stern, W. F. (2013). An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dynamics*, *40*(3–4), 775–803. https://doi.org/10.1007/s00382-012-1412-2

Chen, C., Cane, M. A., Wittenberg, A. T., & Chen, D. (2017). ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *Journal of Climate*, *30*(2), 775–801. https://doi.org/10.1175/JCLI-D-15-0901.1

Chepurin, G. A., Carton, J. A., & Dee, D. P. (2010). Forecast model bias correction in ocean data assimilation. Retrieved from https://doi.org/10.1175/MWR2920.1

Danforth, C. M., Kalnay, E., & Miyoshi, T. (2007). Estimating and correcting global weather model error. *Monthly Weather Review*, *135*(2), 281–299. https://doi.org/10.1175/MWR3289.1

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3323–3343. https://doi.org/10.1256/qj.05.137

Dee, D. P., & da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, *124*(545), 269–295.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643–674. https://doi.org/10.1175/JCLI3629.1

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (2020). SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001895. https://doi.org/10.1029/2019ms001895

Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., et al. (2012). Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*, *25*(8), 2755–2781. https://doi.org/10.1175/JCLI-D-11-00316.1

Deser, C., Alexander, M. a., & Timlin, M. S. (2003). Understanding the persistence of sea surface temperature anomalies in midlatitudes. *Journal of Climate*, *16*(1), 57–72. https://doi.org/10.1175/1520-0442(2003)016<0057:UTPOSS>2.0.CO;2

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2020). Relating CMIP5 model biases to seasonal forecast skill in the tropical Pacific. *Geophysical Research Letters*, *47*, e2019GL086765. https://doi.org/10.1029/2019GL086765

Doblas-Reyes, F. J., Balmaseda, M., Weisheimer, A., & Palmer, T. N. (2011). Decadal climate prediction with the European Centre for Medium-Range Weather Forecasts coupled forecast system: Impact of ocean observations. *Journal of Geophysical Research*, *116*, D19111. https://doi.org/10.1029/2010JD015394

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS002015. https://doi.org/10.1029/2019ms002015

Evensen, G., & van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, *128*(6), 1852–1867. https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2

Fan, Y., & van den Dool, H. (2008). A global monthly land surface air temperature analysis for 1948–present. *Journal of Geophysical Research*, *113*, D01103. https://doi.org/10.1029/2007JD008470

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

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*, 6704–6716. https://doi.org/10.1002/2013JC009067

Harris, L. M., & Lin, S.-J. (2013). A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, *141*(1), 283–306. https://doi.org/10.1175/MWR-D-11-00201.1

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*, 3691–3727. https://doi.org/10.1029/2019MS001829

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), 2418–2427. https://doi.org/10.1175/2009jcli3466.1

IOC, SCOR, & IAPSO. (2010). The international thermodynamic equation of seawater—2010: Calculation and use of thermodynamic properties. *Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, 1–196*. UNESCO.

Jayne, S., Roemmich, D., Zilberman, N., Riser, S., Johnson, K., Johnson, G., & Piotrowicz, S. (2017). The Argo program: Present and future. *Oceanography*, *30*(2), 18–28. https://doi.org/10.5670/oceanog.2017.213

Jia, L., Vecchi, G. A., Yang, X., Gudgel, R. G., Delworth, T. L., Stern, W. F., et al. (2016). The roles of radiative forcing, sea surface temperatures, and atmospheric and land initial conditions in U.S. summer warming episodes. *Journal of Climate*, *29*(11), 4121–4135. https://doi.org/10.1175/JCLI-D-15-0471.1

Jia, L., Yang, X., Vecchi, G., Gudgel, R., Delworth, T., Fueglistaler, S., et al. (2017). Seasonal prediction skill of northern extratropical surface temperature driven by the stratosphere. *Journal of Climate*, *30*(12), 4463–4475. https://doi.org/10.1175/JCLI-D-16-0475.1

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, *12*(3), 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Karamperidou, C., Cane, M. A., Lall, U., & Wittenberg, A. T. (2014). Intrinsic modulation of ENSO predictability viewed through a local Lyapunov lens. *Climate Dynamics*, *42*(1–2), 253–270. https://doi.org/10.1007/s00382-013-1759-z

Kessler, W. S., Wijffels, S. E., Cravatte, S., & Smith, N. (2019). TPOS 2020 second report. https://doi.org/10.13140/RG.2.2.17161.29289

Kim, H.-M., Webster, P. J., & Curry, J. A. (2012). Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. *Geophysical Research Letters*, *39*, L10701. https://doi.org/10.1029/2012GL051644

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L. III, Paolino, D. A., Zhang, Q., et al. (2014). The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Knutson, T. R., Delworth, T. L., Dixon, K. W., Held, I. M., Lu, J., Ramaswamy, V., et al. (2006). Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *Journal of Climate*, *19*(9), 1624–1651. https://doi.org/10.1175/JCLI3709.1

Kumar, A., Chen, M., Zhang, L., Wang, W., Xue, Y., Wen, C., et al. (2012). An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) Version 2. *Monthly Weather Review*, *140*(9), 3003–3016. https://doi.org/10.1175/MWR-D-11-00335.1

L'Heureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., di Liberto, T. E., et al. (2017). Observing and predicting the 2015/16 El Niño. *Bulletin of the American Meteorological Society*, *98*(7), 1363–1382. https://doi.org/10.1175/BAMS-D-16-0009.1

Liu, J., & Curry, J. A. (2010). Accelerated warming of the Southern Ocean and its impacts on the hydrological cycle and sea ice. *Proceedings of the National Academy of Sciences*, *107*(34), 14,987–14,992. https://doi.org/10.1073/pnas.1003336107

Locarini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., & Garcia, H. E. (2006). *World Ocean Atlas 2005, Volume 1: Temperature*. In S. Levitus (Ed.). Washington, DC: NOAA Atlas NESDIS 61.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., et al. (2015). Global Seasonal forecast system Version 5 (GloSea5): A high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1072–1084. https://doi.org/10.1002/qj.2396

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, *41*(9–10), 2393–2409. https://doi.org/10.1007/s00382-012-1599-2

McPhaden, M. J. (2015). Playing hide and seek with El Niño. *Nature Climate Change*, *5*(9), 791–795. https://doi.org/10.1038/nclimate2775

Merlis, T. M., Held, I. M., Stenchikov, G. L., Zeng, F., & Horowitz, L. W. (2014). Constraining transient climate sensitivity using coupled climate model simulations of volcanic eruptions. *Journal of Climate*, *27*(20), 7781–7795. https://doi.org/10.1175/JCLI-D-14-00214.1

Newman, M., Sardeshmukh, P. D., & Penland, C. (2009). How important is air–sea coupling in ENSO and MJO evolution? *Journal of Climate*, *22*(11), 2958–2977. https://doi.org/10.1175/2008JCLI2659.1

Ogata, T., Xie, S.-P., Wittenberg, A., & Sun, D.-Z. (2013). Interdecadal amplitude modulation of El Niño–Southern Oscillation and its impact on tropical Pacific decadal variability. *Journal of Climate*, *26*(18), 7280–7297. https://doi.org/10.1175/JCLI-D-12-00415.1

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., et al. (2004). Development of a European Multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, *85*(6), 853–872. https://doi.org/10.1175/BAMS-85-6-853

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, *15*(13), 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. https://doi.org/10.1175/2007JCLI1824.1

Rosati, A., Miyakoda, K., & Gudgel, R. (1997). The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Monthly Weather Review*, *125*(5), 754–772. https://doi.org/10.1175/1520-0493(1997)125<0754:TIOOIC>2.0.CO;2

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058. https://doi.org/10.1175/2010BAMS3001.1

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Shu, Q., Song, Z., & Qiao, F. (2015). Assessment of sea ice simulations in the CMIP5 models. *The Cryosphere*, *9*(1), 399–409. https://doi.org/10.5194/tc-9-399-2015

Smith, D. M., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, *41*(11–12), 3325–3338. https://doi.org/10.1007/s00382-013-1683-2

Stockdale, T., Doblas-reyes, F. J., & Ferranti, L. (2008). EUROSIP: Multi-model seasonal forecasting. *ECMWF Newsletter* (pp. 10–16). https://doi.org/10.21957/7wc0nybvir

Stockdale, T. N. (1997). Coupled ocean–atmosphere forecasts in the presence of climate drift. *Monthly Weather Review*, *125*(5), 809–818. https://doi.org/10.1175/1520-0493(1997)125<0809:COAFIT>2.0.CO;2

Sun, C., Thresher, A., Keeley, R., Hall, N., Hamilton, M., Chinn, P., et al. (2010). The data management system for the Global Temperature and Salinity Profile Programme. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (pp. 931–938). European Space Agency. https://doi.org/10.5270/OceanObs09.cwp.86

Tietsche, S., Balmaseda, M., Zuo, H., Roberts, C., Mayer, M., & Ferranti, L. (2020). The importance of North Atlantic Ocean transports for seasonal forecasts. *Climate Dynamics*, *55*(7–8), 1995–2011. https://doi.org/10.1007/s00382-020-05364-6

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., et al. (2018). El Niño–Southern Oscillation complexity. *Nature*, *559*(7715), 535–545. https://doi.org/10.1038/s41586-018-0252-6

Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., et al. (2014). On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, *27*(21), 7994–8016. https://doi.org/10.1175/JCLI-D-14-00158.1

Vidard, A., Anderson, D. L. T., & Balmaseda, M. (2007). Impact of ocean observation systems on ocean analysis and seasonal forecasts. *Monthly Weather Review*, *135*(2), 409–429. https://doi.org/10.1175/MWR3310.1

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*, L12702. https://doi.org/10.1029/2009GL038710

Wittenberg, A. T., Rosati, A., Delworth, T. L., Vecchi, G. A., & Zeng, F. (2014). ENSO modulation: Is it decadally predictable? *Journal of Climate*, *27*(7), 2667–2681. https://doi.org/10.1175/JCLI-D-13-00577.1

Wittenberg, A. T., Rosati, A., Lau, N.-C., & Ploshay, J. J. (2006). GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *Journal of Climate*, *19*(5), 698. https://doi.org/10.1175/JCLI3631.1

Xue, Y., Chen, M., Kumar, A., Hu, Z.-Z., & Wang, W. (2013). Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System Version 2. *Journal of Climate*, *26*(15), 5358–5378. https://doi.org/10.1175/JCLI-D-12-00600.1

Yang, X., DelSole, T., & Pan, H.-L. (2008). Empirical correction of the NCEP global forecast system. *Monthly Weather Review*, *136*(12), 5224–5233. https://doi.org/10.1175/2008MWR2527.1

Yang, X., Jia, L., Kapnick, S. B., Delworth, T. L., Vecchi, G. A., Gudgel, R., et al. (2018). On the seasonal prediction of the western United States El Niño precipitation pattern during the 2015/16 winter. *Climate Dynamics*, *51*(9–10), 3765–3783. https://doi.org/10.1007/s00382-018-4109-3

Yang, X., Rosati, A., Zhang, S., Delworth, T. L., Gudgel, R. G., Zhang, R., et al. (2013). A predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system. *Journal of Climate*, *26*(2), 650–661. https://doi.org/10.1175/JCLI-D-12-00231.1

Zhang, G., Murakami, H., Gudgel, R., & Yang, X. (2019). Dynamical seasonal predictions of tropical cyclone activity: Roles of sea surface temperature errors and atmosphere/land initialization. *Geophysical Research Letters*, *46*, 5506–5515. https://doi.org/10.1029/2019gl082529

Zhang, J., & Rothrock, D. A. (2003). Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Monthly Weather Review*, *131*(5), 845–861. https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. T. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, *135*(10), 3541–3564. https://doi.org/10.1175/MWR3466.1

Zhang, S., & Rosati, A. (2010). An inflated ensemble filter for ocean data assimilation with a biased coupled GCM. *Monthly Weather Review*, *138*(10), 3905–3931. https://doi.org/10.1175/2010mwr3326.1

Zhang, S., Rosati, A., & Harrison, M. J. (2009). Detection of multidecadal oceanic variability by ocean data assimilation in the context of a "perfect" coupled model. *Journal of Geophysical Research*, *114*, C12018. https://doi.org/10.1029/2008JC005261

Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018a). The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, *10*, 691–734. https://doi.org/10.1002/2017MS001208

Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018b). The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*, 735–769. https://doi.org/10.1002/2017MS001209

Zuo, H., Alonso-Balmaseda, M., Mogensen, K., & Tietsche, S. (2018). OCEAN5: The ECMWF Ocean Reanalysis System and its real-time analysis component. ECMWF Technical Memorandum. https://doi.org/10.21957/la2v0442

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, *15*(3), 779–808. https://doi.org/10.5194/os-15-779-2019