The Navy's Earth System Prediction Capability: A New Global Coupled Atmosphere-Ocean-Sea Ice Prediction System Designed for Daily to Subseasonal Forecasting

Neil Barton[1] , E. Joseph Metzger[2] , Carolyn A. Reynolds[1] , Benjamin Ruston[1], Clark Rowley[2] , Ole Martin Smedstad[3] , James A. Ridout[1] , Alan Wallcraft[4], Sergey Frolov[1,5] , Patrick Hogan[6] , Matthew A. Janiga[1], Jay F. Shriver[2] , Justin McLay[1] , Prasad Thoppil[2], Andrew Huang[7] , William Crawford[1] , Timothy Whitcomb[1], Craig H. Bishop[8] , Luis Zamudio[4], and Michael Phelps[3]

[1]Naval Research Laboratory, Marine Meteorology Division, Monterey, CA, USA, [2]Naval Research Laboratory, Ocean Sciences Division, Stennis Space Center, MS, USA, [3]Perspecta, Stennis Space Center, MS, USA, [4]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA, [5]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO, USA, [6]National Oceanic and Atmospheric Administration, National Centers for Environmental Information, Stennis Space Center, MS, USA, [7]Science Applications International Corporation (SAIC), Monterey, CA, USA, [8]School of Earth Sciences and ARC Centre of Excellence for Climate Extremes, University of Melbourne, Parkville, Victoria, Australia

Abstract This paper describes the new global Navy Earth System Prediction Capability (Navy-ESPC) coupled atmosphere-ocean-sea ice prediction system developed at the Naval Research Laboratory (NRL) for operational forecasting for timescales of days to the subseasonal. Two configurations of the system are validated: (1) a low-resolution 16-member ensemble system and (2) a high-resolution deterministic system. The Navy-ESPC ensemble system became operational in August 2020, and this is the first time the NRL operational partner, Fleet Numerical Meteorology and Oceanography Center, will provide global coupled atmosphere-ocean-sea ice forecasts, with atmospheric forecasts extending past 16 days, and ocean and sea ice ensemble forecasts. A unique aspect of the Navy-ESPC is that the global ocean model is eddy resolving at 1/12° in the ensemble and at 1/25° in the deterministic configurations. The component models are current Navy operational systems: NAVy Global Environmental Model (NAVGEM) for the atmosphere, HYbrid Coordinate Ocean Model (HYCOM) for the ocean, and Community Ice CodE (CICE) for the sea ice. Physics updates to improve the simulation of equatorial phenomena, particularly the Madden-Julian Oscillation (MJO), were introduced into NAVGEM. The low-resolution ensemble configuration and high-resolution deterministic configuration are evaluated based on analyses and forecasts from January 2017 to January 2018. Navy-ESPC ensemble forecast skill for large-scale atmospheric phenomena, such as the MJO, North Atlantic Oscillation (NAO), Antarctic Oscillation (AAO), and other indices, is comparable to that of other numerical weather prediction (NWP) centers. Ensemble forecasts of ocean sea surface temperatures perform better than climatology in the tropics and midlatitudes out to 60 days. In addition, the Navy-ESPC Pan-Arctic and Pan-Antarctic sea ice extent predictions perform better than climatology out to about 45 days, although the skill is dependent on season.

1. Introduction

Traditional boundaries between weather and climate prediction have been recognized as artificial (e.g., Hurrell et al., 2009; Shapiro et al., 2010) given the wide range of interactions between Earth system components on many timescales. This has motivated the seamless earth-system prediction approach, where the same or similar earth system models are used for both short and long lead-time forecasts (Brunet, 2010; Ruti et al., 2019). The importance of atmosphere-ocean-ice coupling for monthly and seasonal forecasts is well recognized (MacLachlan et al., 2015; Molteni, 2011). More recently, operational centers such as the European Centre for Medium-range Weather Forecasts (ECMWF), U.K. Met Office, and Environment and Climate Change Canada (ECCC) are also finding benefits in ocean coupling for lead times as short as a few days (Smith, 2018), which further supports the seamless approach to environmental prediction.

Along these lines, the U.S. Naval Research Laboratory (NRL) has developed a new global coupled data assimilation and prediction system, the Navy Earth System Prediction Capability (Navy-ESPC), with the goals of both improving short-term forecasts and extending forecasts to multiweek timescales. The Navy-ESPC effort is part of the U.S. National ESPC effort, which is a five-agency effort to coordinate and accelerate the national numerical environmental prediction capability (Carman, 2017). As noted in that article and others, there are many needs for environmental predictions over various timescales. U.S. Navy applications that would benefit from skillful environmental predictions include the planning and conducting of Navy ship sorties on timescales less than a week, ship routing and prepositioning on timescales out to a month, and humanitarian assistance planning and force deployment management on monthly to seasonal timescales. Unlike many other forecasting centers, the U.S. Navy is interested in high-fidelity forecasts of the ocean itself, including the ocean interior. As described below, this need for accurate ocean forecasts drives Navy-ESPC design, and the high-horizontal resolution ocean and sea-ice components of Navy-ESPC distinguishes it from other global coupled prediction systems.

Navy-ESPC is built upon existing atmosphere and ocean/sea ice operational systems that are being used at Fleet Numerical Meteorology and Oceanography Center (FNMOC). The atmospheric model is the NAVy Global Environmental Model (NAVGEM) (Hogan, 2014), and the ocean and sea ice models are part of the Global Ocean Forecasting System version 3.1 (GOFS 3.1) (Metzger et al., 2014). Navy-ESPC is unique compared to many other global coupled modeling systems (Infanti & Kirtman, 2016; Lin et al., 2016; Saha, 2014; Sun et al., 2018; Vitart et al., 2017; Zhu, 2018) in that the ocean model horizontal resolution is eddy resolving in both the ensemble (i.e., probabilistic) and deterministic configurations. In many operational coupled modeling systems, the ocean and sea ice components are included primarily to provide better atmospheric bottom/surface boundary conditions but only at eddy-permitting resolution (~1/4° or coarser). However, users of Navy-ESPC products need high-resolutions predictions from each component.

A high-resolution ocean model is needed to simulate ocean energetics (Chassignet & Xu, 2017; Hogan & Hurlburt, 2000; Thoppil et al., 2011) as the Rossby radius of deformation is much smaller in the ocean than in the atmosphere (Feliks, 1985; Hallberg, 2013). Thoppil et al. (2011) compared ocean modeling results from a 1/12° and 1/25° ocean and found that the Eddy Kinetic Energy (EKE) in the 1/25° nonassimilative simulation and 1/12° data assimilative reanalysis matched observations more closely than the 1/12° nonassimilative simulation. Chassignet and Xu (2017) examined Gulf Stream separation in ocean models ranging from 1/4° to 1/50° and determined that the Gulf Stream was simulated most accurately in the 1/50° ocean. In examining the impact of high-resolution ocean modeling on atmosphere predictions, Hewitt (2017) determined that the high-resolution ocean was beneficial to the atmosphere near western boundary currents.

The main objective of this paper is to describe the various components of the Navy-ESPC system and highlight key metrics used in the validation and verification process as it was transitioned to FNMOC. Due to resource limitations, the changes between the coupled system and the current operational systems could not be tested independently. The purpose of this article is to provide a detailed description of the new system, which will serve as a reference for future hypotheses-driven papers and other studies that may use the system output. The modeling components, coupling technologies, and data assimilation in sections 2 and 3; forecasts used for the diagnostics in section 4; results in section 5; and conclusions and future directions in section 6.

2. Navy-ESPC

The global atmosphere component of Navy-ESPC is NAVGEM (Hogan, 2014) and the ocean and sea ice components are the HYbrid Coordinate Ocean Model (HYCOM) (Bleck, 2002) and Community Ice CodE (CICE) (Hunke &> Lipscomb, 2008), respectively (Figure 1). HYCOM coupled to CICE currently runs at FNMOC as stand-alone GOFS 3.1. NAVGEM and HYCOM are developed at NRL and CICE is from the Los Alamos National Laboratory. These models, which are also referred to as components, are coupled together using Earth System Module Framework (ESMF) tools, in conjunction with the National Unified Operational Prediction Capability (NUOPC) layer (Theurich, 2016).
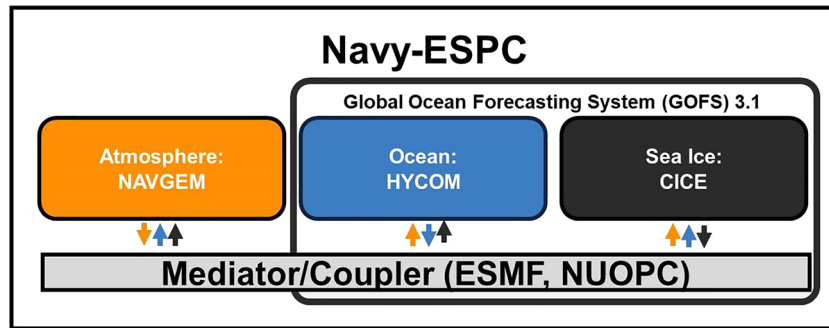
Figure 1. Schematic of the Navy-ESPC modeling system. The atmospheric, ocean, and sea ice model components are the NAVy Global Environmental Model (NAVGEM), the HYbrid Coordinate Ocean Model (HYCOM), and the Community Ice CodE (CICE), respectively. HYCOM and CICE are part of the operational Global Ocean Forecasting System version 3.1 (GOFS 3.1). To couple the models, the Earth System Module Framework (ESMF) tools are used in conjunction with the National Unified Operational Prediction Capability (NJUOPC) framework. The colored arrows represent variables that export and import from each system. Note the land model is imbedded inside of NAVGEM.

2.1. NAVGEM

NAVGEM (Hogan, 2014) is the Navy's global weather prediction system, run operationally at FNMOC and is used as the atmospheric component in Navy-ESPC. The forecast model uses both grid point and spectral (i.e., spherical harmonic) representations to perform the forecast. Grid point calculations are performed for the Semi-Lagrangian (SL) advection and in all physical parameterizations. Calculations in spectral space are performed for the Semi-Implicit (SI) corrections to the divergent component of the winds, virtual potential temperature, and surface pressure. The dynamical core of NAVGEM is a three-time level, SL/SI numerical integration of the hydrostatic equations of motion and the first law of thermodynamics. The SL/SI formulation was first implemented at ECMWF based on the work of Ritchie (1987, 1988, 1991) and Ritchie et al. (1995). The major difference between the Ritchie and NAVGEM formulations is that NAVGEM uses potential temperature while ECMWF uses temperature.

The Navy-ESPC uses two different resolutions: for NAVGEM, the high resolution deterministic (Navy-ESPC$_{DET}$) is T681L60 (~18 km horizontal resolution) and the ensemble resolution (Navy-ESPC$_{ENS}$) is T359L60 (~37 km horizontal resolution) (Table 1). The dynamics framework utilized for the Navy-ESPC$_{DET}$ is slightly different from the Navy-ESPC$_{ENS}$ configuration. NAVGEM in Navy-ESPC$_{ENS}$ followed then operational stand-alone NAVGEM 1.2 (T359L50) in retaining the virtual potential temperature as the prognostic temperature variable, and using the full Gaussian grid. The adiabatic correction option introduced in NAVGEM 1.2 is turned off for the ensemble resolution. The adiabatic correction option in NAVGEM is a method to address the lack of conservation of mass-weighted virtual potential temperature, which occurs largely in the stratosphere in NAVGEM. Mass loss/gain is added/subtracted back into model. NAVGEM in Navy-ESPC$_{DET}$ followed then operational stand-alone NAVGEM 1.4 (T425L60) and uses the three-time level SL/SI dynamical core, perturbation virtual potential temperature for the prognostic temperature variable, and uses the reduced Gaussian grid. A 60-level vertical grid was used in both horizontal resolutions in Navy-ESPC. This is the same vertical resolution used in the recent operational version of NAVGEM.

For the implementation of NAVGEM into NAVY-ESPC, we developed a new suite of model physics referred to as Coupled Version Physics (CVP) (Table 2). The CVP physics have been tailored for the coupled system with two primary objectives: (1) to improve the representation of the Madden-Julian Oscillation (MJO), which is recognized as a key contributor to extended range predictability (Waliser, 2009), and (2) to improve the consistency between surface fluxes computed in NAVGEM and in HYCOM. The MJO is a mode of atmospheric variability in the tropics with a timescale of 20–100 days and is one of the few phenomena affecting global predictability on extended-range timescales (Lim et al., 2018). The phase of the MJO is known to have substantial impact on many other phenomena such as tropical cyclone genesis and frequency (Camargo et al., 2008; Kim et al., 2008; Sobel & Maloney, 2000), atmospheric rivers (Bond & Vecchi, 2003; Guan et al., 2012; Mundhenk et al., 2016), and even the atmospheric circulation in the Arctic and Antarctic (Flatau & Kim, 2013; Henderson et al., 2014, 2018; Lin et al., 2009). As such, it is important for extended-range forecasting systems to be capable of realistic simulations and accurate predictions of the MJO.

Table 1
Configurations of the Components/Models in the High Resolution Deterministic (Navy-ESPC$_{DET}$) and Ensemble (Navy-ESPC$_{ENS}$)

Navy-ESPC configuration		Navy-ESPC$_{DET}$	Navy-ESPC$_{ENS}$
NAVGEM	Resolution	T681 60 Levels	T359 60 Levels
	Parameterization Differences	*Base Version:* NAVGEM 1.4 (resolution T425L60)	*Base Version:* NAVGEM 1.2 (resolution T359L50)
		Prognostic Temperature Variable: perturbation virtual potential temperature	*Prognostic Temperature Variable:* virtual potential temperature
			Grid Type: Gaussian
		Grid Type: reduced Gaussian grid	*Adiabatic Correction:* None
		Adiabatic Correction: Yes	
HYCOM	Resolution	1/25° tripole 41 layers	1/12° tripole 41 layers
	Parameterization Differences	*Tides:* M2,S2,K1,O1,N2	*Tides:* none
CICE	Resolutions	1/25° tripole	1/12° tripole grid
		4 ice categories	4 ice categories
	Parameterization Differences	No differences	No differences

Note. NAVGEM versions in Navy-ESPC$_{DET}$ and Navy-ESPC$_{ENS}$ both use the physics changes listed in.

To improve the representation of the MJO in Navy-ESPC, we replaced the operational NAVGEM Simplified Arakawa-Schubert (SAS) convection scheme (Han & Pan, 2011) with a modified version of the Kain-Fritsch scheme (Kain & Fritsch, 1990, 1993). The new scheme is an extension of Ridout et al. (2005) and incorporates both turbulence and dynamically forced modes. As in the 2005 version, the scheme includes a modified closure formulation for the Kain-Fritsch dynamically forced mode based on an assumed quasi-balance of updraft parcel buoyancy at the cloud base level. This closure formulation requires the scheme to be called at every time step to adjust the cloud base mass flux in a similar manner as in the Emanuel convection scheme (Emanuel, 1991; Emanuel & Zivkovic-Rothman, 1999). For the turbulence-forced mode of convection, triggering by boundary layer plumes is represented in part using the mixed-layer Richardson number convective trigger formulation described by Ridout and Reynolds (1998). Other added enhancements for Navy-ESPC include a treatment of convective momentum transport and a modified representation of the rate of updraft-environment mixing adapted from Peng et al. (2004).

In addition to the convection scheme changes, the other update for the CVP physics is in making the fluxes more consistent between the atmosphere and ocean. The HYCOM surface flux scheme is the Kara et al. (2005) adaptation of the COARE 3.0 scheme of Fairall et al. (2003) and is incorporated in NAVGEM within Navy-ESPC. The coupled system framework enables for the first time in NAVGEM the inclusion of surface currents in the surface flux computation. Navy-ESPC versions of NAVGEM and HYCOM also account for depression of the saturated vapor pressure due to salinity using the correction of Sud and Walker (1997). See Figure 2 for the list of variables coupled to NAVGEM.

Table 2
Summary of NAVGEM Physics Changes Between NAVGEM v1.4 and the NAVGEM Used in the Navy-ESPC system

Parameterization/ scheme	NAVGEM v1.4	NAVGEM in Navy-ESPC
Convection parameterization	SAS (Moorthi et al., 2001)	Modified Kain-Fritsch
Boundary layer scheme	Louis et al. (1982)	COARE (Kara et al., 2005)

Note. The main changes between the two atmospheric models include changes in the convection parameterization and boundary layer scheme. The Modified Kain-Fritsch is based on Ridout et al. (2005).

The CVP has contributed to an improved representation of the MJO with respect to stand-alone NAVGEM. Stand-alone NAVGEM's representation of the MJO was not competitive compared to other global models, as shown in Jiang (2015). With the implementation of the CVP and coupling, the Navy-ESPC model compares favorably in this regard to other state-of-the-art forecast as seen through evaluation of the 17+ years of forecasts in Janiga et al. (2018) and Kim et al. (2019). These 17+ years reforecasts were part of the National Oceanic and Atmospheric Administration (NOAA) Subseasonal eXperiment (SubX) (Pegion, 2019). The CVP physics contributes to this notable improvement, as does the atmosphere-ocean coupling in Navy-ESPC. Figure 3 displays Hovmöller diagrams of precipitation during the Dynamics of the Madden-Julian Oscillation (DYNAMO) observation campaign (Johnson & Ciesielski, 2013) for operational stand-alone NAVGEM

Figure 2. List of fields passed to and from the model components (i.e., variables that are coupled). The color represents the component/model the field is from and the arrows represent if the fields are going into or out of the component. The long names for the atmospheric fields include: U @ 10 m = zonal winds at 10 m; V @ 10 m = meridional winds at 10 m; Temp @ 2 m = air temperature at 2 m; Spec Hum @ 2 m = specific humidity at 2 m; Density @ 2 m = air density at 2 m; Temp @ Sur = temperature of the surface (ground, ice, ocean); SW down = downward surface shortwave radiative flux; LW down = downward surface longwave radiative flux; Precipitation = total precipitation; MSLP = mean sea level pressure. The long names for the ocean variables include the following: SST = sea surface temperature; SSU = zonal sea surface current; SSV = meridional sea surface current; SSS = sea surface salinity. The long names for the sea ice variables include the following: SIC = sea ice concentration; X Ice Stress @ B = x direction ice stress at sea ice basal; Y Ice Stress @ B = y direction ice stress at sea ice basal; SW Flux @ B = shortwave radiative flux at sea ice basal; LW Flux @ B = longwave radiative flux at sea ice basal; Salt Flux @ B = flux of salt at sea ice basal; Water Flux @ B = flux of fresh water at sea ice basal; Ice U = zonal velocity of sea ice at the surface; Ice V = meridional velocity of sea ice at the surface; Ice Thick = sea ice thickness; Ice Albedo = sea ice albedo; Ice Temp = temperature of sea ice at surface.

v1.4, NAVGEM v1.4 with the CVP option, and Navy-ESPC. As shown, operational NAVGEM v1.4 did not represent the first or second MJO for this time period. With addition of the CVP physics, precipitation for
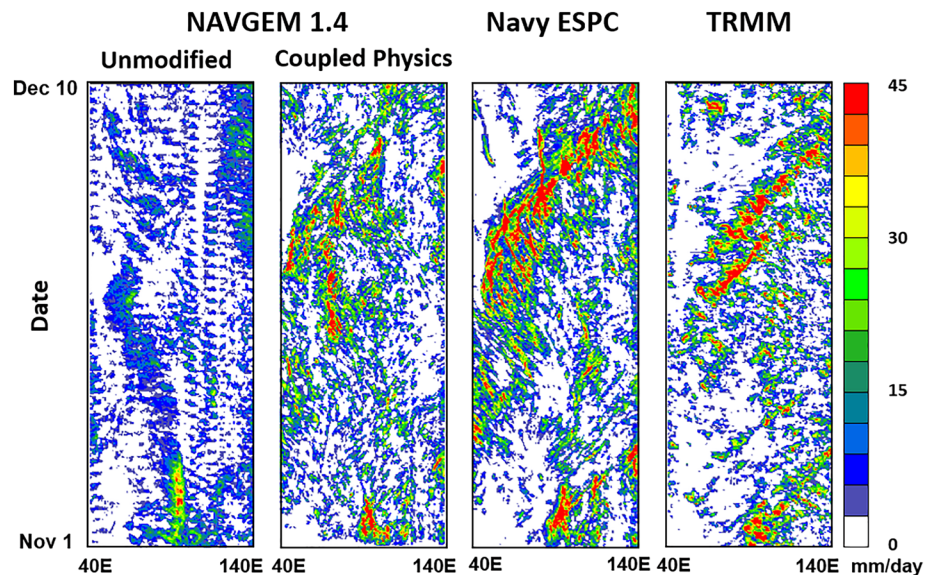
Figure 3. Hovmöller diagram of precipitation from NAVGEM v1.4 with NAVGEM physics, NAVGEM v1.4 with the coupled physics, Navy-ESPC, and The Tropical Rainfall Measuring Mission (TRMM) from 1 November 2011 to 10 December 2011. Precipitation is averaged from 5°S to 5°N and the Hovmöller diagram shows precipitation from 40°E to 140°E. NAVGEM and Navy-ESPC output represent long nonassimilative forecasts starting on 1 November 2011. NAVY-ESPC is configured as a deterministic Navy-ESPC$_{ENS}$ run with 50 vertical levels.

this period is more similar to the observations than without the CVP physics, and including coupling within Navy-ESPC further improves the representation of the MJO precipitation.

2.2. HYCOM

The HYCOM is the ocean model in Navy-ESPC. HYCOM is a primitive equation ocean general circulation model capable of nowcasting and forecasting the three-dimensional temperature, salinity, and current structure of the global ocean. It employs potential density referenced to 2,000 m and includes the effects of thermobaricity (Chassignet et al., 2003). HYCOM's grid is uniform cylindrical from 78.64°S to 66.0°S, on a Mercator projection from 66.0°S to 47°N and curvilinear north of this as it employs an Arctic dipole patch where the poles are shifted over land to avoid a singularity at the North Pole. This grid is referred to as a tripole grid.

HYCOM uses a mixture of vertical coordinates, hence the HYbrid as part of its name. Vertical coordinates can be (1) isopycnals (density tracking), often the best coordinate in the deep stratified ocean; (2) levels of equal pressure (nearly fixed depths), best used in the mixed layer and unstratified ocean; and (3) sigma-levels (terrain following), often the best choice in shallow water. HYCOM combines all three approaches by choosing the optimal distribution at every time step. The model makes a dynamically smooth transition between coordinate types by using the layered continuity equation. The hybrid coordinate extends the geographic range of applicability of traditional isopycnic coordinate circulation models toward shallow coastal seas and unstratified parts of the world ocean. It maintains the significant advantages of an isopycnal model in stratified regions while allowing more vertical resolution near the surface and in shallow coastal areas, hence providing a better representation of the upper ocean physics.

HYCOM is configured with options for a variety of mixed layer submodels (Halliwell, 2004) and here uses the K-Profile Parameterization (Large et al., 1994). A more complete description of HYCOM physics can be found in Bleck (2002). The equatorial horizontal resolution of HYCOM is 1/25° in Navy-ESPC$_{DET}$ and 1/12° in Navy-ESPC$_{ENS}$, and both configurations have 41 layers (Table 1). The number of vertical layers is a compromise between accuracy and computational efficiency when run in an operational setting.

HYCOM versions in Navy-ESPC$_{DET}$ and Navy-ESPC$_{ENS}$ are identical except Navy-ESPC$_{DET}$ turns on astronomical tidal forcing that generates internal tides (Arbic et al., 2018) and the Navy-ESPC$_{ENS}$ does not. HYCOM implementation in stand-alone operational GOFS 3.1 is very similar to that in Navy-ESPC, and the main difference is that the Navy-ESPC HYCOM uses the coupled sea level pressure forcing instead of constant sea level pressure forcing as used in GOFS 3.1. This change in sea level pressure forcing is needed for the internal tides in Navy-ESPC$_{DET}$.

2.3. CICE

Version 4.0 of the Los Alamos-developed CICE model (Hunke & Lipscomb, 2008) is the sea ice component of Navy-ESPC. CICEv4 is also currently used in stand-alone operational GOFS 3.1. CICE includes sophisticated ice thermodynamics such as multiple ice thickness layers, multiple snow layers, and the capability to forecast multicategories of ice thickness according to World Meteorological Organization definitions. In addition, CICE has several interacting components including a thermodynamic model that computes local growth rates of snow and ice due to snowfall, vertical conductive, radiative, and turbulent fluxes; a model of ice dynamics that predicts the velocity field of the ice pack based on a model of the material strength of the ice; a transport model that describes advection of the areal concentration, ice volumes, and other state variables; and a ridging parameterization that transfers ice among thickness categories based on energetic balances and rates of strains. The CICE horizontal grid for Navy-ESPC$_{DET}$ and Navy-ESPC$_{ENS}$ is the same as HYCOM (Table 1) for conservation of variables when coupling. There are no differences in CICE physics between the deterministic and ensemble configurations of Navy-ESPC. A more recent version of CICE (i.e., version 5 or 6) was not used in the Navy-ESPC code simply because the code needed to be locked for testing to meet timelines for transition to operations.

2.4. ESMF/NUOPC

Navy-ESPC uses the NUOPC tools wrapped around ESMF for coupling. This framework allows the components to dynamically interact at a specified coupling time step. A 1-hr coupling time step is implemented in Navy-ESPC. We tested a more frequent time step; however, the results were not improved enough to warrant

the additional computational cost. As the analyzed ocean and sea-ice fields are held constant during the stand-alone NAVGEM integrations, the coupled system holds the promise of improved surface forcing for the atmosphere, particularly for forecast lead times greater than a few days. Figure 2 displays the coupling variables in Navy-ESPC. Each model component performs its own flux calculations. The implementation of a single coordinated flux calculation was considered, but found to be impractical for this initial rollout of Navy-ESPC with time constraints set by our operational partners. NAVGEM's use of a fully implicit mixing scheme increases the computational demands in this context, as does the large mismatch between the high-resolution ocean grid and the lower-resolution atmosphere grid, which results in an inefficient interpolation of variables. Mixing in NAVGEM's scheme is carried out through the depth of the atmospheric column in conjunction with the computation of surface fluxes. Surface exchange coefficients are computed based on Kara et al. (2005), and then a tridiagonal system is solved to obtain the mixing throughout the column, including the surface fluxes. To perform this calculation in the mediator, three-dimensional fields would need to be interpolated onto the high-resolution mediator grid, which is the same as the HYCOM grid, and ideally be performed at every NAVGEM time step (5–6 min). With the mismatch between the high-resolution HYCOM grid and the NAVGEM grid, this interpolation was deemed to be too computationally expensive and memory intensive. The lack of a consistent flux calculation between components results in the loss of conservation. However, note that our forecasts do not go beyond 45 days, and our experience with stand-alone NAVGEM has shown that the implicit approach provides a skill advantage for numerical weather prediction, one that we wish to retain.

Another difficulty in the interpolations is that the ESMF remapping routines initially led to grid points with undefined values near coastlines. In Navy-ESPC, the land/ocean interpolation boundary is complicated thanks to differences in coastline representation in NAVGEM and HYCOM driven by (1) differences in horizontal resolution and (2) the definition of the land or ocean point. For example, the HYCOM land/ocean boundary is defined by a depth in the ocean in which HYCOM will begin calculations, and the NAVGEM land/ocean boundary is defined where surface fluxes are greatly altered by the surface state, which is generally at a shallower ocean depth than HYCOM's land/ocean boundary. With the ESMF remapping routines, unassigned points occurred in the HYCOM to NAVGEM and NAVGEM to HYCOM interpolations, and we developed a routine to address this shortcoming. First, we find the grid points that do not have values (from the default ESMF scheme), and call these "hole cells." Next, we create a stencil of grid points around the hole cell and assign each stencil point as land or ocean. If the hole cell is ocean, the ocean points in the stencil are used to interpolate a value to the hole grid cell. If the hole cell is land, land points in the stencil are used for the interpolation. In our new routine, the stencil size and the number of times to search for hole cells are user defined and is only used at the start of the model simulation. A simple implementation of this algorithm has been added into version 8 of ESMF called ESMF_EXTRAPMETHOD_CREEP.

3. Weakly Coupled Data Assimilation and Ensemble Design

The Navy-ESPC system utilizes two mature assimilation code bases that have been developed for GOFS 3.1 and NAVGEM forecast systems separately: the Navy Coupled Ocean Data Assimilation (NCODA) for the ocean and sea-ice, and the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System Accelerated Representer (NAVDAS-AR) for the atmosphere.

3.1. NCODA: Ocean and Sea Ice Data Assimilation

NCODA is a three-dimensional, multivariate (3DVAR) Data Assimilation (DA) scheme (Cummings, 2005; Cummings & Smedstad, 2013) for the ocean/ice variables temperature, salinity, geopotential, vector velocity components, and sea ice concentration. All observations are processed at FNMOC through a quality control system that checks for gross error and assigns each observation a likelihood of error based on comparison to climatology and ocean analyses. Observation errors are assigned by observation type and include both an instrument error and a representativeness error that account for the observation spatial sampling and the model grid resolution. Sea surface height anomaly (SSHA), sea surface temperature (SST), and ice concentration observations are pooled into superobservations at the resolution of the model grid before assimilation. Along-track SSHA observations from available altimeters referenced to a 1993–2015 mean height are used to derive synthetic subsurface temperature and salinity profiles via the Improved Synthetic Ocean

Profile (ISOP) methodology of Helber et al. (2013), and these are assimilated together with satellite and in situ SSTs from ships, and moored and drifting buoys. Profile temperature and salinity observations come from Argo profiling floats, XBTs, CTDs, gliders, and marine mammals. Satellite sea ice concentration is assimilated in the ice analysis. Ocean data assimilation windows are typically longer than in atmospheric application, and subsurface ocean observations provide useful information for days after the observation time. The data windows for profile and SSHA observations may be extended back in time both to capture late-arriving observations and to include a larger set of observations more capable of capturing the ocean mesoscale structure. While the atmosphere is "data rich," the ocean is "data poor" with regard to having enough subsurface observations to accurately constrain the smaller scale (than the atmosphere) mesoscale features. We use a 5-day window for in situ profile observations and a 4-day window of satellite SSHA observations. For satellite and in situ SST and satellite ice concentration, the data window extends back only to the most recent observation assimilated previously. We do not presently apply bias corrections to the satellite SSTs. SST bias corrections may be explored in the future as these corrections have been shown to aid in surface flux estimates (While & Martin, 2019). The data are compared against a time dependent background field using the First Guess at Appropriate Time (FGAT) method (Cummings & Smedstad, 2014; Fox et al., 2002).

NCODA can be run in stand-alone mode but here is cycled with HYCOM and CICE to provide updated initial conditions for the next model forecast using an incremental analysis update (IAU) procedure (Bloom et al., 1996). In an uncoupled model (i.e., GOFS 3.1), the first guess is generated by the ocean model forecast from the end of the last DA window to the end of the current DA window.

3.2. NAVDAS-AR: Atmosphere Data Assimilation

NAVDAS-AR (Rosmond & Xu, 2006; Xu et al., 2005) has the ability for both weak and strong constraint variational assimilation, and is formulated in the terms of dual variables (i.e., observation space), the dimension of which is generally much smaller than the corresponding state (i.e., model) space. This restriction of the observation space is accomplished by discarding the unobservable degrees of freedom in the system (Bennett, 2002).

NAVDAS-AR can process well over 100 million observations in every 6-hr data assimilation window. After quality control and data thinning it assimilates approximately 4 million observations to create the final analysis. Within the 4DVAR data assimilation cycle, NAVDAS-AR approximates the full nonlinear atmosphere in the Navy-ESPC coupled forecast with a simplified version of the stand-alone NAVGEM system to allow for a computationally efficient solution of the assimilation problem.

3.3. Weakly Coupled Dual Cycle Data Assimilation

Navy-ESPC implements a weakly coupled formulation of the coupled data assimilation system with two separate windows for the atmosphere and ocean. In a weakly coupled data assimilation, the coupled model forecast is used as a first guess for the independent data assimilation systems. Given this coupled first guess, the ocean (NCODA) and atmospheric (NAVDAS-AR) data assimilation systems compute the assimilation increment independently, ignoring any potential cross-correlations between ocean and the atmosphere. To simultaneously accommodate the 6-hr DA cycle that NAVDAS-AR has been designed for and the 24-hr DA cycle that has proven optimal NCODA, a new dual cycle DA has been developed (Figure 4, Table 3).

The centers for the 6-hr atmospheric DA widows are on 00Z, 06Z, 12Z, and 18Z. Coupled forecast from the previous update cycle is performed to at least 3 hr after the center point of the next update cycle, with the 6-hr section of the forecast centered on the next analysis time used as the first guess to initialize the atmospheric portion of the coupled model for the subsequent center point of the next update cycle. For example, from the 00Z coupled forecast, we use forecast hours valid between 03Z and 09Z as the first guess for the initialization at 06Z. Next, we run NAVDAS-AR on this first guess field to obtain the analysis increments/corrections valid at 06Z. The correction is then inserted into the coupled model by IAU over the 3 hr prior to the analysis time. For example, for the 06Z centered window, the coupled forecast starts at 03Z with the increment incorporated over the 03Z to 06Z using IAU. Thus, 12-hr coupled forecasts are needed for each atmospheric DA cycle, 3 hr needed for IAU plus the 9 hr needed for the next first guess. The 3-hr IAU in NAVDAS-AR is a new capability implemented specifically for the Navy-ESPC as the FNMOC
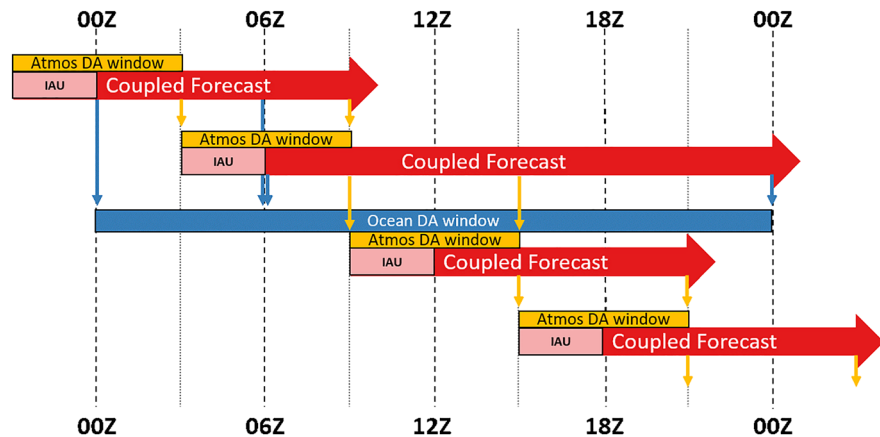
Figure 4. Schematic of the weakly coupled DA system for Navy-ESPC. The red arrows represented the Navy-ESPC first guess/forecast from the coupled model. The yellow bars represent the observational window used for NAVDAS-AR and the blue bar represents the ocean and sea ice observational window used for NCODA. The light red boxes represent the 3-hr IAU period when the increments produced by NCODA and NAVDAS-AR are inserted into the Navy-ESPC coupled forecast. The yellow and blue arrows pointing down from the coupled forecasts represent the times of the forecast that are used for first guess for (yellow) NAVDAS-AR and (blue) NCODA. Note that NCODA appends the forecast together at 00Z and 06Z for its first guess. Also note that sea ice increments are directly inserted at 09Z.

operational NAVDAS-AR implementation does not use an IAU. The forecasts started at 06Z are run 12 hr longer than the forecasts started at 00Z, 12Z, and 18Z to provide a first guess for the 24-hr ocean DA window centered at 12Z, described in the next paragraph. Unlike ECMWF's system (Laloyaux et al., 2016), but similar to our stand-alone atmospheric system, Navy-ESPC uses only a single outer-loop in the atmospheric NAVDAS-AR system.

The NCODA ocean DA window center is at 12Z and uses a 24-hr window that is 12 hr before and after the 12Z center (i.e., 0 to 24Z) with an extended look backward for late arriving observations as described above. The first guess for NCODA combines the 0 to 6 hr coupled forecast center at 00Z, and the 0 to 18 hr coupled forecast at 06Z, which corresponds to times 06Z to 24Z. As in the atmosphere, a 3-hr IAU is used to update the ocean at the 12Z center point, which corresponds to the coupled model forecast starting 3 hr before 12Z, at 9Z. For the sea ice update, the analysis produced by NCODA is directly inserted into CICE at 09Z without any incremental insertion. The ocean does not update with increments/corrections at 00Z, 06Z, and 18Z.

Using the IAU method for both the atmosphere and ocean leads to the coupled model transferring some of the initialization information from the atmospheric observations to the ocean and vice versa during the IAU period. Since both the ocean and atmosphere are updated at 12Z, only the forecasts initialized at 12Z are used for the forecast verification.

To manage the complexity of the weakly coupled data assimilation system, such as task dependencies, we rely on the Cylc workflow manager (Oliver, 2019). The workflow developed for Navy-ESPC represents the combination of workflows developed for each standalone component, with the coupled forecast model as a common component. Additionally, the ensemble configuration makes use of independent workflow execution for each ensemble member.

3.4. Ensemble of Data Assimilations

To generate initial conditions for the subseasonal ensemble forecast system, we employ the ensemble of data assimilations approach (EDA) (Houtekamer et al., 1996). In our implementation, one member is used

Table 3
Description of Atmosphere and Ocean/Sea Ice Data Assimilation Systems Used in Navy-ESPC

	NAVDAS-AR	NCODA
Stand-Alone and Weakly Coupled Implementation		
Forecast Model(s):	NAVGEM	HYCOM and CICE
Method:	4DVAR Hybrid	3DVAR FGAT
Update Window:	6 hr (±3 hr)	24 hr (±12 hr)
Stand-Alone Implementation		
Updating Method:	Direct Insertion	3-hr IAU
Weakly Coupled Implementation		
Updating Method:	3-hr IAU	3-hr IAU

as a control member and is cycled using the ensemble-resolution version of the Navy-ESPC system. In addition to the control member, we also cycle 15 perturbed members, which assimilate "perturbed" observations that included Gaussian noise errors consistent with the observational error statistics prescribed by the data assimilation system. The number of ensemble members was determined by the computer resources available at FNMOC. In NAVDAS-AR, these random errors are added to the innovation vector after thinning and gross quality control checks but before the bias correction is applied to the radiances. The NAVDAR-AR system uses a Variational Bias Correction (VarBC), method which adds the bias coefficient predictors to the control vector for the data assimilation, solving for them at each cycle point (Auligne et al., 2007; Dee, 2005). For NCODA, the raw observations are perturbed before NCODA generated synthetic profiles based on sea surface height anomalies and before thinning and superobservations are computed. Due to the relative numbers of real and synthetic subsurface observations, the largest impact of the ocean observation perturbations is through the ISOP profiles generated by perturbing the SSHA and SST values use as inputs to the ISOP system. This generates observation spreads consistent with the measured model forecast errors (not shown). Although the perturbation of in situ profile observations has limited impact in areas regularly sampled by the altimeters, a method is being developed that adjusts the profile vertically to fit a random perturbation of the profile steric height, as is the simpler method of perturbing the profile location on the model grid after the profile selection process is complete. Either of these should improve on the present approach of perturbing profile level observations independently with no accounting for correlated error.

Unlike traditional EDA systems (El Ouaraini & Berre, 2011; Houtekamer et al., 1996), the Navy-ESPC system does not feed the ensemble variance back into the covariance model used for data assimilation. Furthermore, we do not center the ensemble system on either the control member or the high-resolution analysis from the Navy-ESPC$_{ENS}$ system. We made these choices because of the extreme computational expense of the system that made it impractical to synchronize all the ensemble members at every update cycle.

Prior research (Bowler, 2017) indicated that EDA systems tend to be underdispersive at initial time with sluggish growth of unstable baroclinic modes in the atmosphere. While fully realizing this drawback of the pure EDA system, we chose not to include either stochastic noise or dynamically condition posterior ensembles through inflation or addition of atmospheric singular vectors. This choice was motivated by a compressed development timeline for the first version of the system, and we plan to address these shortcomings in upcoming versions through methods for representing the stochastic effect of unresolved processes (Crawford et al., 2020).

4. Diagnostic Runs

For this paper, we run Navy-ESPC with two different configurations/resolutions (Table 1). For each configuration, reanalysis cycles started on 15th December 2016 at 12Z. The atmospheric Initial Conditions (ICs) for the starts of the reanalyses were taken from the operational NAVGEM for both configurations. For HYCOM and CICE, the Navy-ESPC$_{ENS}$ ICs are from operational stand-alone GOFS 3.1 and the ICs for Navy-ESPC$_{DET}$ are from a 1/25° stand-alone GOFS 3.5 (Metzger et al., 2019) that has been approved for operational testing at FNMOC.

For the ensemble configuration, the ensemble reanalysis started from the same initial condition on 15th December 2016. We allowed our EDA system to develop spread in the ensemble initial conditions by cycling the system for 1.5 months (until 1st February 2017). Preliminary results suggested that the ensemble spread in the EDA system stabilized after 1–3 days in the atmosphere and after 1–2 month in the ocean (for most variables). Some small increase in the sea surface height spread continued in the Western boundary regions up to 6 months into the reanalysis.

The reforecasts results presented in this paper for a single member of the Navy-ESPC$_{DET}$ are from the period 1 January 2017 to 31 December 2017. A series of 16-day forecasts ran every Wednesday at 12Z. For the ensemble configuration, the 60-day reforecasts were performed every Wednesday at 12Z, starting from 1st February 2017 and through 24 January 2018.
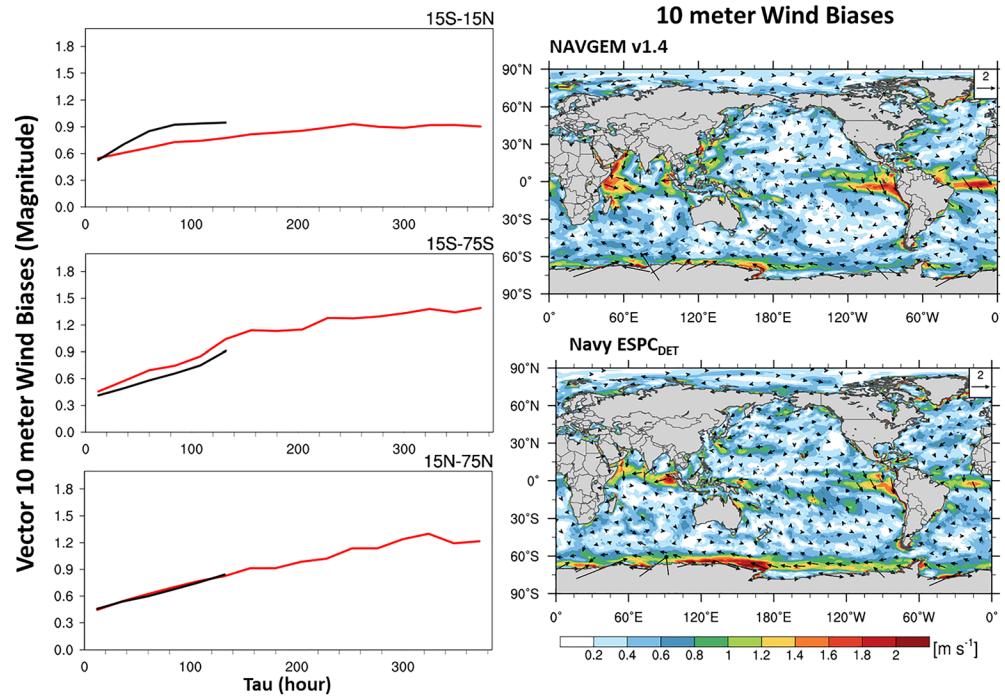
Figure 5. (left) Ten-meter vector wind-speed biases for (black) NAVGEMv1.4 from FNMOC operations and (red) Navy-ESPC$_{DET}$ from 1 January 2017 to 31 December 2017. Averages over (top left) 15°S to 15°N, (middle left) 15°S to 75°S, and (bottom left) 15°N to 75°N are shown. The vector wind-bias is on the y axis and the forecast hour is on the x axis. (right) Spatial maps of biases for (top-right) FNMOC operational NAVGEMv1.4 and (bottom-right) Navy-ESPC$_{DET}$ averaged over the first seven forecast days. In all panels, biases are calculated using ECMWF analyses as verification over the same time frame.

5. Results

5.1. Deterministic/High Resolution Configuration

5.1.1. Atmosphere Verification

We compared the atmospheric results of the deterministic configuration (Navy-ESPC$_{DET}$) to the stand-alone operational NAVGEM version 1.4. The spectral resolutions of NAVGEM in Navy-ESPC$_{DET}$ and NAVGEMv1.4 are different (T681 vs. T425, respectively), and, as noted in section 2.1, there are differences in the physics parameterizations between these models. Because of these differences, insight into how the addition of coupling affects the results is difficult to deduce independently, and these comparisons represent differences between Navy-ESPC$_{DET}$ and the current operational version of NAVGEM. A large amount of analysis was performed in comparing these systems and a very brief summary of some of these results is shown in the following section.

An example of some lower atmospheric verification is shown in Figure 5 examining the 10-m wind biases for Navy-ESPC$_{DET}$ and NAVGEMv1.4. For the 10-m wind speed, the Navy-ESPC$_{DET}$ has a lower bias in the tropical region, while the biases in the southern extratropics (15°S to 75°S) are higher than NAVGEMv1.4. The differences in the tropical wind biases may be due in part to the coupling as the SST gradients are more realistic compared to the stand-alone NAVGEM, and wind stress and SSTs are tightly coupled in this region (Chelton, 2001). The spatial maps show a slight improvement in the Navy-ESPC$_{DET}$ over the stand-alone system in the western boundary current locations, such as the Gulf Stream and Kuroshio Extension, and these slight improvements are consistent with more realistic SST gradients in Navy-ESPC$_{DET}$ compared to NAVGEMv1.4. Southern hemispheric winds near the sea ice edge have a greater bias in NAVGEM-ESPC$_{DET}$ compared to NAVGEMv1.4.

The verification of tropical storms was performed, but lacked robustness due to the limited number of storms in our validation data set. While the sample size does not support an in-depth analysis, no significant

improvement or degradation for the Navy-ESPC$_{DET}$ against the stand alone NAVGEM v1.4 was found for tropical storm tracks (not shown).

The standard FNMOC global atmosphere scorecard (Hogan, 2014) is an example of a summary which seeks to provide a more comprehensive atmospheric comparison between the Navy-ESPC$_{DET}$ and FNMOC operational NAVGEMv1.4. The scorecard was constructed by the U.S. Navy Administrative Model Oversight Panel (AMOP), used to evaluate model upgrades for targeted decision makers, and not necessarily model developers concerned about physics, though the information is useful in determining general improvements in a model. The scorecard is defined by a set of specific variables (e.g., wind, air temperature, geopotential height), levels (e.g., 500 hPa, surface, 100 hPa), error metrics (e.g., anomaly correlation, ACOR; root-mean-square error, RMSE), regions (e.g., Northern Hemisphere, tropics), verifying references (e.g., radiosonde observations, self-analysis, buoy observations), and lead times (e.g., 5-day forecast). For each scorecard entry, we compute pairwise differences between the error metrics of two sets of forecasts over the comparison period, then use a two-sided t test with a null hypothesis of no difference (i.e., mean pairwise error difference is zero). To account for serial error correlation in subsequent forecasts, we use a lag-1 autocorrelation correction to compute an effective sample (Wilks, 2011). If the null hypothesis is rejected (and, for some scorecard elements, if the mean difference exceeds a threshold value), then the specified number of points are assigned: positive points for improvement, negative points for degradation. We use a two-sided t test to evaluate if the forecast errors are distinguishable, rather than sequential one-sided tests for improvement or degradation. This strategy is consistent with historical practice and decreases the probability that we will erroneously reject the null hypothesis. A neutral change is common for many experiments such as adding an observing system, or minor changes to the model or data assimilation system. Scores ranging from 0–5 are small, 5–10 moderate, and it is rare to see a score larger than 10. The maximum achievable score is 25.

When comparing forecast lead-times of 4 and 5 days, NAVGEMv1.4 outperforms Navy-ESPC$_{DET}$ with an overall scorecard difference of 5. However, these degradations are generally small and become even smaller at longer lead times. When considering the same scorecard and extending the lead-time to 6 days, Navy-ESPC$_{DET}$ outperforms NAVGEMv1.4 with positive scores in the tropics. These results reflect the design of the Navy-ESPC physics suite to perform well at longer lead times. The overall findings are that the Navy-ESPC$_{DET}$ performances are broadly neutral compared with the operational FNMOC NAVGEMv1.4 performance, while providing some significant improvements in the tropics.

5.1.2. Ocean Verification

Navy-ESPC$_{DET}$ ocean output is compared to GOFS 3.1 and GOFS3.5. HYCOM in GOFS 3.5 is identical to the version in Navy-ESPC$_{DET}$ as they both have 1/25° horizontal resolution, 41 vertical hybrid layers and internal tides, but GOFS 3.5 is not two-way coupled to the atmosphere, i.e., it is a stand-alone ocean/sea ice prediction system which is forced by operational stand-alone NAVGEM forecast fields. HYCOM in stand-alone GOFS 3.1 has 1/12° horizontal resolution and 41 vertical hybrid layers, which is the same as the version in Navy-ESPC$_{ENS}$. Another difference between Navy-ESPC and the GOFS systems is that the GOFS systems include flux corrections to account for biases in the NAVGEM fields (Metzger et al., 2013) and biases that develop in SSTs, while Navy-ESPC does not include flux corrections.

The temperature and salinity structure in the upper 500 m of the globe (spanning 50°S–50°N) is evaluated against independent (unassimilated) profile observations of various types, e.g., Argo profiles (Roemmich & Gilson, 2009). For a given observation, model outputs are sampled at the nearest model grid point and the bias (mean error, model minus observations) and RMSE are computed as a function of forecast length (Figure 6). Verification is further broken down into various depth ranges. All systems show an overall cool bias and RMSE that grow as a function of forecast length. The highest bias and RMSE are in the depth range of the thermocline (50–150 m). The positive impact of atmospheric coupling can be seen in the shallowest depth range (0–50 m) in that Navy-ESPC$_{DET}$ shows lower bias and RMSE than the stand-alone ocean forecast systems.

Accurate knowledge of the underwater acoustical environment is also important for naval operations. The three-dimensional structure of temperature and salinity and the surface mixed-layer depth (MLD) determine the sound speed profile, which characterizes the acoustical ducts in the ocean. Thus, the forecast system must be able to accurately predict the MLD, sonic layer depth (SLD), and other acoustical proxies. The
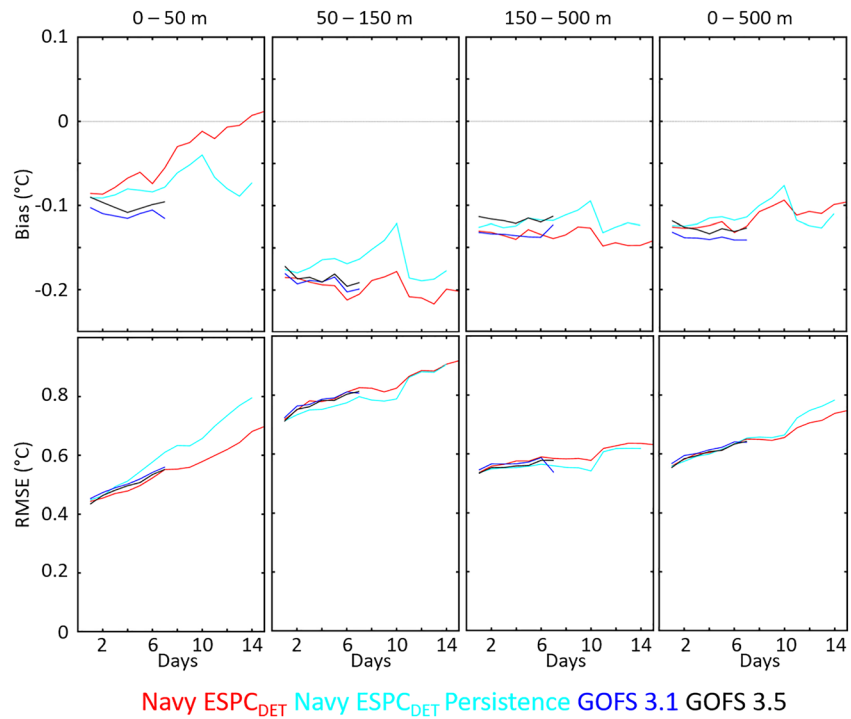
Figure 6. Temperature (°C) bias (mean error, model minus observation) (top row) and RMSE (bottom row) evaluated against unassimilated profile observations as a function of forecast length (days) for the global region spanning 50°S–50°N for calendar year 2017. The analyses are grouped by depth ranges: 0–50, 50–150, 150–500, and 0–500 m (first through fourth columns, respectively). The red curves are Navy-ESPC$_{DET}$, cyan curves are Navy-ESPC$_{DET}$ persistence, blue curves are GOFS 3.1, and black curves are GOFS 3.5.

MLD is defined by a density difference of 0.15 kg/m^3 between the surface and a given depth if both temperature and salinity profiles are available or a difference of 0.5°C between the surface and a given depth if only T profiles are available. The SLD is the distance between the surface and the depth of the sound speed maximum, which is often, but not always, at the base of the mixed layer. The MLD and SLD bias and RMSE as a function of forecast length are shown in Figure 7. All systems show a shallow bias with similar performance and all are superior to persistence forecasts of Navy-ESPC$_{DET}$ analyses, i.e., an unchanging forecast. RMSE grows with increasing forecast length but with similar trends for all systems. However, for this particular metric both stand-alone GOFS show slightly better performance to Navy-ESPC$_{DET}$. The reasons for this are not yet fully understood.

5.1.3. Sea Ice Verification

For sea ice validation, we compared the Navy-ESPC$_{DET}$ ice output to GOFS 3.1 and GOFS 3.5. A major difference between GOFS3.5 and the other systems is that GOFS 3.5 uses CICEv5.1.2 while GOFS 3.1 and Navy-ESPC use CICEv4. This newer CICE version includes the correction of known software bugs, two new melt pond parameterizations, an improved parameterization for form drag, use of sea ice velocity in the coupling updates for high frequency coupling, and Elastic-Anisotrophic-Plastic rheology to name a few of the changes. The reason for the difference in CICE versions was one of timing, i.e., the ESPC code was locked down for testing before the GOFS 3.5 code.

Figure 8 shows the sea ice edge error compared to the independent analysis of the National Ice Center (NIC) (Partington et al., 2003) as a function of forecast day for the Pan-Arctic and multiple Northern Hemisphere subregions shown in the inset. GOFS3.5 with CICEv5.1.2 generally outperforms both GOFS 3.1 and Navy-ESPC$_{DET}$. However, at forecast lead-times of 1 and 2 days, Navy-ESPC$_{DET}$ has similar skill to GOFS 3.5 for the GIN Sea, Barents/Kara Seas, and Beaufort Sea subregions. Note for the Pan-Arctic and some subregions the error is higher at the nowcast time and decreases as forecast length increases (out to about day 2) before eventually increasing again. We hypothesize this is due to the order in which the Interactive
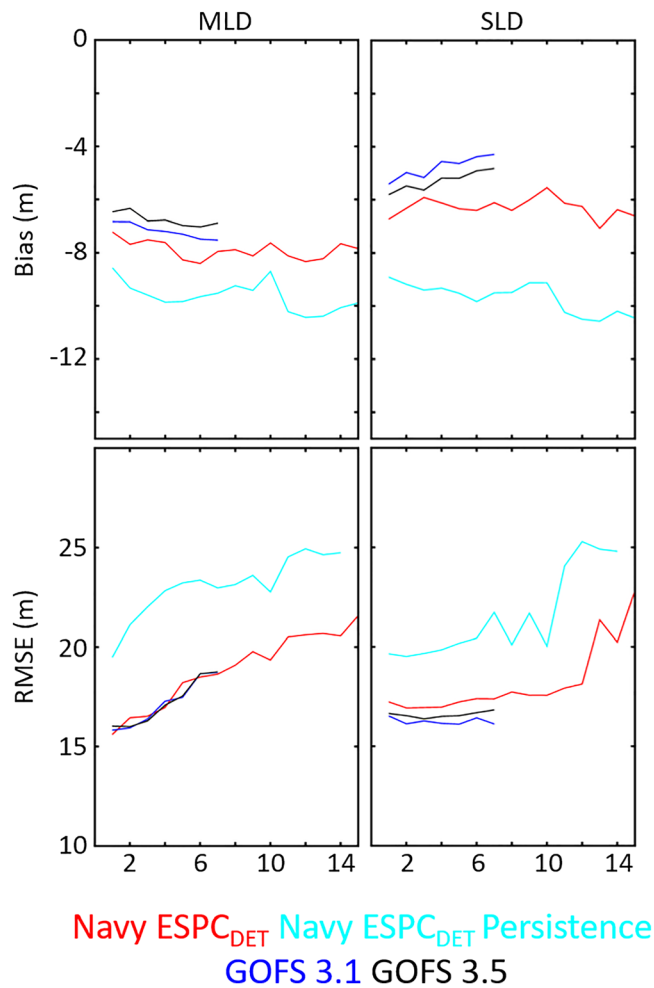
Figure 7. Mixed layer depth (m, left column) and sonic layer depth (m, right column) bias (top row) and RMSE (bottom row) against unassimilated profile observations vs. forecast length (days) for Navy-ESPC$_{DET}$ (red), Navy-ESPC$_{DET}$ persistence (cyan), GOFS 3.1 (blue), and GOFS 3.5 (black). Note that the results start at 24 hr after the forecast. Hence, differences between the start of the lines.

Multisensor Snow and Ice Mapping System (IMS) (Helfrich et al., 2007) sea ice mask is applied, i.e., after the NCODA sea ice analysis completes. IMS typically does not extend as far equatorward as the NIC sea ice edge and so the nowcast time may have higher error. Subsequent testing has shown that bringing the IMS mask inside of NCODA and only applying it to passive microwave sensors that are known to suffer from nondifferentiation of sea ice and summertime melt ponds improves ice edge error performance at the nowcast time.
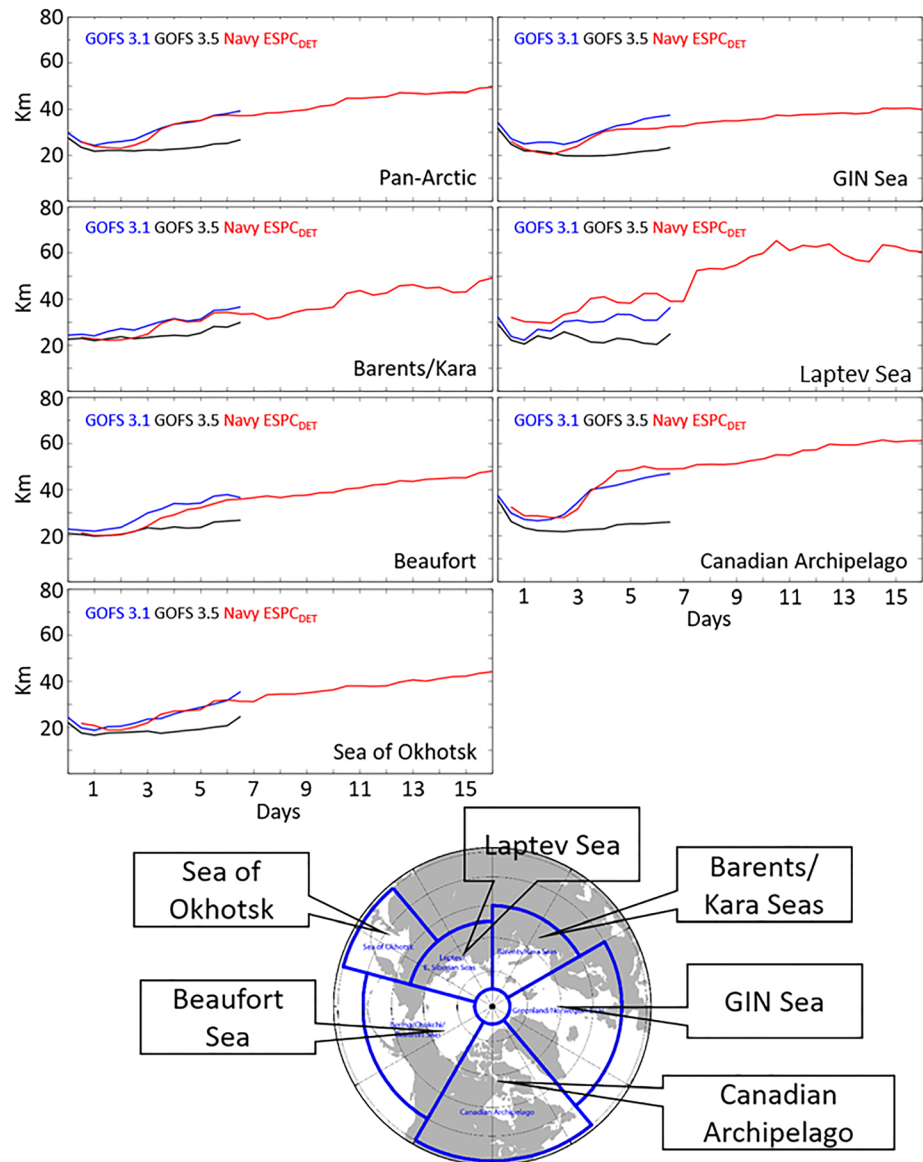
Sea ice drift from the three systems is compared against ice-bound drifting buoys from the International Arctic Buoy Program (http://iabp.apl.washington.edu/index.html). A total of 207 drifting buoys exist over 2017 with the majority of the drifters located north of Alaska and the Canadian Archipelago. Figure 9 shows mean absolute error and RMSE of drifter speed along with direction vector correlation at the 12-hr forecast time. Navy-ESPC$_{DET}$ has the lowest speed bias and often the lowest RMSE, but the vector correlation of direction is typically worse than both GOFS 3.1 and 3.5.

5.2. Ensemble Resolution Configuration
5.2.1. Atmosphere Verification
As noted above, one goal of the NAVGEM physics updates was to obtain skillful long forecasts of the MJO; therefore, we compare Navy-ESPC$_{ENS}$ MJO predictions to state-of-the art systems from other NWP centers. In order to evaluate the skill of the MJO forecasts, we use the Real-time Multivariate MJO (RMM) index (Wheeler & Hendon, 2004). This commonly used index has two components. RMM1 is positive when the MJO is active over the Maritime Continent and negative when the MJO is active over South America to Africa. RMM2 is positive when the MJO is active over the Pacific Ocean and negative when the MJO is active over the Indian Ocean.

We compare Navy-ESPC$_{ENS}$ to the FNMOC operational atmosphere only NAVGEM ensemble transform (NAVGEM-ET) (McLay et al., 2010; Reynolds et al., 2011), Australian Bureau of Meteorology (BOM), Meteo-France (CNRM), the European Centre for Medium-range Weather Forecasts (EC), and the NOAA Climate Forecast System version 2 (CFSv2) (Table 4). We downloaded the forecasts from the other centers from the Subseasonal to Seasonal (S2S) database (Vitart et al., 2017). The models from the other centers vary in ensemble size as noted in Table 4. In addition, we show results for the Navy-ESPC$_{DET}$ configuration. Verification was derived from NOAA satellite-derived Outgoing Longwave Radiation, and the ECMWF ERA-Interim reanalyses.

Figure 10 shows violin diagrams indicating the distributions of the day when the ACOR of the individual ensemble member falls below 0.6 for RMM1 and RMM2. The Navy-ESPC$_{ENS}$ ensemble individual members perform exceptionally well in terms of RMM1 ACOR compared to the other centers. In terms of the ensemble mean ("X"s in Figure 10), Navy-ESPC$_{ENS}$ performance is slightly below ECMWF, comparable to CNRM, and better than the other centers and NAVGEM ensemble. Navy-ESPC$_{DET}$ also performs relatively well.

For RMM2, the performance of individual ensemble members of Navy-ESPC$_{ENS}$ is not quite as good as ECMWF, but on average better than the other centers. For the ensemble mean ("X"s), Navy-ESPC$_{ENS}$ outperforms CFSv2 and BOM. For several of the forecast systems, the average RMM2 ACOR does not fall below 0.6 before the end of their forecast range, hence the lack of "X"s for these systems in the bottom panel of Figure 10.

The MJO forecasts have ACOR skill above 0.6 for about 20 days for both indices, which is comparable to other centers. It should be noted that the sample size here is quite small given that the MJO is not always

Figure 8. Northern Hemisphere sea ice edge error (km) vs. the independent NIC ice edge as a function of forecast length for (red) Navy-ESPC$_{DET}$, (blue) GOFS 3.1, and (black) GOFS 3.5. The geographic locations of the regional analysis are shown in the map at the bottom. Note that Navy-ESPC$_{DET}$ and GOFS 3.1 use CICEv4.0 whereas GOFS 3.5 uses CICE v5.1.2.

active and the time period considered is only a year. Thus, these results should be used to give a general idea of performance only. When examining 17 years of Navy ESCP forecasts produced for the SubX experiment, Janiga et al. (2018) and Kim et al. (2019) found that the Navy-ESPC$_{ENS}$ system overestimates the amplitude of the MJO whereas most other models underestimate the amplitude of the MJO. The ACOR of the RMM1 and RMM2 is more indicative of predictions of the phase of the MJO where other metrics such as the RMSE of RMM1 and RMM2 also reflect errors in MJO amplitude.

In addition to the MJO, we anticipated that teleconnection patterns might be predictable on longer time scales compared to fields at one location because of their large scale. Teleconnection patterns evolve on relatively slow time scales and as a result are more predictable than the atmospheric state at any single point. The teleconnection patterns considered include the Arctic Oscillation (AO) and Antarctic Oscillation (AAO), which provide information on the strength and waviness of the northern and southern jet streams, and the Pacific North American Oscillation (PNA) and North Atlantic Oscillation (NAO), which provide
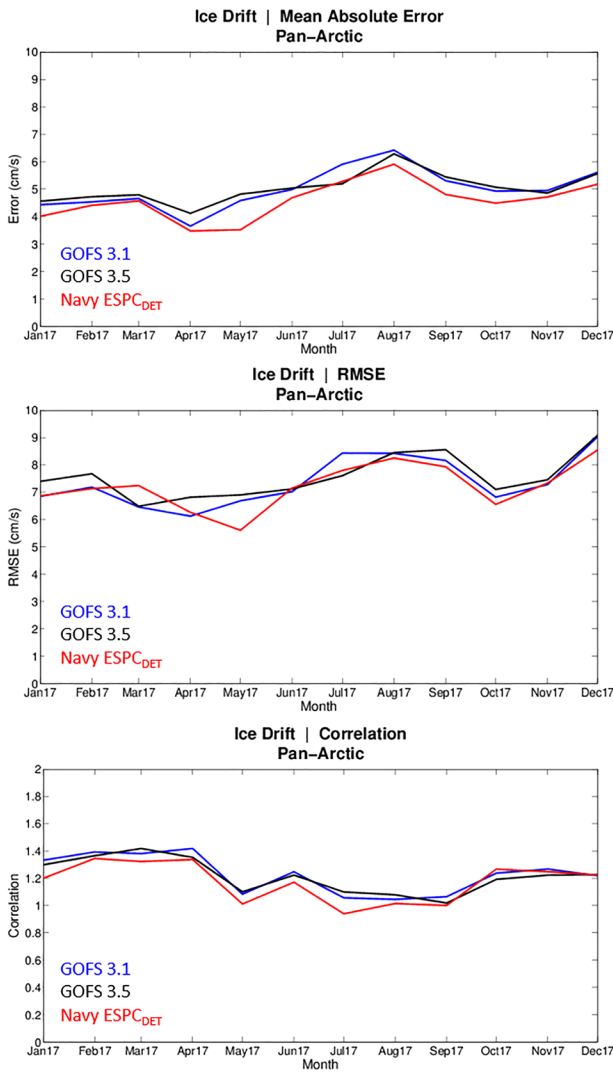
Figure 9. Monthly (top) speed mean absolute error (cm/s), (middle) RMSE (cm/s), and (bottom) direction vector correlation of unassimilated IABP ice-bound drifting buoys for the Pan Arctic region compared against GOFS 3.1 (blue), GOFS 3.5 (black), and Navy-ESPC$_{DET}$ (red) for tau 012 at all ice concentrations. The vector correlation is based on Crosby et al. (1993) with 0 indicating no correlation and 2 indicating perfect correlation. This analysis spans all days of 2017.

information on dominant modes of variability over these respective regions. The models that we compare these indices to are the same as the MJO and listed in Table 4. For MJO, however, the verifying scores were computed using NOAA interpolated OLR (Liebmann & Smith, 1996) and ERA-Interim (Dee, 2011) zonal and meridional wind fields, while the other teleconnection patterns' verifying scores were computed using the National Centers for Environmental Prediction (NCEP) reanalysis height fields.

Figure 11 shows Violin plots describing the distribution of the forecast day when the ACOR skill score of these ensemble forecasts (considered individually, i.e., in a deterministic sense) drops below 0.6. The individual Navy-ESPC$_{ENS}$ forecasts are comparable or better than the performance of the other systems. The day the Navy-ESPC$_{ENS}$ ensemble mean forecast skill falls below 0.6 is also included for these systems, and is indicated by the "X". Using this measure, the Navy-ESPC$_{ENS}$ ensemble mean remains skillful usually about a half-day or day longer than the median of the individual forecasts, while other centers are usually able to gain 2, sometimes 3, days in skill using the ensemble mean. NAVGEM-ET gain in skill between deterministic and ensemble for the AO and PNA is particularly large, attesting to the merits in the ensemble design. Thus, the ensemble mean skill between NAVGEM-ET and Navy-ESPC$_{ENS}$ is comparable as NAVGEM-ET has a slight edge for the AO, and Navy-ESPC$_{ENS}$ has an edge for the AAO. The skill of the Navy-ESPC$_{DET}$ is comparable to Navy-ESPC$_{ENS}$, and generally within a day difference of the ensemble mean score.

Using the rule-of-thumb threshold of 0.6 for useful ACORs, the predictive skill for these patterns typically dips below that threshold between 8 and 10 days. We also wish to note a caveat when interpreting the average skill of the individual members. The relatively good performance of the individual members of Navy-ESPC$_{ENS}$ is in part due to a relative lack of spread in the ensemble. Figure 12 displays the ratio of average variance to average squared error of the ensemble mean, referred to as the spread-skill ratio, for the Navy-ESPC$_{ENS}$ and NAVGEM ET for selected variables and regions. Here the errors are calculated using ECMWF ERA Interim reanalyses. Neglecting analysis error variance, an ideal ensemble would exhibit a spread-skill ratio of one. As shown, the Navy-ESPC$_{ENS}$ is less dispersive compared to the NAVGEM ET at lead times before ~15 days in the forecast. This lack of ensemble spread means that the best members in the Navy-ESPC$_{ENS}$ tend not to be as skillful as the best members at other centers, and likewise the worst members in the Navy-ESPC$_{ENS}$ ensemble tend to have smaller errors than the worst members of the other centers. As noted, there is currently a dedicated effort to improve the spread of the Navy-ESPC ensemble.

5.2.2. Ocean Verification

When diagnosing the 60-day forecasts of the ocean in Navy-ESPC$_{ENS}$, the primary reference comparison is climatology. Here we use the Generalized Digital Environmental Model (GDEM) version 4 (Carnes et al., 2010) that is the U.S. Navy's gridded monthly full-depth climatology of temperature and salinity and their standard deviations on a 0.25° global grid, and was built from 8.3 million profiles spanning multiple decades.

Ensemble spread of Sea Surface Height is first examined to understand the spatial variability created by the ensemble method (Figure 13). Larger spread concentrated in the vicinity of the major current systems associated with the Gulf Stream, Kuroshio (off Japan), and, in the southern hemisphere, the Antarctic Circumpolar Current (ACC), Agulhas Current (off southeast Africa), Eastern Australian Current. The larger

Table 4
Description of Modeling Systems Used in the MJO and Teleconnection Intercomparisons

Model acronym	Modeling Center	# of ensemble members	Frequency of forecasts
BOM	Australian Bureau of Meteorology	33	3.5 days from 2017-01-04
CFSv2	NOAA Climate Forecast System Version 2	16	Daily from 2016-12-31, but match Navy-ESPC$_{ENS}$
CNRM	Meteo-France	51	Weekly from 2017-02-01
EC16	European Centre for Medium Range Weather Forecasting (ECMWF)	16	3.5 days from 2017-01-02
EC51	European Centre for Medium Range Weather Forecasting (ECMWF)	51	3.5 days from 2017-01-02
NAVGEM-D	NAVGEM Operational Deterministic System	1	Weekly from 2017-02-01
NAVGEM-ET	Navy NAVGEM Operational Ensemble System	20	Weekly from 2017-02-01
Navy-ESPC$_{DET}$	Deterministic Navy-ESPC	1	Weekly from 2017-01-04
Navy-ESPC$_{ENS}$	Ensemble Navy-ESPC	16	Weekly from 2017-02-01

*Note.* Dates are in the format year-month-day.

spread indicates growing dynamical instabilities associated with the mesoscale features. The ensemble spread grows with forecast length as evidenced from the spread at 240-hr forecast.

Ocean current validation from Navy-ESPC$_{ENS}$ was performed using independent drifting buoy observations with a drogue at 15 m depth from the NOAA Global Drifter Program (https://www.aoml.noaa.gov/phod/gdp/index.php) vs. HYCOM output at this same depth level. Care was taken to use only those buoys that maintained their drogue. Metrics included speed bias and RMSE along with directional vector correlation, and these latter two are shown in Figure 14 for the nowcast time and as a function of forecast length for stand-alone GOFS 3.1, the control member of Navy-ESPC$_{ENS}$, and the Navy-ESPC$_{ENS}$ ensemble mean. The control member of Navy-ESPC$_{ENS}$ slightly outperforms GOFS 3.1 (lower RMSE and higher vector correlation), indicating the atmosphere-ocean coupling is leading to improved near surface currents, but a clear benefit is seen when using the ensemble mean. The forecast error grows modestly out through the day 6 forecast.

We show the extended-range ensemble-mean temperature bias and RMSE of the top 500 m of the water column in Figure 15. Navy-ESPC$_{ENS}$ mean forecast has skill relative to the climatology, which is indicated by a lower RMSE than the GDEM4 climatology, out to about 35 days in the forecast, as indicated by the crossing
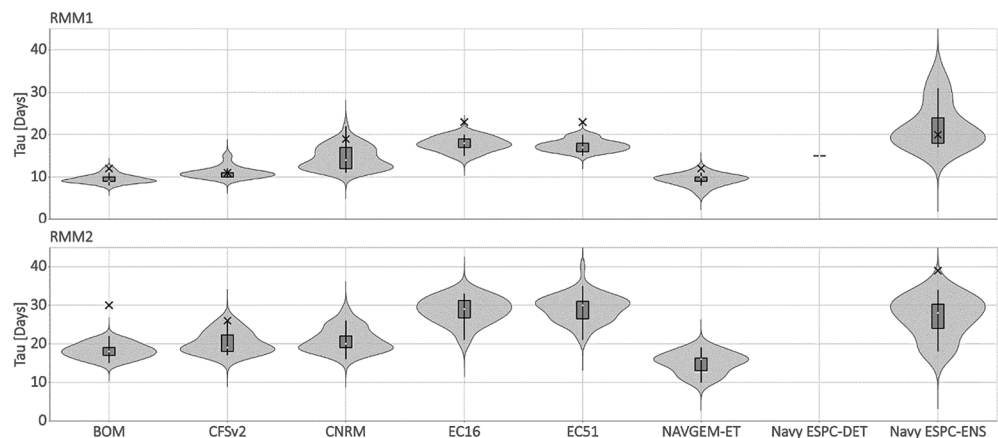

Figure 10. Violin plots showing the distribution of the day that the forecast of RMM1 (top panel) and RMM2 (bottom panel) anomaly correlation (ACOR) falls below 0.6. Each ensemble member is treated as an individual (deterministic) forecast, and the width of the violin is proportional to the number of ensemble members that fall below 0.6 on that day. The probability density is represented by the width of the plot and is proportional to the number of members that have the value. The box-and-whisker plot inside the violin plot shows the interquartile range with the whiskers out to the 95% confidence intervals. The ensemble mean is represented by the "X". Note for RMM2 that the Navy-ESPC$_{DET}$ forecast is only 16 days and the ACOR did not fall below 0.6 before the end of the forecast.
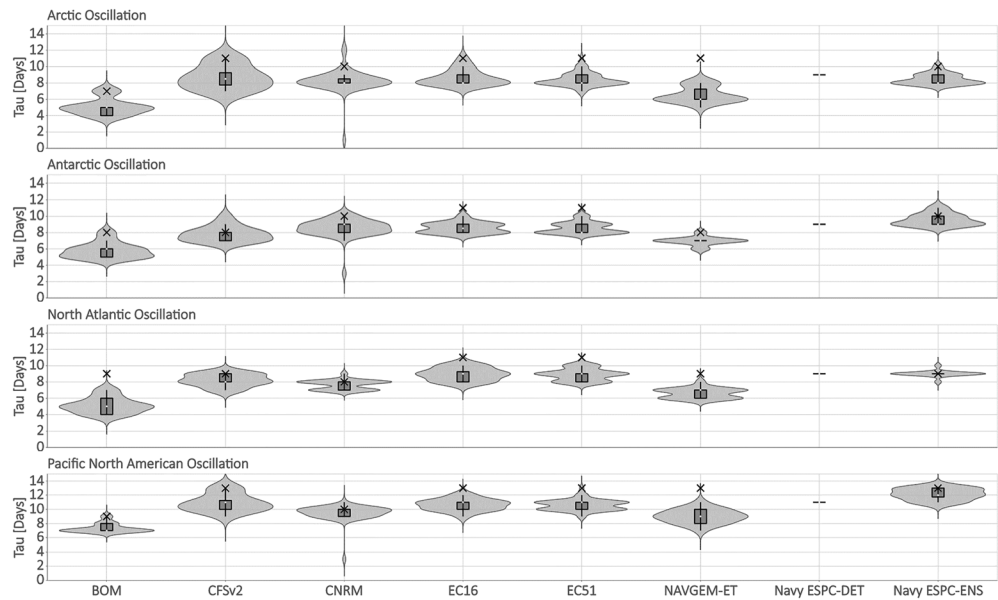
Figure 11. Violin plots showing the distribution of the day that the forecast of the large-scale oscillation (as denoted in panels) ACOR falls below 0.6. Each ensemble member is treated as an individual (deterministic) forecast, and the width of the violin is proportional to the number of ensemble members that fall below 0.6 on that day. The probability density is represented by the width of the plot and is proportional to the number of members that have the value. The box-and-whisker plot inside the violin plot shows the interquartile range with the whiskers out to the 95% confidence intervals. The day on which the ensemble mean forecast ACOR falls below 0.6 is denoted by the "X".

of the RMSE lines in the left panel. The RMSE of the control member (without perturbed observations) crosses climatology at a lead-time of 10 days, which shows the benefit of an ensemble for this metric. The right panel shows the RMSE, bias, and absolute error for the Navy-ESPC$_{ENS}$ ensemble mean and the
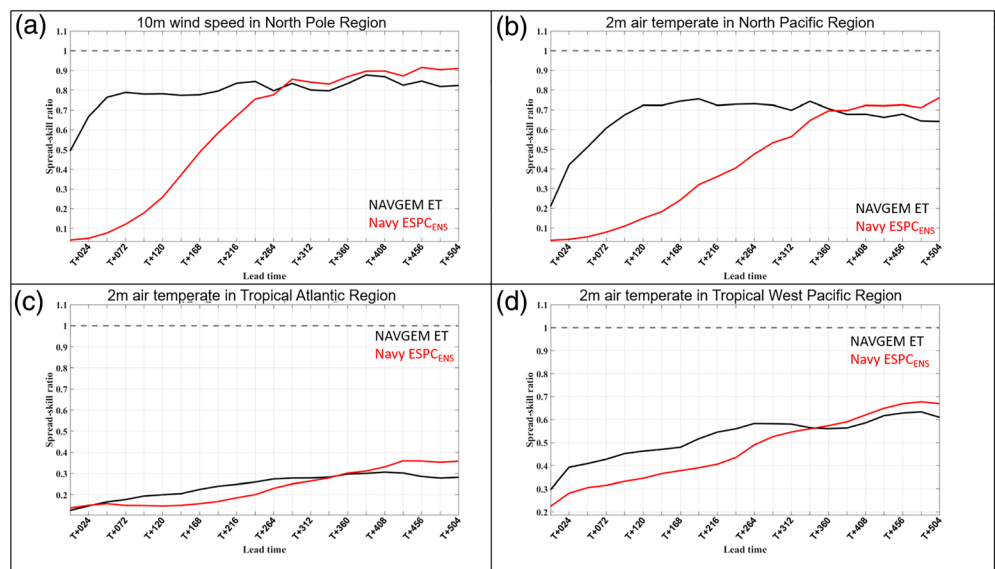

Figure 12. Average ratio of ensemble variance and squared-error of the ensemble mean (Spread Skill Ratio) for the (black) NAVGEM ET and the (red) Navy-ESPC$_{ENS}$. Error is calculated from ERA-interim analysis. The x axis is the forecast length of the forecast. Four variables for specific regions are shown: (a) 10-m winds in the North Pole region (65–90°N, 0–360°E), (b) 2-m air temperatures in the North Pacific Region (30–60°N, 130–240°E), (c) 2-m air temperature in the Tropical Atlantic Region (20°S–20°N, 290–360°E), and (d) 2-m air temperature in the Tropical West Pacific Region (20°S–20°N, 110–200°E). The dashed horizontal line at 1 represents a perfect spread-skill ratio. Note the forecast length is only to 504 hr because this is the forecast length for the NAVGEM ET ensemble.
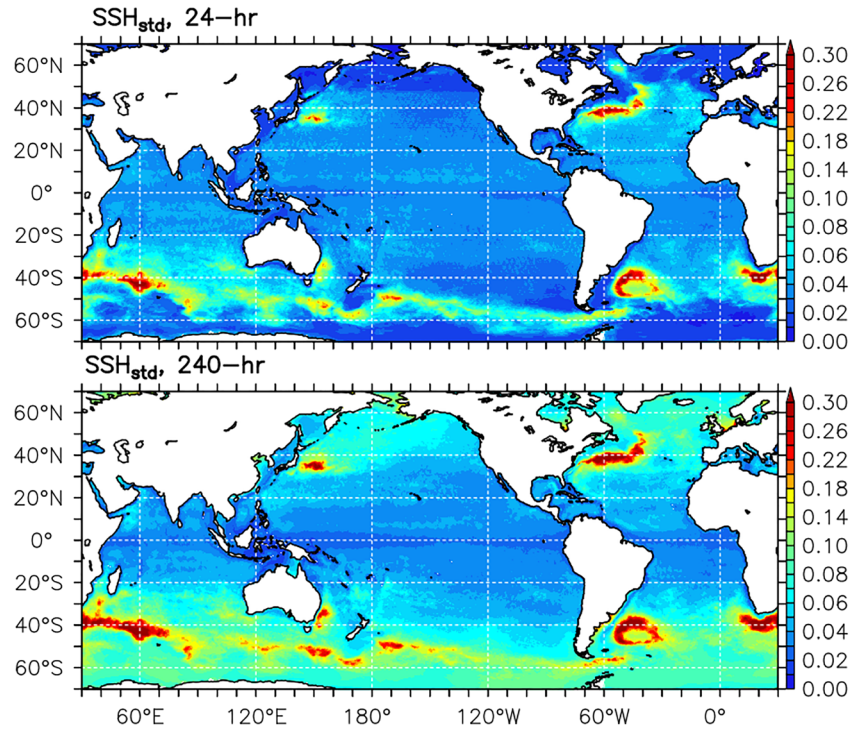
Figure 13. The standard deviation of sea surface height (SSH) in meters for the 16 members of Navy-ESPC$_{ENS}$ at forecast hours of (top) 24 and (bottom) 240 hr. Forecasts are described in section 4.

GDEM4 climatology for the upper ocean to 500 m. Navy-ESPC$_{ENS}$ ensemble mean has a small negative bias at the initial time that decreases overall and becomes positive near the surface, which is consistent with the longwave radiation bias (not shown), although the total heat flux drives the surface temperature. The GDEM4 climatology has a consistent and somewhat larger negative bias. The RMSE of the Navy-ESPC$_{ENS}$ mean at the start of the forecast is noticeably lower than the GDEM4 reference, grows quickly in the upper thermocline, and at 100 m becomes larger than climatology (>1.3°C) between forecast days 26 and 40. Vertical profiles of temperature RMSE and bias averaged over the 60-day forecast show the overall better performance of the ensemble mean compared to GDEM4. The time average spread to RMSE ratio
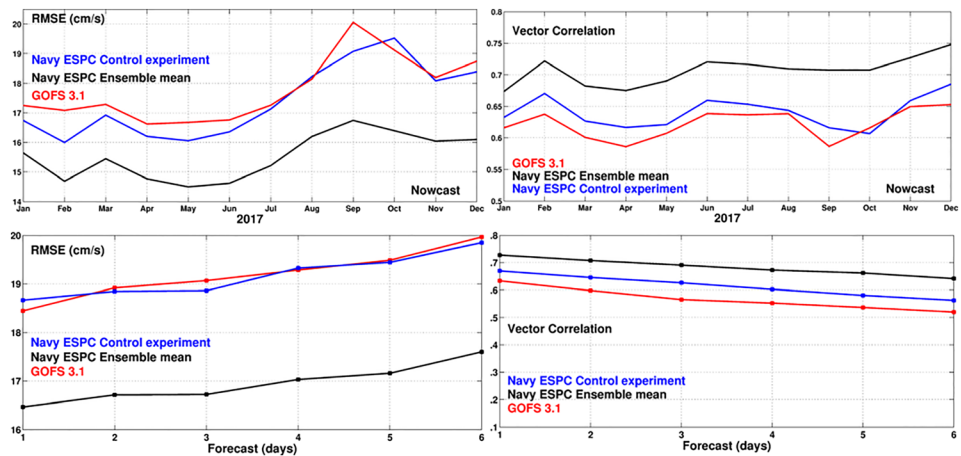

Figure 14. Monthly upper ocean (15 m) speed root-mean-square error (RMSE, cm/s) (upper left) and vector correlation (upper right) at the nowcast time and speed RMSE (lower left) and vector correlation (lower right) as a function of forecast length for GOFS 3.1 (red), the control member of Navy-ESPC$_{ENS}$ (blue), and the Navy-ESPC$_{ENS}$ ensemble mean (black) against independent drifting buoy observations. The statistics span 2017.
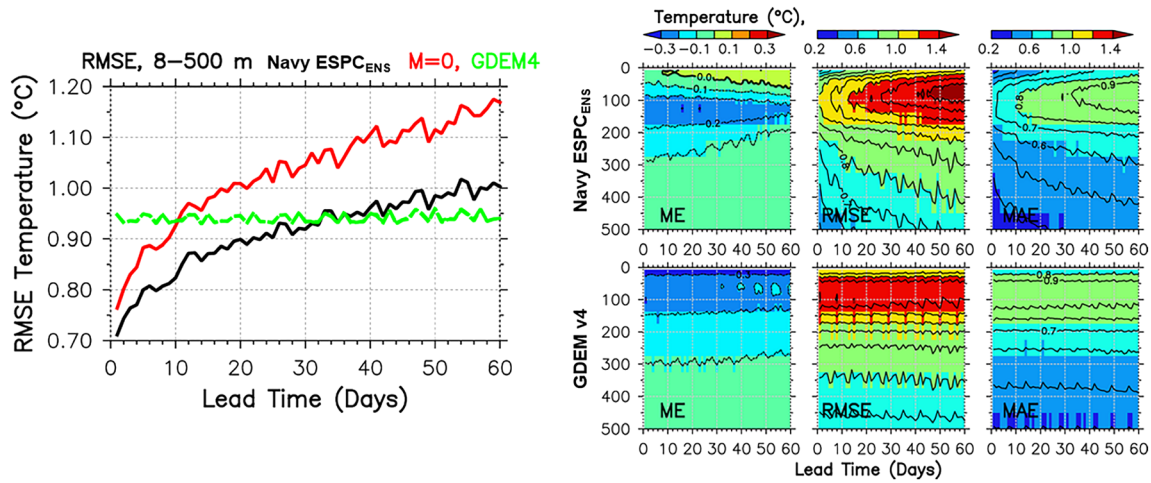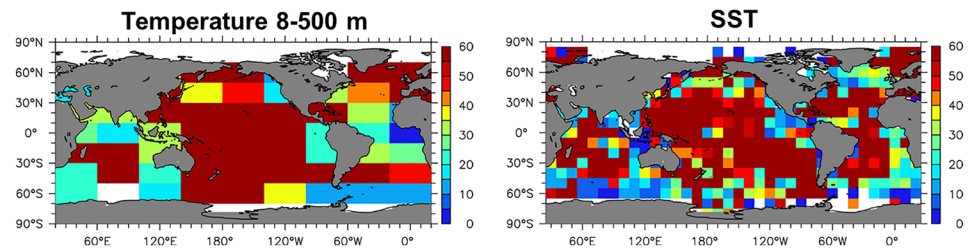
Figure 15. (left) Mean RMSE of ocean temperature over 8–500 m depth as a function of forecast lead time relative to unassimilated profiles for the ensemble (black line), control member (red line), and GDEM4 climatology (green line), and (RIGHT) temperature bias (ME), RMSE, and mean absolute error (MAE) for the upper 500 m for 1- to 60-day forecast time for the ensemble mean (top row) and the GDEM4 monthly climatology (bottom row).

for temperature in the upper 500 m is 0.60 and increases to 0.67 after subtracting bias from RMSE (RMSE − |ME|), suggesting the Navy-ESPC$_{ENS}$ is accounting for 67% of uncertainty.

In situ surface temperature observations, mostly from drifting buoys and ships of opportunity, are used to evaluate the ensemble forecasts along with the GDEM4 climatology (Figure 16). We are aware of larger errors in ship SST observations compared to errors in the drifter observations. However, given the majority of the observations used in our validation come from drifters (~76%), the conclusions drawn are not likely impacted. Note that number of surface observations is significantly larger (~30,000 profiles/day) than pro-files observations (~2,500/day). The RMSE of the ensemble mean (black line), control member (unperturbed member) (red line), and GDEM4 climatology are shown in the left panel of Figure 16 and the right panel shows the bias and the standard deviation. For comparison, the errors from 15 ensemble members are shown as gray symbols. Navy-ESPC$_{ENS}$ has lower bias and RMSE than climatology at all lead times out to 60 days. Navy-ESPC$_{ENS}$ bias is near zero at lead times up to 40-days, which is significantly smaller than GDEM4 at 0.4°C. The control member has skill relative to the climatology out to 20 days in the forecast. The ensemble standard deviation (spread) growth is consistent with the RMSE. However, Navy-ESPC$_{ENS}$ is underdispersive as indicated by the standard deviation lower than RMSE, suggesting that dispersion of the ensembles does not account for the uncertainty defined by climatology. The time-average spread to
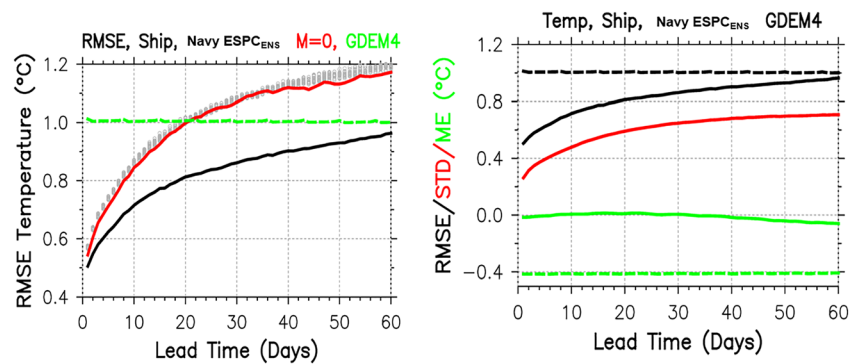
Figure 16. (left) RMSE of surface temperature relative to ship and drifting buoy observations for the ensemble mean (back line), control member (red line), and GDEM4 climatology (green line). Gray symbols are RMSE of the 15 ensemble members treated independently. (right) The bias (ME) and standard deviation (STD) as a function of forecast lead time. Dashed lines indicate GDEM4 climatological ME (green) and RMSE (black).

Figure 17. Spatial distribution of the forecast day when the ensemble mean RMSE crosses the climatological RMSE for ocean temperature in the top 500 m (left) and surface temperature (right) for the Navy-ESPC$_{ENS}$ forecasts.

RMSE ratio of ~0.72 (ranges between 0.52 and 0.76) suggests that the ensemble is accounting for 72% of uncertainty.

To evaluate the spatial distribution of ensemble temperature forecast skill, the observations and model are binned into boxes spanning 40° of longitude and 20° of latitude for temperature observations between 8 and 500 m and 10° by 10° for SSTs. For each bin, the RMSE is calculated with respect to the ensemble mean (EM), control member (M = 0), and climatology. Next, the RMSE in each bin is averaged over 8–500 m depth or SST. Last, the forecast day on which the ensmble mean RMSE crosses the climatology is identified for each bin, and plotted in Figure 17. We excluded bins with fewer than 50 matchup comparisons for the temperatures from 8 to 500 m (Figure 17, left) and fewer than 100 comparisons for the SSTs with surface data (Figure 17, right). We smooth the time-series in each bin using a 5-day boxcar method before finding the day of crossing, and in the event of multiple crossings, the earliest occurrence is selected. The ensemble mean performs better than climatology out to 60 days in the Pacific Ocean except in the Kuroshio Extension east of Japan for both near surface (8 to 500 m) temperatures and SSTs. The western boundary current regions have larger ensemble spread compared to other regions (Figure 13), which may be limiting the predictability. The ensemble mean seems to have a slightly lower predictability to 25–35 days relative to climatology in the Atlantic, Northern Indian Ocean, and Indonesian Seas. The control member has lower predictability than the ensemble mean almost everywhere in the globe (not shown).

For the temperature observations from 8 to 500 m, the relatively lower predictability in the western Arabian Sea, Bay of Bengal, and Indonesia regions can be attributed to the monsoon winds. Part of the lower predictability in the Indian Ocean regions likely stems from the dominance of the seasonally reversing monsoon. The gap winds along the west coast of Mexico appear to influence the SST predictability, lowering predictability out to about 15 days. Lower SST predictability is also evident in the Agulhas current region (20–25 days), where the generation of mesoscale eddies dominates the SST variability. When assimilating satellite altimeter data, HYCOM accurately represents Agulhas eddy shedding; however, in forecast mode, it has a tendency to produce and shed more eddies into the South Atlantic than are observed. This is a common unrealistic phenomenon in most eddy-resolving ocean models (Thoppil et al., 2011). The control member has lower predictability than the ensemble mean almost everywhere in the ocean (not shown).

5.2.3. Sea Ice Verification

Extended-range forecasts of sea ice extent are evaluated using two metrics. The first is the Integrated Ice Edge Error (IIEE) (Goessling et al., 2016), which is the total of the area over which the model has overforecast the extent of the ice edge plus the area over which the model has underforecast the extent of the ice edge. Here we define the ice edge as 15% ice concentration. We also consider a Brier Skill Score (BSS) (Brier, 1950) for ice concentration below 15% following Wayand et al. (2019). The BSS is the degree of improvement of the Brier Score (BS) of the Navy-ESPC$_{ENS}$ forecast over the BS produced using a persisted analysis or climatology. The BSS of one is a perfect forecast, while a BSS of 0 indicates there is no improvement as compared to persistence or climatology.

For verification, we follow Hebert (2015) and use self-analysis as truth. In addition, the climatological forecasts are produced using SSMR SSM/I data from the National Snow and Ice Data Center (NSIDC) (NSIDC 0051) from the 2007–2017 time-period. Including analyses from years prior to 2007 may bias the climatology on the high side given the recent observed trends in sea ice. We calculate these scores for both the Arctic and Antarctic, and consider the scores for the entire 2017 period, and for each individual season.
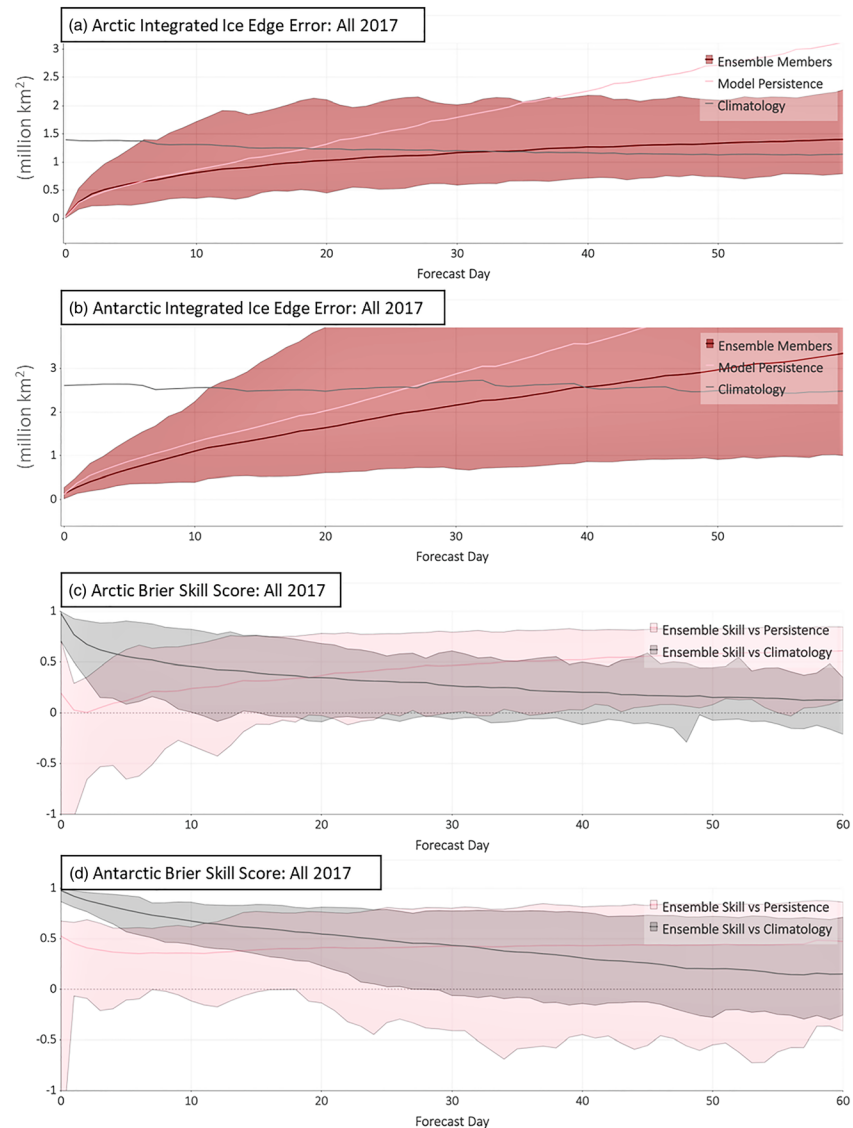
Figure 18. (a, b) Integrated Ice Edge Error (10^6 km^2) for the Arctic and Antarctic, respectively, for Navy-ESPC ensemble mean (maroon), persisted analyses (pink), and 2007–2017 climatology (gray). Range of skill of individual ensemble members indicated by shaded region. (c, d) Brier Skill Score for ice concentration less than 15% for the Arctic and Antarctic, respectively, for Navy-ESPC ensemble as compared to persistence (pink curve) and 2007–2017 climatology (gray curve). BSS range for individual ensemble members indicated by pink and gray areas.

Figure 18 shows the IIEE and BSS for both the Arctic and Antarctic as a function of forecast day averaged for the entire 2017 test period. For IIEE, Navy-EPSC$_{ENS}$ outperforms climatology out to 32 days in the Arctic and out to 40 days in the Antarctic. It outperforms a persistence forecast in the Antarctic for the entire time. In the Arctic, Navy-EPSC$_{ENS}$ outperforms persistence past 5 days, and is comparable to (slightly worse than) persistence during the first 5 days. In terms of BSS, the Navy-EPSC$_{ENS}$ outperforms both persistence and climatology in both the Arctic and Antarctic for the entire 60-day forecast.

However, the skill depends on season. For the Arctic, Navy-EPSC$_{ENS}$ retains skill above climatology for the entire 60-day forecast period for the BS (Table 5). For the Antarctic, Navy-EPSC$_{ENS}$ becomes less skillful than climatology after 46 days for both September-October-November and December-January-February for the BS. There is considerable seasonal variation in the IIEE metric, ranging from a low of 11 days during June-July-August to a high of 46 days for March-April-May in the Arctic. In the Antarctic, the Navy-ESPC

Table 5
Forecast Day for Navy-ESPC$_{ENS}$ Sea Ice Metrics per Season

Statistics for hemisphere	MAM	JJA	SON	DJF	All
Arctic					
BSS < 0	>60	>60	>60	>60	>60
IIEE > Climatology	46	11	35	37	32
Antarctic					
BSS < 0	>60	>60	46	46	>60
IIEE > Climatology	>60	28	28	39	39

Note. Brier Skill Score (BSS) < 0 represents the day when the ratio the Navy-ESPC$_{ENS}$ Brier Score over the climatology Brier Score is less than zero. The Brier Score is calculated with respect to 15% sea ice concentration. Integrated Ice Edge Error (IIEE) > Climatology represents when the IIEE of the Navy-ESPC$_{ENS}$ becomes larger than the IIEE from climatology. The top row represents the season in which the forecast started: MAM = March-April-May; JJA = June-July-August; SON = September-October-November; DJF = December-January-February; and all months.

becomes less skillful than climatology at 28 days for June-July-August and September-October-November, but retains skill above climatology for the entire 60-day forecast in March-April-May.

The IIEE metric can be broken down into overpredictions and underpredictions and the largest signal in the results is the tendency of the Navy-ESPC$_{ENS}$ to overpredict sea-ice extent in the Antarctic for all months except March-May-April, and during the other seasons, the overprediction of sea ice is 3 to 5 times larger than the underprediction (not shown). In the Arctic, the overprediction and underprediction areas are generally more balanced, although June-July-August have underpredictions that are about 2.5 times as large as overprediction area, which is contributing to the relative high IIEEs during these months.

6. Conclusions and Future Directions

Navy-ESPC is the first global coupled atmosphere-ocean-sea ice prediction system developed at NRL. It is a new tool for predicting operational and strategic environmental information for the Department of the Navy. The component models include the atmosphere—NAVGEM; ocean—HYCOM; and sea ice—CICE. Since a motivation for developing Navy-ESPC is subseasonal forecasting, we introduced physics updates into NAVGEM to improve the subseasonal simulation of equatorial phenomena, particularly the MJO, by updating NAVGEM's deep convection physics from the SAS scheme to a modified Kain-Fritsch scheme. In addition, we implemented the COARE surface flux algorithm into NAVGEM to provide more consistent air-sea fluxes with HYCOM, as HYCOM uses the same COARE algorithm. No major physics updates were included in HYCOM or CICE component models to accommodate the longer forecast time horizon. Data assimilation is weakly coupled with NAVDAS-AR for the atmosphere and NCODA for the ocean/ice components.

In this paper, we described the deterministic (Navy-ESPC$_{DET}$) and ensemble (Navy-ESPC$_{ENS}$) configurations. For Navy-ESPC$_{DET}$, we analyzed a database of 16-day reforecasts spanning January 2017 to December 2017. For Navy-ESPC$_{ENS}$, we analyzed a 16-member 60-day ensemble of reforecasts spanning February 2017 to January 2018 starting every Wednesday at 12Z.

For the Navy-ESPC$_{DET}$ configuration, the main findings are as follows:

- For the FNMOC global forecast scorecard, operational NAVGEM performs slightly better than Navy-ESPC$_{DET}$ for 4- to 5-day forecasts. However, the Navy-ESPC$_{DET}$ outperforms operational NAVGEM at 6-day forecasts. This reflects the design priorities for subseasonal forecasting in Navy-ESPC compared to the short-term (i.e., 5 days) forecasting for NAVGEM.
- For the ocean, Navy-ESPC$_{DET}$ performance is broadly comparable to the current operational system at tactical forecast lead times (0–7 days). However, performance of a few key acoustical measures (SLD, below layer sound speed gradient) important to Navy operations show a small degradation.
- Sea ice edge errors in Navy-ESPC$_{DET}$ are generally larger than the sea ice edge errors in GOFS 3.5. This is likely because GOFS 3.5 uses an updated version of CICE (version 5.1.2) compared to CICE version 4 in Navy-ESPC$_{DET}$.

For the Navy-ESPC$_{ENS}$ configuration, the main findings are as follows:

- Deterministic forecasts of the MJO are comparable to other leading centers (and better than NOAA CFSv2), with the ACOR dropping below 0.6 level after 20–30 days of the forecast.
- Deterministic forecasts for the AO, AAO, NAO, and PNA are also comparable to other leading centers and have ensemble mean ACORs greater than 0.6 in the 8- to 12-day period depending on index.
- The gain realized by using an ensemble mean over a deterministic forecast is smaller in Navy-ESPC than in other systems, illustrating the current substantial atmospheric underdispersion in the Navy-ESPC ensemble system.
- Global ocean temperatures averaged between 8 and 500 m in depth have RMSE values lower than climatology out to about 30 days in the forecast.
- Ensemble mean forecasts of SST compared to surface in situ observations have lower RMSE values than climatology out to 60 days, while deterministic (i.e., single member) forecasts RMSE values are lower than climatology out to 20 days.
- Pan Arctic and Antarctic sea ice extent diagnostics show skill over climatology from 11 to 60 days depending on metric and season. The BSS showed more skillful results compared to the IIEE metric, and Navy-ESPC$_{ENS}$ was more skillful in the northern hemisphere winter than summer for Arctic sea ice extent.

As noted above, a limitation of the Navy-ESPC$_{ENS}$ system is that it is underdispersive, particularly for atmospheric variables. As expected, the EDA system used is not sufficiently accounting for uncertainty in the initial state (i.e., underdispersive), and currently no method to account for model uncertainty is incorporated. Improving ensemble design is a major area of development for the next iteration of Navy-ESPC. For the atmosphere, methods successfully employed to account for model uncertainty in the stand-alone NAVGEM ensemble such as Stochastic Kinetic Energy Backscatter (Reynolds et al., 2011; Shutts, 2005) will be tested in Navy-ESPC. In addition, we are currently testing Analysis Correction-based Additive Inflation (ACAI) based on the work of Bowler (2017) (Crawford et al., 2020). Further, to improve the Navy-ESPC ensemble's sampling of initial condition uncertainty we are evaluating the Relaxation To-Prior Perturbations (RTPP) method (Whitaker & Hamill, 2012; Zhang et al., 2004) as well as the ET method (McLay et al., 2010). In the ocean, spread increases are currently being tested by introducing stochastic perturbations based on Lermusiaux (2006). We expect we will be able to realize even more benefit from the Navy-ESPC$_{ENS}$ forecasts once issues with underdispersion have been addressed (McLay & Shafer, 2016). We are also currently examining how ocean resolution affects spread in the boundary current regions by running an ensemble system at a ¼°.

In addition to ensemble design, multiple other aspects will be added or developed into Navy-ESPC. We plan on the addition of the wave model, Watch Watch III (Tolman, 1991; Tolman et al., 2002; Tolman & Chalikov, 1996; WW3DG, 2019), and upgrading CICE to a more recent version. Increased vertical and horizontal resolutions and possible physics upgrades will be tested for NAVGEM and HYCOM. In particular, we are testing updates for simulations of the stratosphere in NAVGEM, as a better simulation of the stratosphere may aid in subseasonal forecasting (Scaife et al., 2005). In addition, ocean tides will be added to the ensemble configuration. For data assimilation, we are exploring methods to share covariances across the air-ocean interface based on the work of Frolov et al. (2016).

Data Availability Statement

The S2S database is hosted at ECMWF as an extension of the TIGGE database and can be accessed online (https://apps.ecmwf.int/datasets/data/tigge). The Navy ESPC analysis and forecast data are stored at the Navy DoD Supercomputing Resource Center (DSRC).

References

Arbic, B. K., Alford, M. H., Ansong, J. K., Buijsman, M. C., Ciotti, R. B., Farrar, J. T., et al. (2018). *A primer on global internal tide and internal gravity wave continuum modeling in HYCOM and MITgcm*. Tallahassee, FL: Florida State University. Retrieved from http://purl.flvc.org/fsu/fd/FSU_libsubv1_scholarship_submission_1536242074_55feafcc

Auligne, T., McNally, A. P., & Dee, D. P. (2007). Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, *133*, 631–642. https://doi.org/10.1002/qj.56

Bennett, A. F. (2002). *Inverse modeling of the ocean and atmosphere*. Cambridge, UK: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/inverse-modeling-of-the-ocean-and-atmosphere/5CFE830AEB412531F6419D17DB620B90#:~:text=Book%20description,the%20hydrosphere%20or%20solid%20earth

Bleck, R. (2002). An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modelling*, *4*, 55–88. https://doi.org/10.1016/s1463-5003(01)00012-9

Bloom, S. C., Takacs, L. L., Silva, A. M. D., & Ledvina, D. (1996). Data assimilation using incremental analysis updates. *Monthly Weather Review*, *124*, 1256–1271. https://doi.org/10.1175/1520-0493(1996)124<1256:dauiau>2.0.co;2

Bond, N. A., & Vecchi, G. A. (2003). The influence of the Madden-Julian oscillation on precipitation in Oregon and Washington. *Weather and Forecasting*, *18*, 600–613. https://doi.org/10.1175/1520-0434(2003)018<0600:tiotmo>2.0.co;2

Bowler, N. E., Clayton, A. M., Jardak, M., Lee, E., Lorenc, A. C., Piccolo, C., et al. (2017). Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Quarterly Journal of the Royal Meteorological Society*, *143*, 1280–1302. https://doi.org/10.1002/qj.3004

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*, 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2

Brunet, G., Shapiro, M., Hoskins, B., Moncrieff, M., Dole, R., Kiladis, G. N., et al. (2010). Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bulletin of the American Meteorological Society*, *91*, 1397–1406. https://doi.org/10.1175/2010bams3013.1

Camargo, S. J., Robertson, A. W., Barnston, A. G., & Ghil, M. (2008). Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochemistry, Geophysics, Geosystems*, *9*, Q06V05. https://doi.org/10.1029/2007gc001861

Carman, J. C., Eleuterio, D. P., Gallaudet, T. C., Geernaert, G. L., Harr, P. A., Kaye, J. A., et al. (2017). The national Earth system prediction capability: Coordinating the giant. *Bulletin of the American Meteorological Society*, *98*, 239–252. https://doi.org/10.1175/bams-d-16-0002.1

Carnes, M., Helber, R. W., Barron, C. N., & Dastugue, J. M. (2010). *Validation test report for GDEM4* (NRL Memorandum Report NRL/MR/7330-10-9271). Mississippi, USA: Stennis Space Center. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA530343.pdf

Chassignet, E. P., Smith, L. T., Halliwell, G. R., & Bleck, R. (2003). North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *Journal of Physical Oceanography*, *33*, 2504–2526. https://doi.org/10.1175/1520-0485(2003)033<2504:NASWTH>2.0.CO;2

Chassignet, E. P., & Xu, X. B. (2017). Impact of horizontal resolution (1/12° to 1/50°) on Gulf Stream separation, penetration, and variability. *Journal of Physical Oceanography*, *47*, 1999–2021. https://doi.org/10.1175/JPO-D-17-0031.1

Chelton, D. B., Esbensen, S. K., Schlax, M. G., Thum, N., Freilich, M. H., Wentz, F. J., et al. (2001). Observations of coupling between surface wind stress and sea surface temperature in the eastern tropical Pacific. *Journal of Climate*, *14*, 1479–1498. https://doi.org/10.1175/1520-0442(2001)014<1479:OOCBSW>2.0.CO;2

Crawford, W., Frolov, S., McLay, J., Reynolds, C., Barton, N., & Bishop, C. H. (2020). Using analysis corrections to address model error in atmospheric forecasts. *Monthly Weather Review*, *148*. https://doi.org/10.1175/MWR-D-20-0008.1

Crosby, D. S., Breaker, L. C., & Gemmill, W. H. (1993). A proposed definition for vector correlation in geophysics: Theory and application. *Journal of Atmospheric and Oceanic Technology*, *10*, 355–367. https://doi.org/10.1175/1520-0426(1993)010<0355:APDFVC>2.0.CO;2

Cummings, J. A. (2005). Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*, 3583–3604. https://doi.org/10.1256/qj.05.105

Cummings, J. A., & Smedstad, O. M. (2013). Variational data assimilation for the global ocean. In S. K. Park, & L. Xu (Eds.), *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* (Vol. II, pp. 303–343). Berlin Heidelberg: Springer.

Cummings, J. A., & Smedstad, O. M. (2014). Ocean data impacts in global HYCOM. *Journal of Atmospheric and Oceanic Technology*, *31*, 1771–1791. https://doi.org/10.1175/JTECH-D-14-00011.1

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

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

El Ouaraini, R., & Berre, L. (2011). Sensitivity of ensemble-based variances to initial background perturbations. *Journal of Geophysical Research*, *116*, D15106. https://doi.org/10.1029/2010JD015075

Emanuel, K. A. (1991). A scheme for representing cumulus convection in large-scale models. *Journal of the Atmospheric Sciences*, *48*, 2313–2335. https://doi.org/10.1175/1520-0469(1991)048<2313:ASFRCC>2.0.CO;2

Emanuel, K. A., & Zivkovic-Rothman, M. (1999). Development and evaluation of a convection scheme for use in climate models. *Journal of the Atmospheric Sciences*, *56*, 1766–1782. https://doi.org/10.1175/1520-0469(1999)056<1766:DAEOAC>2.0.CO;2

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., & Edson, J. B. (2003). Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. *Journal of Climate*, *16*, 571–591. https://doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2

Feliks, Y. (1985). On the Rossby radius of deformation in the ocean. *Journal of Physical Oceanography*, *15*, 1605–1607. https://doi.org/10.1175/1520-0485(1985)015<1605:OTRROD>2.0.CO;2

Flatau, M., & Kim, Y. J. (2013). Interaction between the MJO and Polar Circulations. *Journal of Climate*, *26*, 3562–3574. https://doi.org/10.1175/JCLI-D-11-00508.1

Fox, D. N., Teague, W. J., Barron, C. N., Carnes, M. R., & Lee, C. M. (2002). The modular ocean data assimilation system (MODAS). *Journal of Atmospheric and Oceanic Technology*, *19*, 240–252. https://doi.org/10.5670/oceanog.2002.33

Frolov, S., Bishop, C. H., Holt, T., Cummings, J., & Kuhl, D. (2016). Facilitating strongly coupled ocean-atmosphere data assimilation with an interface solver. *Monthly Weather Review*, *144*, 3–20. https://doi.org/10.1175/MWR-D-15-0041.1

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., & Jung, T. (2016). Predictability of the Arctic sea ice edge. *Geophysical Research Letters*, *43*, 1642–1650. https://doi.org/10.1002/2015GL067232

Guan, B., Waliser, D. E., Molotch, N. P., Fetzer, E. J., & Neiman, P. J. (2012). Does the Madden-Julian Oscillation influence wintertime atmospheric rivers and snowpack in the Sierra Nevada? *Monthly Weather Review*, *140*, 325–342. https://doi.org/10.1175/MWR-D-11-00087.1

Hallberg, R. (2013). Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, *72*, 92–103. https://doi.org/10.1016/j.ocemod.2013.08.007

Halliwell, G. R. (2004). Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). *Ocean Modelling*, *7*, 285–322. https://doi.org/10.1016/j.ocemod.2003.10.002

Han, J., & Pan, H. L. (2011). Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Weather and Forecasting*, *26*, 520–533. https://doi.org/10.1175/WAF-D-10-05038.1

Hebert, D. A., Allard, R. A., Metzger, E. J., Posey, P. G., Preller, R. H., Wallcraft, A. J., et al. (2015). Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the US Navy's Arctic Cap Nowcast/Forecast System. *Journal of Geophysical Research: Oceans*, *120*, 327–8345. https://doi.org/10.1002/2015JC011283

Helber, R. W., Townsend, T. L., Barron, C. N., & Carnes, M. R. (2013). *Validation test report for the Improved Synthetic Ocean Profile (ISOP) System, Part I: Synthetic profile methods and algorithm* (NRL/MR/7320—13-9364). Mississippi, USA: Stennis Space Center. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA585251.pdf

Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., & Kasheta, T. (2007). Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrological Processes*, *21*, 1576–1586. https://doi.org/10.1002/hyp.6720

Henderson, G. R., Barrett, B. S., & Lafleur, D. M. (2014). Arctic sea ice and the Madden-Julian Oscillation (MJO). *Climate Dynamics*, *43*, 2185–2196. https://doi.org/10.1007/s00382-013-2043-y

Henderson, G. R., Barrett, B. S., Lois, A., & Elsaawy, H. (2018). Time-lagged response of the Antarctic and high-latitude atmosphere to tropical MJO convection. *Monthly Weather Review*, *146*, 1219–1231. https://doi.org/10.1175/MWR-D-17-0224.1

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., et al. (2017). Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, *120*, 120–136. https://doi.org/10.1016/j.ocemod.2017.11.002

Hogan, P. J., & Hurlburt, H. E. (2000). Impact of upper ocean-topographical coupling and isopycnal outcropping in Japan/East sea models with 1/8° to 1/64° resolution. *Journal of Physical Oceanography*, *30*, 2535–2561. https://doi.org/10.1175/1520-0485(2000)030<2535:IOUOTC>2.0.CO;2

Hogan, T., Liu, M., Ridout, J. A., Peng, M. S., Whitcomb, T. R., Ruston, B. C., et al. (2014). The navy global environmental model. *Oceanography*, *27*, 116–125. https://doi.org/10.5670/oceanog.2014.73

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., & Mitchell, H. L. (1996). A system simulation approach to ensemble prediction. *Monthly Weather Review*, *124*, 1225–1242. https://doi.org/10.1175/1520-0493(1996)124<1225:assate>2.0.co;2

Hunke, E. C., & Lipscomb, W. (2008). *CICE: The Los Alamos sea ice model, documentation and software user's manual, version 4.0. Techincal Report, LA-CC-06-012*. Los Alamos, NM: Los Alamos National Laboratory. Retrieved from http://www.cesm.ucar.edu/models/ccsm4.0/cice/ice_usrdoc.pdf

Hurrell, J., Meehl, G. A., Bader, D., Delworth, T. L., Kirtman, B., & Wielicki, B. (2009). A unified modeling approach to climate system prediction. *Bulletin of the American Meteorological Society*, *90*(12), 1819–1832. https://doi.org/10.1175/2009bams2752.1

Infanti, J. M., & Kirtman, B. P. (2016). Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America. *Journal of Geophysical Research: Atmospheres*, *121*, 12,690–12,701. https://doi.org/10.1002/2016JD024932

Janiga, M. A., Schreck, C. J., Ridout, J. A., Flatau, M., Barton, N. P., Metzger, E. J., & Reynolds, C. A. (2018). Subseasonal forecasts of convectively coupled equatorial waves and the MJO: Activity and predictive skill. *Monthly Weather Review*, *146*, 2337–2360. https://doi.org/10.1175/mwr-d-17-0261.1

Jiang, X., Waliser, D. E., Xavier, P. K., Petch, J., Klingaman, N. P., Woolnough, S. J., et al. (2015). Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations. *Journal of Geophysical Research: Atmospheres*, *120*, 4718–4748. https://doi.org/10.1002/2014JD022375

Johnson, R. H., & Ciesielski, P. E. (2013). Structure and properties of Madden-Julian Oscillations deduced from DYNAMO sounding arrays. *Journal of the Atmospheric Sciences*, *70*, 3157–3179. https://doi.org/10.1175/jas-d-13-065.1

Kain, J. S., & Fritsch, J. M. (1990). A one-dimensional entraining detraining plume model and its application in convective parameterization. *Journal of the Atmospheric Sciences*, *47*, 2784–2802. https://doi.org/10.1175/1520-0469(1990)047<2784:aodepm>2.0.co;2

Kain, J. S., & Fritsch, J. M. (1993). Convective parameterization for mesoscale models: The Kain-Fritsch scheme. In K. A. Emanuel & D. J. Raymond (Eds.), *The representation of cumulus convection in numerical models* (pp. 165–170). Boston, MA: American Meteorological Society. Retrieved from https://link.springer.com/chapter/10.1007/978-1-935704-13-3_16

Kara, A. B., Hurlburt, H. E., & Wallcraft, A. J. (2005). Stability-dependent exchange coefficients for air-sea fluxes. *Journal of Atmospheric and Oceanic Technology*, *22*, 1080–1094. https://doi.org/10.1175/jtech1747.1

Kim, H., Janiga, M. A., & Pegion, K. (2019). MJO propagation processes and mean biases in the SubX and S2S reforecasts. *Journal of Geophysical Research: Atmospheres*, *124*, 9314–9331. https://doi.org/10.1029/2019JD031139

Kim, J. H., Ho, C. H., Kim, H. S., Sui, C. H., & Park, S. K. (2008). Systematic variation of summertime tropical cyclone activity in the western north pacific in relation to the Madden-Julian Oscillation. *Journal of Climate*, *21*, 1171–1191. https://doi.org/10.1175/2007jcli1493.1

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., & Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *142*, 65–78. https://doi.org/10.1002/qj.2629

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

Lermusiaux, P. F. J. (2006). Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *Journal of Computational Physics*, *217*, 176–199. https://doi.org/10.1016/j.jcp.2006.02.010

Liebmann, B., & Smith, C. A. (1996). Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*, *77*, 1275–1277. https://doi.org/10.2307/26233278

Lim, Y., Son, S. W., & Kim, D. (2018). MJO prediction skill of the subseasonal-to-seasonal prediction models. *Journal of Climate*, *31*, 4075–4094. https://doi.org/10.1175/jcli-d-17-0545.1

Lin, H., Brunet, G., & Derome, J. (2009). An observed connection between the North Atlantic Oscillation and the Madden-Julian Oscillation. *Journal of Climate*, *22*, 364–380. https://doi.org/10.1175/2008jcli2515.1

Lin, H., Gagnon, N., Beauregard, S., Muncaster, R., Markovic, M., Denis, B., & Charron, M. (2016). GEPS-based monthly prediction at the Canadian Meteorological Centre. *Monthly Weather Review*, *144*, 4867–4883. https://doi.org/10.1175/mwr-d-16-0138.1

Louis, J. F., Tiedtke, M., & Geleyn, J. F. (1982). A short history of the operational PBL parameterization at ECMWF. In *ECMWF Workshop on Planetary Boundary Parameterizations* (pp. 59–79). Reading, UK: ECMWF. Retrieved from https://www.ecmwf.int/en/elibrary/10845-short-history-pbl-parameterization-ecmwf

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*, 1072–1084. https://doi.org/10.1002/qj.2396

McLay, J., Bishop, C. H., & Reynolds, C. A. (2010). A local formulation of the ensemble transform (ET) analysis perturbation scheme. *Weather and Forecasting*, *25*, 985–993. https://doi.org/10.1175/2010waf2222359.1

McLay, J., & Shafer, J. (2016). *NAVGEM ensemble prediction system (EPS): Upgrades to model, ensemble transform, and resolution. Validation Test Report*. Monterey, CA: Naval Research Laboratory.

Metzger, E. J., Hurlburt, H. E., Wallcraft, A. J., Shriver, J. F., Townsend, T. L., Smedstad, O. M., et al. (2019). *Validation test report for the Global Ocean Forecast System 3.5–1/25° HYCOM/CICE with tides* (NRL Memorandum Report NRL/MR/7320--19-9975). Mississippi, USA: Stennis Space Center.

Metzger, E. J., Smedstad, O. M., Thoppil, P., Hurlburt, H., Cummings, J., Walcraft, A., et al. (2014). US Navy Operational Global Ocean and Arctic Ice Prediction Systems. *Oceanography*, *27*, 32–43. https://doi.org/10.5670/oceanog.2014.66

Metzger, E. J., Wallcraft, A., Posey, P., Smedstad, O. M., & Franklin, D. S. (2013). *The switchover from NOGAPS to NAVGEM 1.1 atmospheric forcing in GOFS and ACNFS* (NRL Memorandum Report NRL/MR/7320-13–9486). Mississippi, USA: Stennis Space Center. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a591601.pdf

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., et al. (2011). *The new ECMWF seasonal forecast system (System 4)*, *Technical Memorandum* (Vol. 656, pp. 1–49). Reading, UK: ECMWF. Retrieved from https://www.ecmwf.int/en/elibrary/11209-new-ecmwf-seasonal-forecast-system-system-4

Moorthi, S., Pan, H.-L., & Caplan, P. (2001). Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. In *NWS Technical Procedures Bulletin* (Vol. 484, pp. 1–14). Washington, DC: National Weather Service. Retrieved from http://144.76.120.80/city/mos/artikel/tpb484.pdf

Mundhenk, B. D., Barnes, E. A., & Maloney, E. D. (2016). All-season climatology and variability of atmospheric river frequencies over the North Pacific. *Journal of Climate*, *29*, 4885–4903. https://doi.org/10.1175/JCLI-D-15-0655.1

Oliver, H., Shin, M., Matthews, D., Sanders, O., Bartholomew, S., Clark, A., et al. (2019). Workflow automation for cycling systems. *Computing in Science & Engineering*, *21*, 7–21. https://doi.org/10.1109/MCSE.2019.2906593

Partington, K., Flynn, T., Lamb, D., Bertoia, C., & Dedrick, K. (2003). Late twentieth century Northern Hemisphere sea-ice record from U. S. National Ice Center ice charts. *Journal of Geophysical Research*, *108*, 3343. https://doi.org/10.1029/2002JC001623

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., et al. (2019). The subseasonal experiment (SubX): A multi-model subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, *100*, 2043–2060. https://doi.org/10.1175/BAMS-D-18-0270.1

Peng, M. S., Ridout, J. A., & Hogan, T. F. (2004). Recent modifications of the Emanuel convective scheme in the Navy Operational Global Atmospheric Prediction System. *Monthly Weather Review*, *132*, 1254–1268. https://doi.org/10.1175/1520-0493(2004)132<1254:RMOTEC>2.0.CO;2

Reynolds, C. A., McLay, J. G., Goerss, J. S., Serra, E. A., Hodyss, D., & Sampson, C. R. (2011). Impact of resolution and design on the US Navy Global Ensemble performance in the tropics. *Monthly Weather Review*, *139*, 2145–2155. https://doi.org/10.1175/2011MWR3546.1

Ridout, J. A., Jin, Y., & Liou, C. S. (2005). A cloud-base quasi-balance constraint for parameterized convection: Application to the Kain-Fritsch cumulus scheme. *Monthly Weather Review*, *133*, 3315–3334. https://doi.org/10.1175/MWR3034.1

Ridout, J. A., & Reynolds, C. A. (1998). Western Pacific warm pool region sensitivity to convective triggering by boundary layer thermals in the NOGAPS atmospheric GCM. *Journal of Climate*, *11*(7), 1553–1573. https://doi.org/10.1175/1520-0442(1998)011<1553:wpwprs>2.0.co;2

Ritchie, H. (1987). Semi-Lagrangian advection on a gaussian grid. *Monthly Weather Review*, *115*, 608–619. https://doi.org/10.1175/1520-0493(1987)115<0608:SLAOAG>2.0.CO;2

Ritchie, H. (1988). Application of the semi-Lagrangian method to a spectral model of the shallow-water equations. *Monthly Weather Review*, *116*, 1587–1598. https://doi.org/10.1175/1520-0493(1988)116<1587:AOTSLM>2.0.CO;2

Ritchie, H. (1991). Application of the semi-Lagrangian method to a multilevel spectral primitive-equations model. *Quarterly Journal of the Royal Meteorological Society*, *117*, 91–106. https://doi.org/10.1002/qj.49711749705

Ritchie, H., Temperton, C., Simmons, A., Hortal, M., Davies, T., Dent, D., & Hamrud, M. (1995). Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Monthly Weather Review*, *123*, 489–514. https://doi.org/10.1175/1520-0493(1995)123<0489:IOTSLM>2.0.CO;2

Roemmich, D., & Gilson, J. (2009). The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Progress in Oceanography*, *82*, 81–100. https://doi.org/10.1016/j.pocean.2009.03.004

Rosmond, T., & Xu, L. (2006). Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus Ser. A-Dynamic Meteorology and Oceanography*, *58*, 45–58. https://doi.org/10.1111/j.1600-0870.2006.00148.x

Ruti, P., Rixen, M., & de Coning, E. (2019). In A. W. Robertson & F. Vitart (Eds.), *Sub-seasonal to seasonal prediction* (pp. 1–585). Elsevier.

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*, 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Scaife, A. A., Knight, J. R., Vallis, G. K., & Folland, C. K. (2005). A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophysical Research Letters*, *32*, L18715. https://doi.org/10.1029/2005GL023226

Shapiro, M., Shukla, J., Brunet, G., Nobre, C., Béland, M., Dole, R., et al. (2010). An Earth-system prediction initiative for the twenty-first century. *Bulletin of the American Meteorological Society*, *91*(10), 1377–1388. https://doi.org/10.1175/2010bams2944.1

Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, *131*, 3079–3102. https://doi.org/10.1256/qj.04.106

Smith, G. C., Bélanger, J.-M., Roy, F., Pellerin, P., Ritchie, H., Onu, K., et al. (2018). Impact of coupling with an ice-ocean model on global medium-range NWP forecast skill. *Monthly Weather Review*, *146*, 1157–1180. https://doi.org/10.1175/MWR-D-17-0157.1

Sobel, A. H., & Maloney, E. D. (2000). Effect of ENSO and the MJO on western North Pacific tropical cyclones. *Geophysical Research Letters*, *27*, 1739–1742. https://doi.org/10.1029/1999GL011043

Sud, Y. C., & Walker, G. K. (1997). Simulation errors associated with the neglect of oceanic salinity in an atmospheric GCM. *Earth Interactions*, *1*, 1–19. https://doi.org/10.1175/1087-3562(1997)001<0001:SEAWTN>2.3.CO;2

Sun, S., Bleck, R., Benjamin, S. G., Green, B. W., & Grell, G. A. (2018). Subseasonal forecasting with an icosahedral, vertically quasi-Lagrangian coupled model. Part I: Model overview and evaluation of systematic errors. *Monthly Weather Review*, *146*(1), 601–1617. https://doi.org/10.1175/MWR-D-18-0006.1

Theurich, G., DeLuca, C., Campbell, T., Liu, F., Saint, K., Vertenstein, M., et al. (2016). The Earth system prediction suite: Toward a coordinated U.S. modeling capability. *Bulletin of the American Meteorological Society*, *97*, 1229–1247. https://doi.org/10.1175/BAMS-D-14-00164.1

Thoppil, P. G., Richman, J. G., & Hogan, P. J. (2011). Energetics of a global ocean circulation model compared to observations. *Geophysical Research Letters*, *38*, L15607. https://doi.org/10.1029/2011GL048347

Tolman, H. L. (1991). A 3rd-generation model for wind-waves on slowly varying, unsteady, and inhomogeneous depths and currents. *Journal of Physical Oceanography*, *21*, 782–797. https://doi.org/10.1175/1520-0485(1991)021<0782:ATGMFW>2.0.CO;2

Tolman, H. L., Balasubramaniyan, B., Burroughs, L. D., Chalikov, D. V., Chao, Y. Y., Chen, H. S., & Gerald, V. M. (2002). Development and implementation of wind-generated ocean surface wave models at NCEP. *Weather and Forecasting*, *17*, 311–333. https://doi.org/10.1175/1520-0434(2002)017<0311:DAIOWG>2.0.CO;2

Tolman, H. L., & Chalikov, D. (1996). Source terms in a third-generation wind wave model. *Journal of Physical Oceanography*, *26*, 2497–2518. https://doi.org/10.1175/1520-0485(1996)026<2497:STIATG>2.0.CO;2

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*, 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Waliser, D., Sperber, K., Hendon, H., Kim, D., Maloney, E., Wheeler, M., et al. (2009). MJO simulation diagnostics. *Journal of Climate*, *22*, 3006–3030. https://doi.org/10.1175/2008JCLI2731.1

Wayand, N. E., Bitz, C. M., & Blanchard-Wrigglesworth, E. (2019). A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophysical Research Letters*, *46*, 3298–3307. https://doi.org/10.1029/2018gl081565

Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, *132*, 1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:aarmmi>2.0.co;2

While, J., & Martin, M. J. (2019). Variational bias correction of satellite sea-surface temperature data incorporating observations of the bias. *Quarterly Journal of the Royal Meteorological Society*, *145*(723), 2733–2754. https://doi.org/10.1002/qj.3590

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*, 3078–3089. https://doi.org/10.1175/mwr-d-11-00276.1

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Elsevier Academic Press. Retrieved from https://www.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-385022-5

WW3DG (2019). User Manual and System Documentation of WAVEWATCH III version 6.07, The {WAVEWATCH III} Development Group. Tech. Note (p. 326). Appendices, NOAA/NWS/NCEP/MMAB

Xu, L., Rosmond, T., & Daley, R. (2005). Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus Ser. A-Dynamic Meteorology and Oceanography*, *57*, 546–559. https://doi.org/10.1111/j.1600-0870.2005.00123.x

Zhang, F. Q., Snyder, C., & Sun, J. Z. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, *132*, 1238–1253. https://doi.org/10.1175/1520-0493(2004)132<1238:ioieao>2.0.co;2

Zhu, Y. J., Zhou, X., Li, W., Hou, D., Melhauser, C., Sinsky, E., et al. (2018). Toward the improvement of subseasonal prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System. *Journal of Geophysical Research: Atmospheres*, *123*, 6732–6745. https://doi.org/10.1029/2018JD028506