GEOS-S2S Version 2: The GMAO High Resolution Coupled Model and Assimilation System for Seasonal Prediction

Andrea Molod[1], Eric Hackert[1], Lauren C. Andrews[1], Robin M. Kovach[1,2], Bin Zhao[1,2], Zhao Li[1,2], Young-Kwon Lim[1,3], Anna Borovikov[1,2], Guillaume Vernieres[4,5], Jelena Marshak[1], Randal Koster[1], James Carton[6], Richard Cullather[1,6], Donifan Barahona[1], Siegfried Schubert[1,2], Deepthi Achuthavarier[1,7], Yury Vikhliaev[1,7], Julliana L. M. Freire[1,7], Karla M. Longo[1,8]

[1]NASA, Goddard Space Flight Center, Greenbelt, MD 20771
[2]SSAI,Science Systems and Applications, Inc. Lanham, MD 20706
[3]Goddard Earth Sciences Technology and Research, I. M. Systems Group, College Park, MD 20740
[4]UCAR Boulder Co.
[5]JCSDA/UCAR/NOAA, College Park, MD
[6]University of Maryland, College Park, MD
[7]Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, MD
[8]Center for Weather Forecast and Climate Studies, National Institute for Space Research (INPE), Cachoeira Paulista, Sao Paulo, Brazil

Key Points:

- Atmosphere-Ocean Modeling
- Atmosphere-Ocean Data Assimilation
- Seasonal and Subseasonal Prediction

Corresponding author: Andrea Molod, `andrea.molod@nasa.gov`

Abstract
TEXT

1 Introduction

The Global Modeling and Assimilation Office (GMAO) has recently released a new version of the Goddard Earth Observing System (GEOS) Sub-seasonal to Seasonal prediction (S2S) system, GEOS-S2S Version 2. The S2S system includes an Atmosphere-Ocean General Circulation Model (AOGCM) an ocean data assimilation system (ODAS), and a methodology for weakly coupled Atmosphere-Ocean Coupled Data Assimilation (AO-CDAS). This new version of GEOS-S2S represents a substantial improvement in performance and system infrastructure relative to the previous version, retroactively named GEOS-S2S Version 1, that was described in *Borovikov et al.* [2017].

The GEOS system has a long history of being successfully employed in seasonal prediction efforts and contributing to multi-system ensemble projects. For example, the GEOS system has been a participating model in the North American Multi-model Ensemble [*Kirtman et al.*, 2014] since that project's inception in 2011. The GEOS forecasts also routinely contribute to various other national and international multi-model ensembles including the multi-model forecast products at the International Research Institute for Climate and Society (IRI) of Columbia University and those of the APEC Climate Center, Korea, thus enabling rigorous evaluations of the forecast skill and model biases and orienting the performance of the GEOS-S2S system relative to other state-of-the art systems. The new system described here (GEOS-S2S Version 2) builds upon the already established experimental seasonal prediction system at the GMAO [*Borovikov et al.*, 2017], and further expands it to produce near-real-time weekly initialized forecasts at the subseasonal timescale, also facilitating GMAO's participation in the NOAA's experimental subseasonal multi-model ensemble project (http://cola.gmu.edu/kpegion/subx/index.html).

In this paper, we describe the GEOS-S2S Version 2 system, emphasizing the improvements over Version 1, and assess the performance of climate, forecasts and data assimilation. Section 2 presents a description of the model and data assimilation system, section 3 describes the experiments performed with the new system that will be analyzed here, results of the experiments are presented and compared with results from the old system and observational estimates in sections 4, 5 and 6 and the study is summarized in section 7.

2 Description of Coupled Model and Data Assimilation System

The GEOS AOGCM and Data Assimilation System (AODAS) are developed to simulate the earth system on a wide range (synoptic to decadal) of time scales. The main components of the GEOS-S2S are the GEOS AOGCM, which in-turn consists of the GEOS atmospheric general circulation model [*Molod et al.*, 2015; *Rienecker et al.*, 2008], the catchment land surface model [*Koster et al.*, 2000], and the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol model REFS. The ocean general circulation model is the Modular Ocean Model-5 (MOM5) developed by the Geophysical Fluid Dynamics Laboratory [*Griffies et al.*, 2005; *Griffies*, 2012], coupled to the Community Ice CodE-4 (CICE4) sea ice model developed at Los Alamos National Lab (LANL, [*Hunke and Lipscomb*, 2008]). The atmospheric data assimilation component is the pre-existing "Forward Processing for Instrument Teams" near-real time assimilation (https://gmao.gsfc.nasa.gov/products/GEOS-5_FP-IT_details.php), and the ocean data assimilation follows the Local Ensemble Transform Kalman Filter (LETKF) [*Penny et al*, 2013]. All components are coupled together using the Earth System Modeling Framework (ESMF, [*Hill et al.*, 2004]) and the MAPL interface layer [*Suarez et al.*, 2007]. Various operational centers are developing AO-CDAS systems [*Dee et al.*, 2014; *Brassington et al.*, 2015], in which different components (e.g., atmosphere and ocean) of the earth system are analyzed separately [*Laloyaux et al.*, 2016; *Lea et al.*, 2015] or simultaneously [*Sluka et al.*, 2016; *Wada and Kunii*, 2017]. The GEOS-S2S AO-CDAS relies on an pre-computed atmospheric analysis, performing an analysis of the ocean state only. Each of the components of the GEOS S2S Version 2 system will be described here in varying amounts of detail.

2.1 Coupled Atmosphere-Ocean General Circulation Model

2.1.1 Atmospheric, Land and Aerosol Models

The version of the GEOS AGCM that is used as part of the GEOS Seasonal to Subseasonal (S2S) prediction system Version 2 contains algorithms that describe atmospheric transport and dynamics, atmospheric physical processes, a land model and an interactive aerosol model. The algorithm for large-scale transport and dynamics in the current GEOS AGCM is an adaptation of the flux-form semi-Lagrangian (FFSL) finite-volume (FV) dynamics of *Lin* [2004], adapted for a cubed sphere horizontal discretization [*Putman and Lin*, 2007]. A comprehensive description of baseline versions of the physical parameterizations is found in *Rienecker et al.* [2008], and the updates to a recent version of the AGCM are found in *Molod et al.* [2015]. GEOS AGCM physical parameterizations include schemes for convection, large scale precipitation and cloud cover, longwave and shortwave radiation, turbulence, and gravity wave drag.

Convection is parameterized using the Relaxed Arakawa-Schubert scheme [*Moorthi and Suarez*, 1992], which contains an updraft-only cloud model and a quasi-equilibrium closure. The frequency and intensity of deep convection is governed by a stochastic Tokioka-type trigger function [*Tokioka et al.*, 1988] as suggested by *Bacmeister and Stephens* [2011]. Prognostic cloud cover and cloud water and ice are determined by the two moment parameterization of *Barahona et al.* [2014], which predicts both the amount and number concentration of cloud water and cloud ice. Processes include cloud particle nucleation, large scale condensation, evaporation, autoconversion and accretion of cloud water and ice, sedimentation of cloud ice, and re-evaporation of falling precipitation. The probability distribution function (PDF) for total water that governs the condensation and evaporation processes is described by *Molod* [2012].

Longwave radiative processes are described by *Chou and Suarez* [1994], and include absorption due to cloud water, water vapor, aerosols, carbon dioxide (CO_2), ozone (O_3), nitrous oxide (N_2O), methane (CH_4), chlorofluorocarbons CFC-11 and CFC-12, and hydroclorofluorocarbon HCFC-22. Shortwave radiative transfer is from *Chou* [1990, 1992], and includes absorption by water vapor, cloud water, O_3, CO_2, molecular oxygen (O_2), and aerosols, and scattering by cloud water and aerosols.

The turbulence parameterization is based on the Lock scheme [*Lock et al.*, 2000] combined with the Richardson-number based algorithm of *Louis and Geleyn* [1982]. The former includes a representation of non-local mixing driven by both surface fluxes and cloud-top processes in unstable layers, either coupled to or decoupled from the surface. It was extended in GEOS to include moist heating and entrainment in the unstable surface parcel calculations which determine the depth of unstable layers. The latter is a first-order local scheme, and its effect is mostly felt just above the surface layer and in regions of shear-generated turbulence. The turbulent length scale that governs its behavior is a function of the planetary boundary layer height at the previous time step [*Molod et al.*, 2015], which is diagnosed based on the profile of eddy diffusivity over the ocean and on a bulk Richardson number threshold over land [*McGrath-Spangler and Molod*, 2014]. The Monin-Obukhov surface layer parameterization is described by *Helfand and Schubert* [1995] and includes the effects of a viscous sublayer for heat and moisture transport over all surfaces except land. Ocean surface roughness is determined by a blend of the algorithms of *Large and Pond* [1981] and *Kondo* [1975], modified in the midrange wind regime according to *Garfinkel et al.* [2011] and in the high wind regime according to *Molod et al.* [2013].

The gravity wave drag parameterization computes momentum and heat deposition due to orographic [*McFarlane*, 1987] and nonorographic [*Garcia and Boville*, 1994] waves. The background drag profile that generates an internal quasi-biennial oscillation (QBO) is described by *Molod et al.* [2015]. They demonstrate that downward propagation of the zonal wind anomalies is realistic, but phase speeds are slower and amplitudes are larger than those observed.

The Land Surface Model from *Koster et al.* [2000] is a catchment-based scheme that treats subgrid scale heterogeneity in surface moisture statistically. Glacial thermodynamic process are parameterized using an adaptation of the *Stieglitz et al.* [2001] snow model to glacial ice [*Cullather et al.*, 2014], and the catchment and glacier models are each coupled to the multi-layer snow model developed by *Stieglitz et al.* [2001]. Sea ice albedos in the northern hemisphere are from the monthly mean observations of Duynkerke and de Roode (2001). add bibitem

GOCART.....

2.1.2 Ocean and Sea Ice Models

The ocean component of the GEOS AOGCM is the Modular Ocean Model version 5 (MOM5) developed at Geophysical Fluid Dynamics Laboratory, [*Griffies et al.*, 2005; *Griffies*, 2012]. It is a hydrostatic primitive equations model with a staggered Arakawa B-grid, [*Mesinger and Arakawa*, 1976] and vertical coordinate based on depth. A tripolar grid is used to resolve the Arctic Ocean without polar filtering, [*Murray*, 1996]. The model uses a three level time stepping scheme. The ocean surface boundary is computed as an explicit free surface with real fresh water forcing. The topography is represented as a partial bottom step to better represent topographically influenced advective and wave processes. Vertical mixing follows non-local K-profile parametrization of *Large et al.* [1994] and includes parametrization of tidal mixing on continental shelves. Horizontal mixing uses the isoneutral method developed by *Gent and McWilliams* [1990]. The horizontal viscosity uses anisotropic scheme of *Large et al.* [2001] for better representation of equatorial currents. The exchange with marginal seas is parametrized under coarse resolution as discussed in *Griffies* [2012].

The sea ice component of the GEOS AOGCM is the CICE 4.1 developed by the Los Alamos National Laboratory, [*Hunke and Lipscomb*, 2008]. The model includes several interacting components: a thermodynamic model that computes local growth rates of snow and ice due to vertical conductive, radiative and turbulent fluxes, along with snowfall; a model of ice dynamics, which 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 strain.

The ocean and atmosphere exchange fluxes of momentum, heat and fresh water through a ”skin layer” interface which includes a parameterization of the diurnal cycle [*Price et al.*, 1978].

2.2 Coupled Atmosphere and Ocean Data Assimilation

2.2.1 Data Assimilation Method

Similar to Version 1, the GEOS S2S Version 2 AO-CDAS is a weakly coupled atmosphere-ocean data assimilation system, as depicted in Figure 1. During all stages of the data assimilation, the coupled AOGCM is performing the simulations, and the AO-CDAS includes an ocean predictor sequence (the green line across the top of the figure), during which the atmospheric state is "replayed" using an "intermittent replay" [**??**] to GMAO’s "Forward Processing for Instrument Teams" (FPIT) atmospheric reanalysis. The land is driven using observed CMAP precipitation [**?**], [*Xie and Arkin*, 1997]. After a 5-day segment, the ocean analysis increments are computed (see below for a description) and the AO-CDAS returns to the beginning of the 5-day segment to perform the corrector segment (blue arrow across the bottom) using the Incremental Analysis Update (IAU) method of *Bloom et al.* [1996].

Throughout the predictor and corrector steps (depicted in Figure 1), as the coupled atmosphere-ocean model is integrated in time, the sea surface temperature (SST) is strongly relaxed (with a 1-day relaxation time-scale) to the MERRA-2 [*Gelaro et al.*, 2017] SST (also used in FPIT) so that the simulated atmosphere in this coupled system is as consistent as possible with the FPIT atmospheric reanalysis. It should be noted that the present GEOS S2S-2 system has no relaxation to salinity. As the predictor segment proceeds, every 6-hours the departure of the model trajectory (i.e., background field) from observations is gathered, via the so-called ocean *observers*. Using the background and monthly averaged anomalies of 20 freely-run AOGCMs (with 0.5° × 0.5° × 40 levels resolution and no assimilation), we generate ensemble members that are centered about the current background state. The ocean observers are run for all the ensemble members, resulting in an ensemble of innovations (i.e., an ensemble of the departure of observations minus the ensemble member). These innovations combined with the above calculated ensemble members are then used to perform an LETKF analysis. Sea ice fraction (AICE) is replaced by concentrations calculated using the NASA Team algorithm [*Cavalieri et al.*, 1996; *Maslanik and Stroeve*, 1999] at the analysis step in order to provide optimal ice fields. The coupled model is then rewound to the start of the assimilation cycle to perform the corrector step, during which time the ocean temperature and salinity increments are evenly applied to the ocean trajectory over the first 18 hours of the corrector step to reduce any shocks to the system. The coupled system is then allowed to evolve constrained by FPIT forcing and its sea surface temperature for the remainder of the 5-day segment. This process is then repeated over the next 5-day data assimilation window, cycling over time.

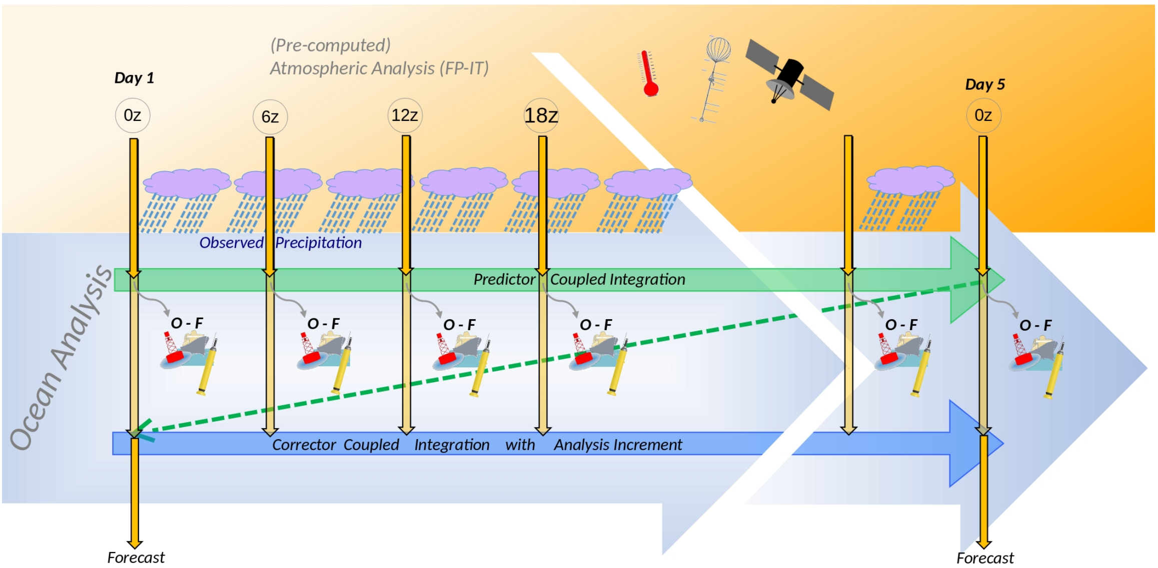


Figure 1. Schematic of Coupled Data Assimilation Methodology

2.2.2 Ocean Data Assimilation Technique

The Ocean Data Assimilation System (ODAS) used by the GEOS S2S-2 system at the GMAO follows the Local Ensemble Transform Kalman Filter (LETKF) developed by *Penny et al* [2013] for ocean applications. Unlike *Penny et al* [2013], our ensemble members are derived from an already existing trajectory of a coupled model run that was computed without any data assimilation, that is, from a free-running model. Thus, our version of the assimilation procedure more closely resembles the Ensemble Optimal Interpolation (EnOI, e.g., *Keppenne et al.* [2008], *Karspeck et al.* [2013], *Vernieres et al.* [2012]). However, the GEOS-S2S-2 ODAS closely matches the LETKF for other aspects like the localized influence of an observation on the analysis. More details regarding the LETKF can be found in *Ott et al.* [2004], *Hunt et al.* [2007], and *Penny et al.* [2015].

The GEOS-S2S-2 ocean analysis includes an assimilation of various in situ profile observations summarized in Table 1. Tropical Atmosphere/Ocean (TAO), Prediction and Research Moored Array in the Atlantic (PIRATA), and Research Moored Array for African-Asian-Australian Monsoon Analysis(RAMA) and Prediction refer to fixed tropical mooring arrays that are designed to observe temperature and salinity at depth within the oceanic waveguide [*McPhaden et al.*, 2010] in the Pacific, Atlantic and Indian Oceans, respectively. Expendable Bathythermograph (XBT) data are instruments released from research cruises or volunteer observing ships and so are generally found in regions of repeat transects. Unlike most of the other profile data, XBT data typically only includes temperature observations. Conductivity, temperature, and depth (CTD) data are collected by cast from research cruises so are also generally sparse in nature. On the other hand, the major profile data source is from the Argo array [*Roemmich et al.*, 2009]. The Argo array is made of thousands of autonomous profiling Lagrangian floats that descend and ascend through the water column on a regular schedule (typically 10 days, 5m - 2000m), saving temperature, salinity and pressure observations as they travel. When the float surfaces, it sends its observations to the Global Telecommunications System and is made available at near-real time.

An example of typical data coverage for the near-real time 5-day assimilation cycle is shown in Figure 2 (in this case for December 21-25, 2017 run on December 29, 2017). Note that most of the spatial resolution comes from the Lagrangian drifters that make up the Argo array (59567 observations whose locations are shown by light blue dots in Figure 2) whereas much of the temporal coverage can be accounted for by the fixed moorings (e.g. TAO in the Pacific, PIRATA in the Atlantic, RAMA in the tropical Indian Ocean - 13530 total number of observations). For this work, the mooring data are averaged daily to reduce the high frequency signal found in the hourly mooring observations. Also note the lack of XBT and CTD data since these data are typically available about 2 months in arrears and so these data were not available to the near-real-time ODAS. It is also interesting to note that there are approximately 20% fewer salinity observations than temperature observations (compare Figure 2 left and middle panels for light blue dots) since salinity sensors are more likely to foul than temperature.

In addition to in situ observations, surface topography observations from satellite altimetry have been utilized to help determine the general ocean circulation, to study seasonal to decadal changes, and to improve global ocean and coupled model initialization. Sea level anomaly is defined as the sea surface height above a mean sea surface (MSS) that is defined over many repeat tracks and independently for each satellite (https://www.aviso.altimetry.fr/en/data/products/auxiliary-products/mss.html). The absolute dynamic topography (ADT) is then the sea level anomaly added to the mean dynamic topography (MDT) which is calculated using a combination of gravity missions (GOCE, GRACE), all available altimetry, and in situ data over 1993-2012 (https://www.aviso.altimetry.fr/en/data/products/auxiliary-products/mdt/mdt-description.html). Since TOPEX/Poseidon (launched in 1992), a series of altimeters have continuously provided ADT observations with varying estimated accuracy of 4 cm [*Shum et al.*, 1995]. Typically the joint US/French series (TOPEX, Jason 1, 2, and 3) have repeat orbits designed to

	INST_ERROR			Data Coverage			
Instrument	**T (°C)**	**S (psu)**	**SLA (m)**	**Start**	**End**	**Availability/ Repeat**	**Source:**
TAO	0.02	0.1		1980	present	daily	PMEL (ftp://taopmelftp@ftp.pmel.noaa.gov/ascii/sites/daily)
PIRATA	0.02	0.1		1997	present	daily	"
RAMA	0.09	0.1		2005	present	daily	"
XBT	0.1	0.1		Prior to 1980	Present – ~1 month	Sparse – VOS*/SOO or research	UK Metoffice (https://www.metoffice.gov.uk/hadobs/en4/download-en4 -2-0.html)
CTD	0.003	0.02		Prior to 1980	Present – ~1 month	Sparse – research (high vertical resolution)	"
ARGO	0.005	0.02		1996	present	10 day cycle - lagrangian	USGODAE (ftp://www.usgodae.org/pub/outgoing/argo/geo/)
Topex			0.02	09/25/92	04/24/02	10day repeat	Copernicus (ftp.sltac.cls.fr/Core/SEALEVEL_GLO_PHY_L3_NRT_OBSERVATIONS_008_044)
ERS-1+2			0.02	10/23/92	10/08/02	35day repeat	"
Geosat FO			0.02	01/07/00	09/07/08	17 day repeat	"
Jason1 **Jason2** **Jason3**			0.02 0.02 0.02	04/24/02 10/19/08 08/22/16	10/19/08 09/12/16 present	10 day repeat	"
Envisat			0.02	10/26/10	04/08/12	35 day repeat	"
Cryosat-2			0.02	01/28/11	present	369 day repeat with 3, 29, 85 sub-repeat	"
Saral			0.02	03/14/13	present	35 day repeat	"
HY-2A			0.02	04/12/14	present	14 day repeat	"
Sentinel 3A			0.02	02/29/16	present	27 day repeat	"

Table 1. Table showing the estimated instrumental error, period and source for all data used in the ODAS. The altimeter products were produced by SSALTO/DUACS and distributed by AVISO, with support from CNES (http://www.aviso.altimetry.fr/duacs/).

measure the complete globe every 10 days. In contrast, the European satellites have 35-day exact repeat orbits (see Table 1 for details).

A unique aspect of this ODAS is that absolute dynamic topography (ADT) is assimilated into the S2S-v2.1 using the same schedule/technique as for the in situ data. Because of the sheer volume of the data, ADT data are thinned along track prior to assimilation. A Gaussian weighted mean is calculated for the central point of +/- 10 along-track observations using a decorrelation scale of 1000 km. This mean is then output for assimilation. The observational error for assimilation for this point is also calculated and output as the standard deviation using the weighted mean and the Gaussian-weighted +/- 10 surrounding along-track points. Within the ODAS, the ADT observational error is minimized to 0.1 m and an additional term is added to increase the ADT observational error away from the equator accounting for the Rossby deformation radius. This term linearly rises from 0.0 m at the equator to 0.1 m at 90° and the resulting observational error is then scaled by a factor which was tuned to produce optimal fields. Finally, the mean of all ADT observations for the assimilation period is removed prior to assimilation so that no mass is added to the solution. The horizontal localization is identical to that used for the profile data described above. Finally, if

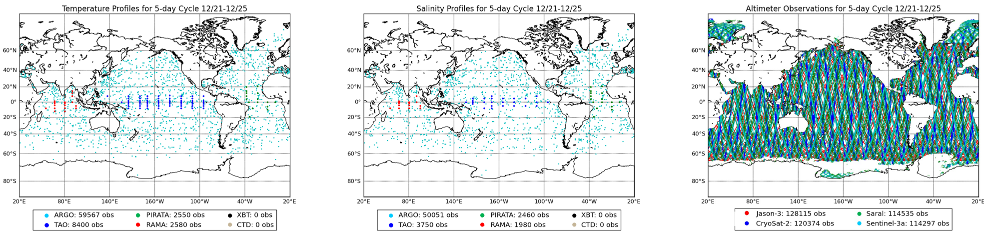


Figure 2. Data coverage for a typical 5-day assimilation (this case is for 5 day cycle centered on Dec. 23, 2017 extracted on Dec 29, 2017). Left panel shows all the profile data for temperature. Middle panel shows salinity profile data and right panel shows the data coverage for ADT data. For the profile data, the dot corresponds to the profile location and the number of observations in the key includes all model depths.

any ADT observation departure from mean background is greater than 1.0 m then the observation is removed from consideration.

The data coverage of ADT data for a typical 5-day operational ODAS is shown in Figure 2. The large volume of satellite ADT data is obvious even with subsampling the ADT data as described above. Relatively large gaps “diamonds” in the coverage occur 5° to 40° off the equator due to the various orbit orientations. Jason-3 data (red dots) cuts off at approximately 66° due to the high repeat orbital frequency (i.e. 10 days) and concomitant satellite orbit inclination. Other satellites with higher inclination orbits observe further north and south (e.g. CryoSat2 - 92°, Saral - 98.5° and Sentinel 3A - 98.6°inclination). Data gaps off Greenland, Labrador Sea, Hudson Bay, and Barents Sea in the north are due to ice coverage. In addition, the ice extent is clearly visible in the Southern Ocean around Antarctica. The valid data in the Ross Sea and eastern Weddell Sea is associated with open water of ice-sheet melting in southern summer. In any case, satellite derived ADT measurements provide a copious amount of observations for the ODAS.

Prior to assimilation into the ODAS, observations are preprocessed to perform quality control and to limit the data to be ingested. All profile data types are treated in the same way. First each profile is thinned to the model depth by simply averaging all the profile information within the model layer and these data are then assigned to the model layer depth. Next, the resulting data are checked against a climatology formulated using all available in situ observations from the World Ocean Atlas 2013 for temperature [*Locarnini et al.*, 2013] and for salinity [*Zweng et al.*, 2013]. Profile data at depth are flagged if the profile observation is more than 6 standard deviations away from the corresponding observation climatology. For the LETKF analysis, data are assigned observational error depending on the depth gradient of the observation. For the S2S-v2.1 ODAS, dT/dz and dS/dz are first required to be greater than or equal to 1.e^{-3} °C/m and psu/m, respectively. Then they are scaled by a factor, determined by a series sensitivity studies, to give the optimal profile observation error. In this way the observational error is greatest at the depth of maximum gradient such as within the thermocline. As a final large-scale error test, any observation with background departure greater than 10° C for temperature and 10 psu for salinity are tossed out.

The LETKF solves the analysis states in a local volume centered on each model grid point and is applied on a regular grid (0.5°x 0.5°x 40 levels). In the S2S-v2.1 formulation of the LETKF, vertical localization is turned off for profile data. This has the benefit of calculating the analysis only once (as opposed to 40 times) and unlike the previous ODAS system [*Vernieres et al.*, 2012] unique vertical localization profiles for each observation type are no longer applied. This technique has the additional benefit of allowing assimilation of vertical profiles and satellite altimetry data within a single process. The horizontal localization func-

OCEAN/ICE MODEL	MOM4.1 (Griffes et al., 2005), 0.5°, CICE 4.0 (Hunke and Lipscomb, 2008), 40 layers down to ~4500m	MOM5.0 (Griffes et al., 2012), 0.5° CICE 4.0, 40 layers down to ~4500m
ASSIM TECHNIQUE	SAFE/EnOI (Keppenne et al, 2014/ Oke et al., 2010, Wan et al. 2010), 5-day IAU	LETKF (Penny et al., 2013), 18 hour IAU
FORECAST ERROR	Pre-specified static forecast error cov. from leading EOFS of an ensemble of forecast anoms (wrt climate drift) from AOCGM	Estimate evolving errors using monthly averaged anoms from 20 freely coupled AOCGMs re-centered about analysis.
DATA	10-day window, binned to model level, reject if >6σ from WOA09, obs. error prescribed to observation and type, decorrelation scales x, y, z, t function of variable, unique vertical localization factors applied for each observation and type	5-day window, binned to model level, reject if >6σ from WOA13, obs. error from $\frac{dT,S}{dz}$ or $\frac{dADT}{d(\text{along track dist.})}$, horizontal localization applied $\propto$ Rossby deform radius, no vertical localization applied
ASSIM SEQUENCE (*SAFE, 'EnOI)	Tz, Sz Clim* SSS Clim* SST* Tz' Sz' AICE	Tz, Sz, ADT
RELAXATION VARS	None	SST (1day), AICE (insertion)
REPLAY FORCING	MERRA (Rienecker et al., 2011) corrected using GPCP (<9/09), CMAP (9/09-present)	"MERRA2-like" (FPIT) (Gelaro et al., 2017) corrected using CMAP precip.

Table 2. Table highlighting the differences between the setup for the S2S-v1 (middle column) versus S2S-v2.1 (right column).

tion is used to scale the observational errors such that observations nearer the central model point result in higher localization weighting function. In addition, the horizontal localization function accounts for the larger Rossby radius of deformation near the equator. This radius varies from 240 km near the equator to less than 10 km near the poles [*Chelton et al.*, 1998]. In practice, the horizontal localization function is parameterized as a Gaussian and as a function of latitude with 1 standard deviation of 3.6° at the equator and 1.8° at the poles. Thus, the impact degrades as the observation point is further from the central point and as the observation latitude increases.

There are many differences between the ocean analysis methods in the GEOS S2S-1 *Borovikov et al.* [2017] and GEOS S2S-2 systems. The major differences are summarized and highlighted in Table 2. For the current system, initial conditions and verification for the land and atmosphere are provided by the NASA FPIT reanalysis, an improved atmospheric forcing field over MERRA which was used to force GEOS S2S-1 (NEED REFERENCE FOR THIS). In GEOS S2S Version 2 the observations are incorporated into the ocean state using a 5-day assimilation cycle and the Local Ensemble Transform Kalman Filter (LETKF) using 20 ensemble members (*Vernieres et al.* [2012]). The advantage of this ensemble Kalman Filter over a less expensive deterministic filter such as the GEOS S2S-1 SAFE/EnOI (*Keppenne et al.* [2014], *Oke et al.* [2010]) techniques is that it allows the error covariances to evolve with the seasonal cycle and the phase of ENSO. Another clear advantage of the current system is that GEOS S2S-2 assimilates all available satellite altimetry whereas the previous system did not assimilate any sea level in the production system. In section 5 we will show some key metrics documenting improvements in the current (GEOS S2S-2) versus the old production system (GEOS S2S-1).

3 Experiments and Initialization Method

In this study, we examine the results of four sets of seasonal and subseasonal forecasts using the new GEOS-S2S_2.1 system, a long-term free running simulation with the GEOS S2S-2 AOGCM, and the AO-CDAS itself. The four sets of forecasts are: 1) near real-time seasonal forecasts, 2) retrospective seasonal forecasts, 3) near real-time sub-seasonal forecasts, and 4) retrospective subseasonal forecasts. The long-term simulation is a 50-year long atmosphere-ocean coupled climate simulation with GEOS-S2S_2.1, which is a "perpetual 2000" simulation, where the external climate forcing is fixed at that of year 2000. GEOS-S2S_2.1 is also utilized in producing the ocean analysis dataset (2012-present), that is used to initialize the seasonal and subseasonal near real-time forecasts.

Here we provide a description of the experimental set-up of the seasonal retrospective forecasts and forecasts. Historically, owing to the availability of the GMAO ocean data assimilation products, the GMAO forecasts were initialized on a fixed set of calendar dates. These begin on Jan 1, and are phased 5 days apart, thus producing a total of 72 start dates per year (Fig 3). In the new system, we follow the same start date calendar, although seasonal forecasts and reforecasts are limited to the last 4 start dates of the month (green and orange boxes in Fig 3). For the retrospective forecasts, no perturbations on initial conditions are employed, rather the lagged start dates form an ensemble of 4 for any month. A set of seasonal reforecasts are conducted for a period of 36 years (1981-2016). A suite of 5-day long ocean data assimilation runs were produced for 1981-2016, initialized at 5 day intervals, in order to generate initial conditions for the retrospective seasonal and subseasonal forecasts.

For near real-time seasonal and subseasonal forecasts, the atmosphere is initialized from GMAO's real-time forward processing for instrument teams (FPIT) analysis (*Lucchessi et al.* [2016]), and the ocean and sea ice initial conditions are taken directly from a continuous GEOS-S2S-2_1 AODAS run initialized in 2012 with MERRA-Ocean fields. *In the land fields, observed precipitation values are incorporated by ....* As compared to the retrospective forecasts, for the near-real-time suite an additional 6 ensemble members are generated around the 4th member (last start date in each month) (orange box in Fig 3), thus producing a total of 10 ensemble members (4 unperturbed and 6 perturbed). The method to perturb initial conditions is based on the difference between two analysis states 5 days apart. The perturbations are re-scaled and the magnitude of the norm reduced to approximately 10 percent of the natural variability of SST over the norm region (i.e. 0.48C); the region for defining the norm is tropical Pacific domain over 120W-90W, and 10S-10N. The rescaling norm is the RMS difference of the instantaneous sea surface temperatures (SSTs) from two analyses 5 days apart. The variables perturbed are on ocean model grid - 3D temperature, salinity and ocean velocities, surface temperature, sea level and frazil, ice velocity and strain components, ice strength, extent U mask and stress tensor components; on atmospheric grid - wind components, potential temperature and specific humidity; on tiles - skin temperature, salinity and depth. Both the near real-time forecasts and retrospective forecasts are run for a period of 9 months.

The subseasonal retrospective forecasts are performed for 17 years (1999-2015). The reforecasts are initialized every 5 days (pink, green and orange boxes in Fig 3) with a total of 73 start dates per year, with 4 ensemble members per start date. Here, one ensemble member consists of unperturbed initial conditions and the remaining three are generated by perturbing the atmospheric initial conditions in horizontal winds, potential temperature and specific humidity. The perturbations are computed as scaled differences (+/- (x-y)/8) between two arbitrary atmospheric states (x, y) taken 1 day apart. The somewhat arbitrary scaling factor of 1/8 is chosen in an attempt to produce perturbations that are of size consistent with initial errors typically found in numerical weather prediction models. The initial conditions, as for the seasonal forecasts, are based on a series of 5-day long ocean data assimilation runs. The near real-time subseasonal forecasts are conducted in a similar manner to the retrospective forecasts, with initialization in every 5 days, and 4 ensemble members around each start date. Both the near real-time forecasts and retrospective forecasts are run for a period of 45 days.

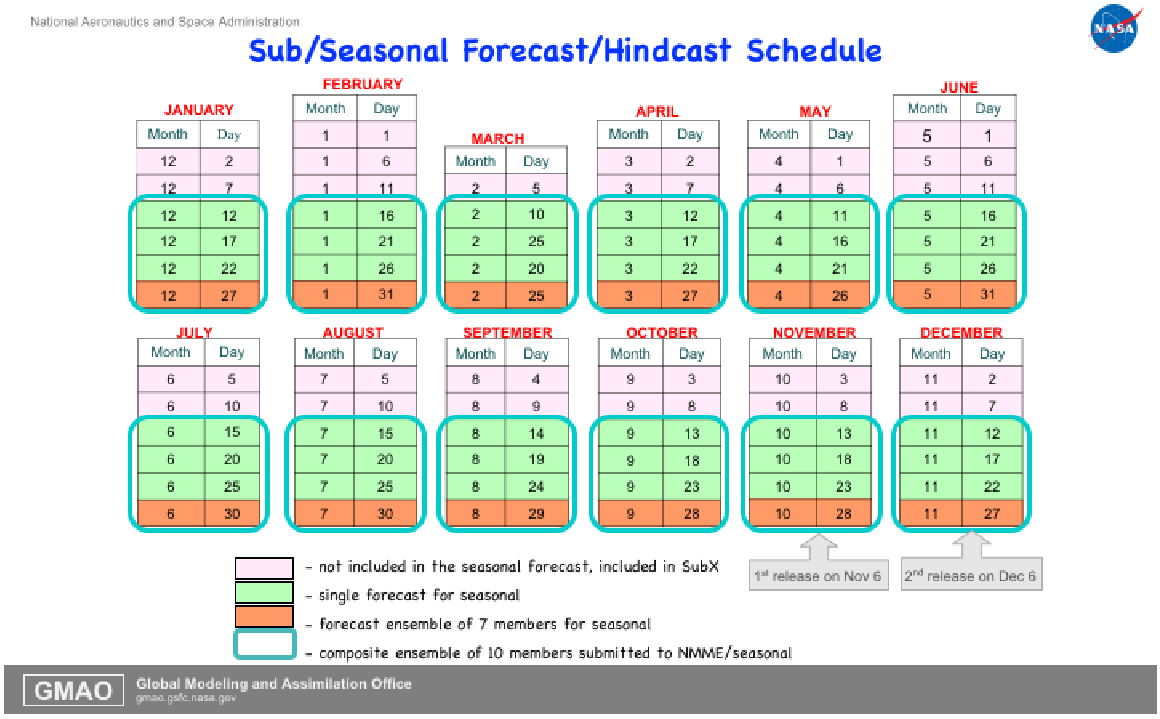


Figure 3. Schedule of Forecasts and Hindcasts

4 Climate of Coupled Model

4.1 Mean Atmospheric Climate

The examination of the results of simulations with GEOS S2S-2 begins with the assessment of the 50-year mean climate and a comparison against the previous version of the model (GEOS S2S-1). This is the climate state towards which the shorter term simulations are drifting at different rates, and as such its assessment is critical to establishing the fidelity of the model. In this section we focus on the quality of the atmospheric component of the GEOS S2S-2 model's long-term climate. Results are presented to assess the quality of the circulation, hydrological cycle and radiative balance, as reflected in the December-January-February (DJF) and June-July-August (JJA) mean circulation, stationary waves, mean wind, and atmospheric transients, as well as precipitation, net radiation and surface air temperature.

Figure 10 shows....

Figure 6 shows that the mean error in the stationary waves (as shown by the 300 hPa eddy heights) is substantially reduced in the new model relative to the old model for both seasons and both hemispheres. In particular, the weak boreal winter stationary waves in the old model (especially over the North Pacific/North American region, Fig. 6a), are now stronger and closer to the observed as seen from the smaller biases (Fig. 6c). During JJA the new model also has reduced biases particularly over Eurasia (Figures 6b and 6d).

The simulated 200 mb zonal winds (Figure 7) show some improvements and some degradations in GEOS S2S Version 2 relative to Version 1. While the middle and high latitudes generally show smaller errors for Version 2, reflecting improvements in the middle latitude jets, there is a subtropical westerly wind bias in the new model in both hemispheres and during both seasons, and especially so for boreal summer. Despite these subtropical zonal wind biases, there are substantial improvements in the middle and high latitude transients (shown in Fig. 8) that we argue are tied to the above-noted improvements in the middle and high latitude zonal winds. In particular, the old GEOS S2S-1 model had weak boreal summer middle latitude storm tracks (see the negative bias in Fig. 8b) linked to the weak North Pacific and North Atlantic jets (Fig. 7b), that are now much improved. The new model does show somewhat excessive wintertime transients in the upper troposphere subtropics, associated with the excessive subtropical westerlies (Figs. 8c and 8d). Figure 8 also shows improvements in the stationary wave variances, reflecting the improvements to the stationary waves shown previously (Fig. 6). The new model also has improved lower tropospheric northward moisture transport (Figure 9) during both seasons and in both hemispheres. During DJF the excessive northward transport in the southern hemisphere (Fig. 9c) is almost completely eliminated in the new model. The negative biases in the North Pacific and North Atlantic are also eliminated but with some positive biases now occurring in the subtropics. During JJA the moisture fluxes are much improved throughout much of the globe, with the new model showing only weak negative biases in most regions, except for a few isolated regions such as over the North Pacific just off the east coast of Asia.

We next examine the biases in precipitation and surface air temperature, that are two quantities that have considerable practical importance since they are the quantities for which society would likely benefit most from improvements in skill. The precipitation biases (Figure 11) show a mix of both improvements and some degradation. During DJF the new model shows generally reduced biases with the main exception being the large positive bias just north of the equator in the Pacific (Figure 11c). There is also little improvement over the Indian Ocean. There is much less improvement during JJA, especially in the tropics with a much increased positive bias in the Atlantic and a much more pronounced split ITCZ in the Pacific (Figure 11d). There is also much less rain over India, reflecting a lack of summer monsoon rainfall in the new model. The main improvements include the elimination of the excessive precipitation over Tibet, and much improved precipitation over the NH storm tracks, presumably reflecting the improved summer transients mentioned earlier (Figures 11b and d).

Long term mean precipitation patterns in GEOS S2S-2_1 show some improvement and some degradation relative to GEOS S2S-1_0. Figure 11 shows the new and old AOGCM simulation precipitation relative to estimates from the Global Precipitation Climatology Project (GPCP, *Huffman et al.* [2009]). The excessive precipitation over high topography seen in the results from the old system in Figure 11d over the Andes, the Maritime Continent and the Tibetan Plateau is still present in Figure 11c, but is reduced by more than a factor of 2. In the tropical Pacific, however, the unrealistic precipitation minimum along the equator is exacerbated in the results using the new GEOS-S2S-2_1 AOGCM.

Before turning to the surface temperature evaluation, the net radiation at the surface is examined, as errors in the net radiation are a dominant source of error in surface temperature. The earth is assumed to be in approximate radiative balance, meaning that the globally averaged temperature is nearly constant during the year. The global net radiation at the top of the atmosphere, therefore, which represents the balance between the net incoming solar and net outgoing terrestrial radiation, is close to zero. In the GEOS S2S-2 simulation the balance is held to within approximately 1 W m^{-2}. The horizontal distribution of the different components of the radiation budget have important implications for the atmospheric general circulation patterns.

The distribution of the net TOA radiation, in particular the latitudinal gradients, is associated with the net driving energy for the atmospheric and oceanic general circulations. The zonal mean cross sections, figure 4, show a pattern that follows the net shortwave, which is a positive maximum in the summertime subtropics, decreasing towards the summer pole and towards the winter hemisphere, crossing zero in the winter subtropics, descending to a negative minimum at the dark winter polar latitudes. The strongest north-south gradients, which imply the strongest north-south heat transports, are in the winter hemisphere. The figure shows that GEOS S2S-2 simulation matches the general distribution of the net radiation quite well. The errors in the shortwave and the errors in the longwave compensate and the resulting net radiation matches the CERES estimate.

Surface radiative balance is shown in Figure 5....

Finally we turn to the surface air temperature over land (Figure ???). The new model shows overall much improved (reduced) biases in both seasons. The strong negative biases during DJF in NH high latitudes in the old model are for the most eliminated (cf. Figs 6a and c). The new model does show some tendency for increased positive biases during DJF especially in the SH. During JJA, the new model has much reduced positive biases over northern Eurasia and reduced negative biases over Greenland, with however some increase in the positive biases in the tropics and SH as well as over the Indian subcontinent (cf. Figs 6b and d).

In summary, the biases in the new model are overall much improved throughout the middle and high latitudes, both for the dynamical quantities considered here and the precipitation, net radiation and surface temperature. The main problems in the new model appear to be linked to excessive tropical precipitation, and that includes the excessive subtropical westerlies and associated exceptionally warm tropical upper tropospheric temperatures (not shown). It is noteworthy that the tropical precipitation biases (and weak precipitation over India) are much reduced if not absent from the new model when run uncoupled (forced with observed SST), indicating these are errors associated with coupled processes. As such, we must also look at the modelâĂŹs ocean climate including the surface fluxes to help us understand the nature of these biases. This is the subject of the next subsection.

4.2 Mean Ocean Climate

The long term mean climate of the AOGCM is examined in comparison to other observationally based estimates of various fields. Figure 13 shows a comparison of the old and new S2S AOGCM long term mean sea surface temperature (SST), using the Reynolds SST analysis for reference. The large region of saturated purple shading in Figure 13d near Greenland and the Labrador Sea are absent in the S2S-2_1 result seen in Figure 13c. In addition, other regions of cold bias greater than 2 ° C, in the northern Pacific and Atlantic and southern Pacific oceans are also absent in the simulation with the new system. Warm bias on the order of 2 ° C close to the west coast of Africa and South America seen in 13d are also removed in the new system (13d). Most of the changes in SST represent clear improvements in GEOS S2S-2 relative to GEOS S2S-1. NEED TO EXPLAIN WHY

need more stuff here

Figure 14 shows.....

Figure 15 shows......

Figure 16 shows.....

4.3 Climate Variability

A sub/seasonal forecast systemâĂŹs success is often viewed in terms of its ability to predict the evolution of climatic modes such as El Nino and the Pacific Decadal Oscillation (PDO) âĂŞ components of the Earth system that have significant intrinsic predictability at seasonal time scales. Much of the societal value of a forecast, however, is arguably tied to its performance in continental population centers, which can be quite distant from the SST patterns representing such modes. A desirable characteristic for a forecast system is indeed the ability to capture realistically any teleconnections that might exist between the predictable modes over the ocean and remote continental meteorology.

This is evaluated in Figure 17, using an approach described by Collow et al. (2017). Figure 17a shows the relationship between an observed index of the PDO during DJF and the observed DJF near-surface air temperature (T2M). The PDO indices were extracted from the Physical Science Division of NOAAâĂŹs Earth Science Research Laboratory (ESRL, https://www.esrl.noaa.gov/psd/data/climateindices/list), while the observed air temperatures were estimated with MERRA-2. The observations indicate a swath of negative correlation from the northeastern US into eastern Mexico. When the same calculation is performed with output from a long-term, free-running coupled simulation using the model underlying the old forecast system (with the PDO index computed from the simulated SSTs using ESRLâĂŹs method), negative correlations are again seen in the eastern US, but unrealistic values appear in the center of the country, and underestimated values are seen along the eastern coast (Figure 17b). In a parallel coupled simulation using the model underlying the new system, the spatial pattern of the negative correlations is generally improved, though the magnitudes of the correlations along the east coast are still too weak.

Figures 17d-f show corresponding results for the Nino3.4 index. The observed teleconnection pattern (Figure 17d; the observed Nino3.4 was also extracted from the ESRL website) shows positive correlations in the northern US and southern Canada and negative correlations in Mexico and Texas. The simulated teleconnections in both the old system (Figure 17e) and the new system (Figure 17f) clearly overestimate the spatial extent of the negative correlations and miss the positive ones. Results for precipitation, however, are more encouraging. Gauge-based precipitation observations (as processed in MERRA-2; see Reichle et al. (2017)) show a positive correlation with Nino3.4 index in the southern tier of the US and patches of negative correlation along the northern border (Figure 17g). Both the old (Figure 17h) and new (Figure 17i) systems capture some of the eastern segment of the positive correlations, and the new system appears to place the negative patches to the north in roughly the right place.

5 Ocean Data Assimilation Results - Comparison of S2S-1 and S2S-2

In this section we will evaluate the S2S-1 and the S2S-2 ocean data assimilation system (ODAS) results. The quality of the ODAS has direct implications for the fidelity of the forecasts that are initialized with the ODAS fields. MORE HERE??

We assess improvements in our current coupled system by comparing a sampling of ocean metrics against those from the previous system. For sea level and geostrophic currents, the two sets of reanalyses are compared against the multi-gridded altimetry product of *AVISO* [2013]. S2S-2 has an overall reduced bias with respect to observations for nearly all regions. This is not surprising considering the fact that the S2S-2 ODAS assimilates sea level whereas the S2S-1 ODAS did not. The improvement with S2S-2 is most dramatic in the western boundary current region of the Kuroshio, Gulf Stream and Brazil Currents. In these regions, assimilation of sea level data is critical for improving the character and location of these turbulent regions. As an example, Figure 18 shows the zonal current speed for August 2017. The bottom right figure shows the observations from the multi-product AVISO sea level. The prominent features of the Loop Current entering the Gulf of Mexico and the eddies and meanders of the Gulf Stream off the east coast of North America are clearly evident in observations. On the other hand, the S2S-v1.0 (top right) barely shows the Loop Current and the Gulf Stream has no eddies nor meanders and it exits the coast too far north as compared to observations. The S2S-v2.1 better reproduces the amplitude and location of the observations (top right). The S2S-v2.1 has a much more realistic pattern of eddies and meanders in the Gulf Stream and the amplitude and Loop current is much better reproduced in the S2S-v2.1 as opposed to S2S-v1.0. In addition, the Gulf Steam exit into the North Atlantic is much better modeled in the S2S-v2.1 experiment. Although the exact locations of the meanders and eddies for the S2S-v2.1 don't exactly match the observed locations, the general character of the S2S-v2.1 experiment better represents the transport of observations. This improvement in the western boundary current is likely due to the combination of the improved forcing of MERRA2 versus MERRA and the assimilation of satellite altimetry for S2S-v2.1.

The location and amplitude of the western boundary currents have important consequences for climate as measured by global heat conveyor belt indices. One such index is the RAPID array along 26.5°N that measures the Atlantic Meridional Overturning Circulation (AMOC). Moorings stretching across the Atlantic at 26.5°N measure temperature, salinity and currents and can thus measure the transport of warm water northward (via the Gulf Stream) and cool water southward (via the North Atlantic deep circulation). Figure 19 top shows that the AMOC for the S2S-v1.0 is consistently weaker than observed. On the other hand, the S2S-v2.1 is initially too strong compared to observed values. However, as the experiment spins down, the S2S-v2.1 reanalysis settles to match the magnitude of observations. In addition, the interannual variability of the S2S-v2.1 looks more realistic with respect to observed values as compared to the S2S-v1.0.

Another major region of western boundary current transport of the global heat conveyer belt is the Indonesian Throughflow (ITF). Warm, fresh water is transported through the ITF due to consistent pressure head from the Pacific to the Indian Ocean. Eleven moorings were deployed across the entrance (Makassar Strait, Lifamatola Passage but not Halmahera) and exit regions (Lombok, Ombai, and Timor) of the ITF from 2004-2006 and are dispersed to accurately measure each passage's contribution to the ITF (*Sprintall et al.* [2009]). In addition, various studies attempted to directly measure the flow of the ITF and estimate the interannual variability. For example, *Meyers et al.* [1995] measured the mean ITF using the geostrophic transport calculated from the IX01 WOCE XBT data (Fremantle-Sunda Straits). Here we calculate an index of the ITF using our gridded optimal interpolation of observed temperature and salinity (*Carton* [1989]), convert temperature and salinity to dynamic height, and then calculate geostrophic currents. The transport is then estimated the across 114°E between 21°S and 9°S (closely matching the IX01 line). The good correspondence between the INSTANT measurements (red dash line in Figure 19 bottom) and our ITF estimates (red solid line) demonstrates the fidelity of this technique. Figure 19 bottom shows that the S2S-v1.0 reanalysis badly underestimates the transport of the ITF. The mean for S2S-v1.0 is about -5 Sv whereas observed values are estimated at -15 Sv (INSTANT) and -14 Sv by *Wijffels et al.* [2008] using QuikScat winds and the Island Rule of *Godfrey* [1989]. On the other hand, the S2S-v2.1 closely matches the mean, seasonal cycle, and the interannual variability of the observations. The ITF transport of the S2S-v2.1 clearly outperforms the S2S-v1.0 values.

Finally we assess the differences between S2S.v1.0 and S2S-v2.1 for the large-scale oceanic Kelvin and Rossby waves for the equatorial Pacific. These waves are instrumental for proper attribution of the buildup and recharge stages of ENSO, respectively (e.g. *Jin* [1997]). In Figure 20, the model and observed sea level data sets are first converted to geostrophic currents (*Picaut and Tournier* [1991]) then the Kelvin and Rossby amplitudes are calculated using the technique of *Decroix et al.* [1994]. The top left panel shows the observed west-to-east propagating Kelvin wave signal for 2013-2015. Early in the time series, negative (blue) Kelvin waves represent the upwelling associated with the weak 2013 La Niña. The downwelling (red) signals of the 2015 El Niño are evident starting from January 2015 and each successive Kelvin wave increases in magnitude as the Bjerknes feedback becomes enhanced throughout the buildup of this big event (see e.g. *Santoso et al.* [2015] for details of this event). The amplitude and timing of these Kelvin waves are well reproduced by the S2S-v2.1 experiment (top middle panel). On the other hand, the S2S-v1.0 shows overall weaker Kelvin wave amplitude throughout the period (right top panel). For example, the big Kelvin wave in summer 2015 is roughly 30% smaller for the S2S-v1.0 than for S2S-v2.1 example. For the Rossby waves (Figure 20 bottom), the lack of amplitude for the S2S-v1.0 is even more evident. The big upwelling (downwelling) Rossby wave in early 2013 (2015) is accurately reproduced by the S2S-v2.1 system (middle bottom) whereas the S2S-v1.0 badly underestimates these signals. For example, the upwelling in spring of 2013 reaches -0.3 m/s for observations and S2S-v2.1 but the S2S-v1.0 only peaks at -0.15 m/s.

In summary, almost all ocean variables examined were improved for the S2S-v2.1 relative to S2S-v1.0. SST and SSS biases were reduced (especially off the equator and in the North Atlantic, respectively) but SSS was somewhat degraded over Indonesia and the Amazon plume. Assimilation of SL in S2S-v2.1 improves western boundary currents. For example, the amplitude, location, and character of Loop Current and the Gulf Stream and both were more realistic with respect to S2S-v1.0. The large scale meridional (AMOC) and zonal (ITF) heat transport indices were significantly improved in S2S-v2.1. The amplitude of the Large-scale Kelvin and Rossby waves were simulated well with S2S-v2.1 whereas S2S-v1.0 badly underestimated the El Niño and La Niña forcing. Improvements are most likely due to sea level assimilation and better forcing (MERRA2 versus MERRA) for S2S-v2.1 as compared to S2S-v1.0.

6 Seasonal and Subseasonal Forecast Assessment

Results of retrospective seasonal forecasts....

While forecast assessment should in general include both a deterministic and probabilistic evaluation, the relatively small ensemble size of our seasonal retrospective forecasts limits what we can do to assess the quality of the ensemble (e.g., reliability diagram, ROC, Brier skill score). As such, we focus here primarily on deterministic measures involving the ensemble mean such as, anomaly correlation, rms, and phenomena-based compositing, though we do provide an initial assessment of the ensemble spread compared to that of the previous system. We begin with a look at the climate drift, keeping in mind that the forecast skill evaluation is done for the anomalies (after removing the climate drift).

6.1 Climate Drift

Climate drift (SST, T2m over land, precipitation) – time series of AMOC

Taken from Anna's paper, needs to be rephrased.

Forecast drift is an artifact of the imperfect models. For the seasonal forecast it is necessary to properly account for the drift and calibrate the forecast accordingly. A complete set of retrospective forecasts for the entire training period are required to consistently de-trend the forecast. Following the convention established by Stockdale (1997) and others the the drift is calculated as the average of these hindcasts from 1981-2010. It is subsequently subtracted from the production forecasts. For the S2S-2.1 a single hindcast was computed on each date, while for the S2S-1.0 multiple hindcasts with perturbed initial conditions we run on the dates corresponding to the forecasts with multiple ensemble members; the same perturbations techniques were used for hindcasts as for the forecasts. For the comparison of the forecast drifts and evaluation of forecast skill only the individual hindcasts done on the same dates (4 per month) were used from both systems.

Figures 21 and 22 show the SST seasonal mean error at 1, 3 and 6 month leads. From the west to east across the equatorial Pacific, the amplitude of the Niño 4 bias is reduced in S2S_2.1 for all initialization times for all leads; in Niño 3.4 the late fall and early winter bias is smaller in S2S_2.1 for all leads, and is generally within 0.5°C, while in S2S_1.0 the bias reached -2°C in DJF, in spring and summer the bias is similar in both systems. In Niño3 region the cold bias in S2S_2.1 lingers from February through early summer, while in S2S_1.0 it is the largest in March, in in summer the bias is near 0°C. The greatest difference between the two systems occurs in the eastern equatorial Pacific Ocean, apparent in Ninño1, Ninño2 and the combined Niño1+2 regions SST indices. Figure 23 shows......

Talk about different AGCM clouds here.

6.2 Forecast skill at subseasonal time scales: Tropical intraseasonal variability

Figure 24 shows bivariate correlation of Real-time Multivariate MJO (Real-time multivariate MJO (RMM)) indices between reforecasts (1999-2016) and MERRA-2 as a function of the forecast lead. The RMM indices are derived from the zonal wind at 850hPa and 200hPa and outgoing longwave radiation following the method addressed in [*Wheeler and Hendon*, 2004] and [*Gottschalck et al.*, 2010]. The result presents that, at short lead (say 1-10 day lead), correlation is greater than 0.85 at all initial condition month. At inter-mediate lead, correlations are generally greater than 0.75, 0.65, and 0.55 at 15, 20 and 25 day lead, respectively. Prediction skill at long lead is particularly higher in summer from June through September, exceeding correlations of 0.5 even at 40 day lead. We compare this prediction skill with those identified from the other prediction models. A number of different models have had a difficulty in forecasting the MJO with correlation reaching 0.5 beyond 25 day lead (e.g., Fig. 1 in [*Saha et al.*, 2014] and Fig. 2 in [*Lim et al.*, 2018]). Our model for the MJO prediction clearly shows that correlations remain to be greater than 0.5 at 25 lead day and especially higher in summer at longer lead, quite comparable to the prediction skill of the other models.

6.3 Forecast skill at seasonal time scales

(anomaly correlation, rms, composites) - MAKE SURE FAIR COMP TO OLD SYSTEM

6.3.1 Skill of Global SST, Nino 3.4

ACC computation is based on the 4 hindcasts from 1982 through 2010 started on the same dates for either seasonal forecast system.

Rank histogram is a tool to assess the consistency of the forecast system, to check whether the observations statistically belong to the distribution of the forecast ensembles. Given the small number of forecasts per month, we consider seasonal samples, i.e. we combine 3 months of 4 lagged ensemble members, initialized on the same dates in both forecast systems (S2S_1 and S2S_2-1) during the same month to a total of 105 forecasts for DJF, MAM, JJA and SON seasons. In an ideal situation, when the distributions of observations and the forecasts coincide, the rank histogram would be close to a uniform, flat shape. A skewed, “L“ shape, of the rank histogram in indicative of a biased forecast ensemble, and a “U“ shape is telling that observations tend to fall outside the ensemble envelope, i.e. the forecast ensemble does not have enough spread. Overall across the equatorial Pacific ocean neither system has sufficient spread, measured by the mean deviation from the ideal uniform histogram (marked by the red horizontal line). Larger values of this measure mean worse ensemble spread. The difference between the S2S_1 and S2S_2-1 is most dramatic in DJF forecasts (shown in figure 27) in the equatorial Eastern Pacific. In Nino1+2 region the difference between the two systems is remarkable: S2S_1-0 is biased warm at all leads (1 and 6 are shown here) in all seasons except JJA (not shown here), while the histogram of the S2S_2-1 ensemble appears much closer to the desired uniform shape, especially at lead 6 (initialized in JJA), with the mean deviation from the uniform is only 0.8. Rank histograms show both systems biased warm in Nino 3.4 regions at lead 1 (initialized in NDJ), but by lead 6 (initialized in JJA) S2S_2-1 warm bias is reduced, and the ensemble is slightly biased cold.

6.3.2 Skill of T2m and Precip

6.3.3 Teleconnections and Low Frequency Mode Prediction

here we can look at 1) how well we reproduce these modes, 2) how well we can forecast them (say by looking at some index), and also 3) how well we reproduce their impacts on say T2m and precipitation - by compositing on those indices)

We try to capture the major teleconnection patterns for boreal winter using the 250mb geopotential height (1981-2016) by applying the Rotated Empirical Orthogonal Function (REOF) analysis technique. We first capture the major teleconnection patterns from the MERRA-2 data as a reference. As shown on the right panel in Figure 30, the North Atlantic Oscillation (NAO), Northern Annular Mode/Arctic Oscillation (NAM/AO), and the Pacific North American (PNA) patterns are identified as the first leading teleconnections over the Northern Hemisphere. The same calculation is then applied to the GEOS-S2S_2.1 hindcast data (one month lead) to assess the capability of the GEOS-S2S_2.1 for producing those teleconnection patterns. Comparison of the teleconnections indicates that the GEOS-S2S_2.1 (left panel) successfully captures the spatial structure of the major teleconnections over the Northern Hemisphere. Geographical locations of the positive/negative anomalies seen in the MERRA-2 are quite realistically reproduced in the GEOS-S2S_2.1 hindcast, though some underestimation of the observed magnitude of the anomalies is found.

We next assess how reliably the GEOS-S2S_2.1 can predict the phase/intensity of the leading teleconnections in winter. Time series in Figure 31 shows the interannual variation of the January/February averaged teleconnection indices (initialized on Dec. 27) computed from the GEOS-S2S_2.1 (blue), the old forecasting model (red), and the MERRA-2 (black), respectively. The teleconnection indices are computed by projection of the anomalous 250mb geopotential height over the Northern Hemisphere onto the spatial REOFs of the teleconnections. Comparison in the indices demonstrates that the GEOS-S2S_2.1 has improved the forecast skill by achieving anomaly correlations greater than 0.5 for all three teleconnections. A little decrease in correlation is found, however, when looking at December/January averaged teleconnections (initialized on Nov. 27). But they are still in an encouraging skill level, with correlations greater than 0.4 (Figure not shown).

6.3.4 TC activity

Predictive skill of seasonal tropical cyclone (TC) activity from the GEOS-S2S_2.1 is assessed in terms of the Genesis Potential Index (GPI) [*Emanuel and Nolan*, 2004]. The GPI is generally larger over the period, when the TC activity (e.g., counts and intensity) is stronger than usual, while the GPI tends to be smaller during the weak TC season. We compute the GPI over the North Atlantic and the Western Pacific region, respectively, from the GEOS-S2S_2.1 hindcast data and then compare the interannual variation of the GPIs with that computed from the MERRA-2. Time series over the period 1982-2016 in Figure 32 show the GPI each year averaged from June through September (initialized on May 31). We adequately assess, based on this figure, the ability of the GEOS-S2S_2.1 in anticipating the TC activity up to the next four months at the beginning of the TC season. It is clear that both GEOS-S2S_2.1 and old forecasting models are capable of predicting reliably the anomalous (above or below average) TC activity for the first four months (JJAS). Comparison in correlations indicates relatively better performance by the GEOS-S2S_2.1. Correlations are 0.55 for the North Atlantic and even up to 0.82 for the Western Pacific basin, while those two values are 0.52 and 0.68, respectively, from the old forecasting system.

6.3.5 Cryosphere

The presence and character of sea ice critically alters the local energy and moisture exchange between the ocean and atmosphere, affecting the local Arctic and Antarctic climate. In the Arctic, the accelerated reduction of sea ice cover in recent years is also associated with a regional amplification in near-surface air temperatures [*Screen and Simmonds*, 2010; *Serreze and Barry*, 2011]. These effects may also influence the larger-scale general circulation [*Alexander et al.*, 2004; *Deser et al.*, 2010; *Screen et al.*, 2018]. Appropriate treatment of sea ice characteristics in seasonal forecasting models may then influence Northern Hemisphere predictive skill [*Jung et al.*, 2014]; however, there is uncertainty in the causal relationship between Arctic sea ice conditions and midlatitude weather variability [*Overland et al.*, 2015], in part due to the limited the atmospheric response to sea ice variability in climate models

[*Screen et al.*, 2018]. Nevertheless, local improvements in sea ice forecasts provide useful information for Arctic stakeholders [*Ban et al.*, 2016].

The GEOS S2S Version 1 forecasting system demonstrated reasonable predictive skill of hemispheric sea ice cover, with June forecasts explaining approximately 50 percent of the observed variance in the September Arctic ice extent (Figure 38a). Forecasts of the minimum sea ice extent also fared well when compared with other dynamical models in the Sea Ice Outlook [*Borovikov et al.*, 2017]. In producing GEOS-S2S Version 2, a key goal was to assess an upper limit on predictive skill with the current model configuration. To this end, forecasts for the retrospective period were initialized with ice thickness values from a validated modeling system (GIOMAS: Global Ice-Ocean Modeling and Assimilation System; *Schweiger et al.* [2011]). The use of GIOMAS sea ice thicknesses resulted in substantial improvement in forecast skill at longer lead time. In the Arctic, the system also benefitted from the use of the GEOS Forward Processing for Instrument Teams analysis (FP-IT), a near-real time derivative of the MERRA-2 reanalysis. The FP-IT incorporates a seasonally-varying sea ice albedo for improved air temperatures in the atmospheric forcing [*Gelaro et al.*, 2017]. The use of GIOMAS and FP-IT atmospheric forcing effectively eliminated a large springtime negative sea ice extent bias found in earlier versions of the seasonal prediction system (Figure 33a).

A credible initial ice thickness field has been widely demonstrated to improve the seasonal forecast skill for sea ice extent (e.g., *Blanchard-Wrigglesworth et al.* [2017]; *Chevallier and Salas-MÃľlia* [2011]; *Day et al.* [2014]). Retrospective forecasts of the GEOS-S2S Version 2 system using GIOMAS explain between 70 and 82 percent of the sea ice variability, a substantial improvement over the previous system (Figure 33). However, much of the skill is derived from predicting the long-term decreasing trend in sea ice extent. The model's skill is substantially reduced when the forecast extent is linearly detrended (after *Bushuk et al.* [2012]). Over the retrospective period, the GEOS S2S Version 2 July forecast explains approximately 30 percent of the detrended sea ice extent variability. This reduction in the detrended forecast skill arises from difficulties in predicting anomalously high or low sea ice extents (i.e., the anomalous summer extents of 1996, 2007, 2012, etc.), and is commonly found in dynamical sea ice forecasting models (e.g., *Hamilton and Stroeve* [2016]). The results nevertheless highlight the importance of deriving an accurate, historical record of ice thickness and methods for incorporating near real-time ice thickness observations in future seasonal forecasting systems.

Changes in glacier and ice sheet surface mass balance (SMB: here, the net of precipitation minus evaporation/sublimation and runoff) may alter the climate on seasonal timescales via local changes to surface energy budget characteristics (i.e. ice surface albedo; [*Box et al.*, 2012] and through the selective discharge of freshwater, which may impact local fjord circulation as well as ocean stratification [*Mortensen et al.*, 2013; *Sciascia et al.*, 2013]. Change in runoff may also impact the pattern and timing of ocean nutrient delivery and dependent phytoplankton production, particularly in the Northern Hemisphere [*Bhatia et al.*, 2013; *Sommaruga*, 2015]. Therefore, the appropriate representation of glacier and ice sheet SMB processes is an important step in improving the complexity of seasonal prediction systems and may provide valuable information to a range of stake holders.

The GEOS S2S Version 2 system incorporates the same snow and ice scheme as used in MERRA-2 [*Cullather et al.*, 2014]. Snow cover is explicitly represented with a modified version of the Stieglitz snow model [*Lynch-Stieglitz*, 1994; *Stieglitz et al.*, 2001], which caps snow depth at 15m and snow density at 500 kg m^{-3}. Snow cover is permitted to be fractional. The underlying ice column is composed of 15 layers, for an adequate representation of surface heat conduction, with a lower boundary condition of zero heat flux. Meltwater runoff may occur both from the snow column and directly from the ice surface. This produces a reasonable representation of SMB for the Greenland Ice Sheet (GrIS) when compared to both in situ measurements and high-resolution regional climate models [*Cullather et al.*, 2014], although the reduced spatial resolution may limit the ability to appropriately represent the surface melt spatial extent and gradients within the ablation zone. Nevertheless, the results compare well with localized observations [*Smith et al.*, 2017].

The current forecasting system reasonably reproduces the spatial pattern of mean SMB during the retrospective forecast period at 1-month lead times (1981-2016; Figure 34). Regions of high snow accumulation in southeastern GrIS are clearly present; however, accumulation is a moderately under-predicted. This occurs primarily due to an under prediction of snowfall during winter months and April. In addition, SMB within the ablation zones of the western GrIS, Iceland, and the northeastern Canadian Arctic are generally over predicted at 1-month lead times, reducing their spatial extent. For the GrIS and nearby areas, this over-prediction of SMB is primarily driven by an under-prediction of summertime ice sheet runoff during the latter part of the retrospective forecast period (Figure 34c). The conditions associated with the more recent retrospective forecasts are generally associated with a strong negative NAO and a strong positive EA during summer months, leading to high pressure blocking and warmer air temperatures, particularly over the western GrIS [*Lim et al.*, 2016], which do not develop as strongly as observed in the GEOS S2S Version 2 system. This decline in SMB forecast skill and the under-prediction of ice surface runoff during years and with a strong negative NAO also corresponds to an overestimate in the predicted summertime sea ice cover at leads of 3-4 months. Localized feedbacks between sea ice extent and ice sheet runoff (e.g., *Liu et al.* [2016]) and initial conditions and ice sheet surface albedo feedbacks may also play a role in summertime SMB forecasting [*Box et al.*, 2012; *Tedesco et al.*, 2013]. However, in forecasting seasonal changes in GrIS surface conditions, the need to accurately predict the phase of the summertime NAO is clearly a limiting factor.

Aside from the trend toward reduced skill in the retrospective period, May forecasts are found to predict approximately 66 percent of the total variance of the JJA (lead times of 1-3 months) SMB relative to MERRA-2. This reduction suggests that much of the skill in predicting summer SMB arises from the ability to predict the recent trend towards enhanced ice sheet melt and runoff [*Khan et al.*, 2015], and not necessarily in predicting the SMB interannual variability.

6.4 Results of Aerosol Forecast

We used the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2, *Gelaro et al.* [2017]) aerosol optical depth (AOD) at 550nm channel reanalysis *Randles et al.* [2017] to evaluate the GEOS-5 sub-seasonal to seasonal (S2S) aerosol hindcast simulations. Our analysis focused on the trimesters December, January, and February (DJF), and July, August, and September (JAS) from 2000 to 2015.

The MERRA-2 aerosol reanalysis applies the Goddard Aerosol Assimilation System (GAAS, *Buchard et al.* [2015]; **?**). The AOD observing system used in MERRA-2 includes ground-based Aerosol Robotic Network (AERONET) direct measurements of AOD **?**, AOD from the Multiangle Imaging SpectroRadiometer (MISR) over bright surfaces **?**, and bias-corrected near-real-time (NNR) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) from Terra and Aqua, and from the Advanced Very High Resolution Radiometer (AVHRR) instruments. *Randles et al.* [2017] describes in details the data and its spatial and temporal coverage. The aerosol emissions fields used for the GEOS-5-S2S hindcast simulations were the same as the applied for the MERRA-2. Emissions of dust and sea salt are wind driven **??**, respectively. Sea salt emission is also modulated with a sea surface temperature (SST)-derived correction following **?**. Biomass burning emissions have daily variability and are from the Quick Fire Emissions Dataset (QFED) version 2.4-r6 **?**. MERRA-2 includes bias corrected aerosol data assimilation and the GEOS5-S2S aerosol hindcast data analysis also includes the bias correction relative to MERRA-2. Therefore, comparing them, we are evaluating the model performance to predict the aerosol distribution as result of transport and removal processes.

Figure 35 shows scatter plots of the globally and monthly averaged AOD from the GEOS5-S2S ensemble mean relative to the MERRA-2 correspondents, from 2000 to 2015. However, climate model has a better performance during the Austral winter (R=0.81, R2=0.65, SE=0.006 and bias=-0.001) compared to boreal winter (R=0.73, R2=0.54, SE=0.005 and bias=0.015).

In general, the GEOS5-S2S aerosol global mean and seasonal spatial distributions are in agreement with MERRA-2, capturing the main patterns of the biomass burning aerosols over South America, Africa austral and the South Atlantic Ocean, the position of the dust plume coming from the Sahara desert, and the Asian pollution plume (Figure 36 panels A-B and D-E). The global mean AOD from GEOS5-S2S for the JAS and DJF trimesters are 0.16 and 0.12, respectively; while the correspondent MERRA-2 values are 0.17 and 0.14. However, we observed biases over a few specific regions (Figure 36, panels C and F). During the Austral winter (JAS) GEOS5-S2S overestimates the AOD on the southwestern coast of Africa, India, and Boreal forest in North America and Asia, and underestimates over South America and southeast Asian, associated to biomass burning emissions. It is noticeable (not shown) that the AOD biases correlate well with the precipitation biases in South America and Africa. Therefore, the climate model overestimation and underestimation over Africa and South America, respectively, is likely related to the model ability to predict precipitation accurately, and therefore the wet removal processes over these regions. For the same period, GEOS5-S2S also overestimates the AOD on the north of Africa and the Middle East, related to dust emissions. The positive AOD biases associated with dust aerosols are related with stronger winds simulated by the GEOS5-S2S compared to the meteorological reanalysis. A similar feature and bias have been previously reported for the GEOS5 model results **???**. During the Boreal winter (DJF), the climate model predicted an intense aerosol loading all over the Arctic region, and over Central Africa, associated with boreal fires. On the other side, the GEOS5-S2S prediction underestimated the AOD over India, northern South America, and eastern Asia. The climate prediction of sea-salt aerosols over oceanic regions are typically underestimated both over the Southern and Northern Hemispheres during the Austral and Boreal winter. This underestimation is likely related to the prediction of meteorological factors influencing sea-salt emissions, such as wind speed at 10m and the sea surface temperature (SST). However, the possible effects of these influences are yet not well understood and therefore will not be discussed here.

7 Summary and Future Directions

Acknowledgments
The authors would also like to thank Nicole Roth for her meticulous editing of the bibliography.

References

Adler, R. F., Huffman, G. J., Chang, A.,Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E. (2003), The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), *J. of Hydrometeorology, 4*(6), 1147-1167.

Alexander, M. A., Bhatt, U. S., Walsh, J. E., Timlin, M. S., Miller, J. S., Scott, J. D. (2004), The Atmospheric Response to Realistic Arctic Sea Ice Anomalies in an AGCM during Winter.*Journal of Climate*, *17*,6, 890–905, doi:10.1175/1520-0442(2004)017<0890:TARTRA>2.0.CO;2.

Antonov, J. I., Seidov, D. and Boyer, T. P. and Locarnini, R. A. and Mishonov, A. V. and Garcia, H. E. and Baranova, O. K. and Zweng, M. M. and Johnson, D. R. (2010), World Ocean Atlas 2009, Volume 2: Salinity, U.S. Government Printing Office, Washington, D.C., *NOAA Atlas NESDIS 69*, 184 pp.

(2013), The altimeter products were produced by Ssalto/Duacs and distributed by Aviso, with support from Cnes (http://www.aviso.oceanobs.com/duacs/). in SSALTO/DUACS User Handbook:(M)SLA and (M)ADT Near-Real Time and Delayed Time Products, edited by AVISO, p. 58, http://www.aviso.oceanobs.com/en/data/products/sea-surface-height-products/global/msla/index.html.

Ban, R. J., Bitz, C. M., Brown, A., Chassignet, E., Dutton, J. A., Hallberg, R., Kamrath, A., Kleist, D., Lermusiaux, P. F. J., Lin, H., Myers, L., Pullen, J., Sandgathe, S., Shafer, M., Waliser, D., and Zhang, C. (2016), Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. *National Academies Press*, doi:10.17226/21873.

Bloom, S. C. and Takacs, L. L. and DaSilva, A. M. and Ledvina, D. (1996), Data assimilation using incremental analysis updates.*Monthly Weather Review*, **124**,6, 1256-1271.

Bhatia, M. P., Kujawinski, E. B., Das, S. B., Breier, C. F., Henderson, P. B., and Charette, M. A. (2013). Greenland meltwater as a significant and potentially bioavailable source of iron to the ocean. *Nature Geoscience*, textit6,4, 274–278, doi:10.1038/ngeo1746.

Blanchard-Wrigglesworth, E., Barthelemy, A., Chevallier, M., Cullather, R., FuÄŊkar, N., Massonnet, F., et al. (2017). Multi-model seasonal forecast of Arctic sea-ice: forecast uncertainty at pan-Arctic and regional scales. *Climate Dynamics*, *49*,4, 1399–1410, doi:10.1007/s00382-016-3388-9.

Borovikov, A., R. Cullather, R. Kovach, J. Marshak, G. Vernieres, Y. Vikhliaev, B. Zhao, Zhao Li, 2017. GEOS-5 seasonal forecast system. *Clim. Dyn.*, https://doi.org/10.1007/s00382-017-3835-2

Bacmeister, J. T., and G. L. Stephens (2011), Spatial statistics of likely convective clouds in CloudSat data, *J. Geophys. Res., 116*, D04104, doi:10.1029/2010JD014444.

Barahona, D., A. Molod, J. Bacmeister, A. Nenes, A. Gettelman, H. Morrison, V. Phillips, and A. Eichmann (2014), Development of two-moment cloud microphysics for liquid and ice within the NASA Goddard Earth Observing System Model (GEOS-5), *Geosci. Model Dev., 7*, 1733–1766, doi:10.5194/gmd-7-1733-2014.

Box, J. E., Fettweis, X., Stroeve, J. C., Tedesco, M., Hall, D. K., and Steffen, K. (2012). Greenland ice sheet albedo feedback: thermodynamics and atmospheric drivers. *The Cryosphere*, *6*(4), 821–839, doi:10.5194/tc-6-821-2012.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A.,and Yang, X. (2012). Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophysical Research Letters*, *44*(10), 4953–4964, doi:10.1002/2017GL073155.

Carton, J.A.,(1989), Estimates of Sea-Level in the Tropical Atlantic-Ocean Using Geosat Altimetry, *J. Geosphys. Oc.*, 94(C6), 8029-8039.

Cavalieri, D. J., Parkinson C. L., Gloersen, P. and Zwally, H. J. (1996) Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. [1978-2017]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:10.5067/8GQ8LZQVL0VL. Accessed January 2018.

Chou, M.-D. (1990), Parameterizations for the absorption of solar radiation by O_2 and CO_2 with applications to climate studies, *J. Climate, 3*, 209–217.

Chelton, D. B. and DeSzoeke, R. A. and Schlax, M. G. and El Naggar, K. and Siwertz, N. (1998), Geographical variability of the first baroclinic Rossby radius of deformation. *Journal of Physical Oceanography*, **28**,3,433-460.

Chevallier, M., and Salas-MÃľlia, D. (2011). The Role of Sea Ice Thickness Distribution in the Arctic Sea Ice Potential Predictability: A Diagnostic Approach with a Coupled GCM. *Journal of Climate*, 25(8), 3025–3038. doi:10.1175/JCLI-D-11-00209.1.

Chou, M.-D. (1992), A solar radiation model for use in climate studies, *J. Atmos. Sci., 49*, 762–772.

Chou, M.-D., and M. J. Suarez (1994), An efficient thermal infrared radiation parameterization for use in general circulation models, NASA Tech. Memorandum, NASA/TM-1994-104606, Vol. 3, 85 pp.

Cullather, R. I., Nowicki, S. M. J., Zhao, B., and Suarez, M. J. (2014). Evaluation of the Surface Representation of the Greenland Ice Sheet in a General Circulation Model. *Journal of*

Climate, 27(13), 4835âĂŞ4856. doi:10.1175/JCLI-D-13-00635.1.

Day, J. J., Hawkins, E., and Tietsche, S. (2014). Will Arctic sea ice thickness initialization improve seasonal forecast skill? *Geophysical Research Letters*, 41(21), 7566âĂŞ7575. doi:10.1002/2014GL061694.

Delcroix, T., J. P. Boulanger, F. Masia, and C. Menkes (1994), Geosat-Derived Sea-Level and Surface Current Anomalies in the Equatorial Pacific During the 1986-1989 El-NiÃśo and La-NiÃśa, *J. Geophys. Res. Oc.*, 99(C12), 25093-25107.

Deser, C., Tomas, R., Alexander, M., and Lawrence, D. (2010), he Seasonal Atmospheric Response to Projected Arctic Sea Ice Loss in the Late Twenty-First Century. *Journal of Climate*, 23(2), 333-351. doi:10.1175/2009JCLI3053.1.

Emanuel, K. A., and Nolan D. S. (2004), Tropical cyclone activity and global climate, *26th Conf. on Hurricane and Tropical Meteorology*, Miami, FL, 240-241.

Garcia, R. R., and B. A. Boville (1994), Downward control of the mean meridional circulation and temperature distribution of the polar winter stratosphere, *J. Atmos. Sci., 51*, 2238–2245.

Garfinkel, C. I., A. Molod, L. D. Oman, and I.-S. Song (2011), Improvement of the GEOS-5 AGCM upon updating the air-sea roughness parameterization, *Geophys. Res. Lett., 38*, L18702, doi:10.1029/2011GL048802.

Gelaro, R., et al. (2017), The Modern-Era Retrospective Analysis for Research and Applications, Version-2 (MERRA-2), *J. Climate, 30*, 5419-5454, doi:10.1175/JCLI-D-16-0758.1.

Gent, P. and J. McWilliams, (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, 20, 150âĂŞ155.

Godfrey, J. S., (1989). A SVERDRUP MODEL OF THE DEPTH-INTEGRATED FLOW FOR THE WORLD OCEAN ALLOWING FOR ISLAND CIRCULATIONS. *Geosphys. and Astro. Fluid Dyn.*, 45(1-2), 89-112.

Good, S. A., M. J. Martin, and N. A. Rayner, (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geosphys. Res. Oc.*, 118(12), 6704-6716, doi:10.1002/2013JC009067.

Evensen, G. (1994), Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research-Oceans*, **99**, C5, 10143-10162.

Gottschalck, J., and coauthors (2010), A framework for assessing operational Madden-Julian Oscillation forecasts: A CLIVAR MJO working group project. *Bull. Amer. Meteor. Soc.*, 91, 1247-1258, doi:10.1175/2010BAMS2816.1.

Griffies, S., A. Gnanadesikan, K. Dixon, J. Dunne, R. Gerdes, M. Harrison, A. Rosatti, J. Russel, B. Samuels, M. Spelman, M. Winton, and R. Zhang, (2005). Formulation of an ocean model for global climate simulations.*Ocean Science*, 1, 1025–1035.

Griffies, S., (2012). Elements of the Modular Ocean Model (MOM). `http://mdl-mom5.herokuapp.com/web/docs/project/MOM5_elements.pdf`.

Hamilton, L. C., and Stroeve, J. (2016), 00 predictions: the SEARCH Sea Ice Outlook 2008âĂŞ2015. *Polar Geography*, 39(4) 274-287. doi:10.1080/1088937X.2016.1234518.

Helfand, H. M., and S. D. Schubert (1995), Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States, *J. Climate, 8*, 784-806.

Hill, C., C. DeLuca, V. Balaji, M. Suarez, and A. da Silva (2004), Architecture of the Earth System Modeling Framework, *Computing in Science and Engineering, 6*, 1, doi:10.1109/MCISE.2004.1255817.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu (2009), Improving the Global Precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.

Hunke, E. and W. Lipscomb, (2008). The Los Alamos sea ice model, documentation and software manual, version 4.0. *Technical report, Los Alamos National Laboratory.*

Hunt, B. R. and Kostelich, E. J. and Szunyogh, I. (2007), Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D-Nonlinear Phe-*

nomena, **230**, 1-2, 112-126, doi:10.1016/j.physd.2006.11.008.

Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette (2000), Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3, *J. Geophys. Res., 105* (D11), 14,873–14,890, doi:10.1029/2000JD900091.

Ingleby, B. and Huddleston, M.(2007), Quality control of ocean temperature and salinity profiles - Historical and real-time data, *Journal of Marine Systems*, **65**, 1-4, 158-175, doi:10.1016/j.jmarsys.2005.11.019.

Jin, F.F. (1997), An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual mode, *J. of Atm. Sci.*, 54(7), 811-829.

Jung, T., Kasper, M. A., Semmler, T., and Serrar, S. (2014). Arctic influence on sub-seasonal midlatitude prediction. *Geophysical Research Letters*, 41(10), 3676-3680. doi:10.1002/2014GL059961.

Karspeck, A. R. and Yeager, S. and Danabasoglu, G. and Hoar, T. and Collins, N. and Raeder, K. and Anderson, J. and Tribbia, J. (2013), An Ensemble Adjustment Kalman Filter for the CCSM4 Ocean Component, *Journal of Climate*, **26**, 19, 7392-7413, doi:10.1175/jcli-d-12-00402.1.

Keppenne, C. L. and Rienecker, M. M. and Jacob, J. P. and Kovach, R. (2008), Error covariance modeling in the GMAO ocean ensemble Kalman filter, *Monthly Weather Review*, **136**, 8, 2964-2982, doi:10.1175/2007mwr2243.1.

Keppenne, C., M. Rienecker, R. Kovach, and G. Vernieres (2014), Background Error Covariance Estimation using Information from a Single Model Trajectory with Application to Ocean Data Assimilation into the GEOS-5 Coupled Model, *NASA/TM-2014-104606*, Goddard Space Flight Center, Greenbelt, MD 20771.

Khan, S. A., Aschwanden, A., BjÃÿrk, A. A., Wahr, J., Kjeldsen, K. K., and KjÃęr, K. H. (2015). Greenland ice sheet mass balance: a review. *Reports on Progress in Physics*, 78(4), 046801. doi:10.1088/0034-4885/78/4/046801.

Kirtman BP, Min D, Infanti JM, Kinter JL III, Paolino DA, Zhang Q et al (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull Am Meteorol Soc* 95(4), 585âĂŞ601.

Kondo, J. (1975), Air-sea bulk transfer coefficients in diabatic conditions, *Boundary Layer Meteor., 9*, 91–112.

Large, W. G., and S. Pond (1981), Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr., 11*, 324–336.

Lean, J. (2000), Evolution of the Sun's spectral irradiance since the Maunder Minimum, *Geophys. Res. Lett., 27*, 2425–2428.

Large, W., J. McWilliams, and S. Doney, (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, 32, 363âĂŞ403.

Large,W., G. Danabasoglu, J. McWilliams, P. Gent, and F. Bryan, (2001). Equatorial circulation of a global ocean climate model with anisotropic horizontal viscosity. *Journal of Physical Oceanography*, 31, 518âĂŞ536.

Lin, S.-J. (2004), A vertically Lagrangian finite-volume dynamical core for global models, *Mon. Wea. Rev., 132*, 2293–2307.

Lim, Young-Kwon and Schubert, Siegfried D. and Nowicki, Sophie M. J. and Lee, Jae N. and Molod, Andrea M. and Cullather, Richard I. and Zhao, Bin and Velicogna, Isabella (2016). Atmospheric summer teleconnections and Greenland Ice Sheet surface mass variations: insights from MERRA-2. *Environmental Research Letters*, 11(2), 024002. doi:10.1088/1748-9326/11/2/024002.

Lim, Y., Son, S., and Kim, D. (2018), MJO prediction skill of the subseasonal-to-seasonal prediction models. *J. Climate*, 31, 4075-4094, doi:10.1175/JCLI-D-17-0545.1.

Locarnini, R. A. and Mishonov, A. V. and Antonov, J. I. and Boyer, T. P. and Garcia, H. E. and Baranova, O. K. and Zweng, M. M. and Johnson, D. R. (2010), World Ocean Atlas 2009, Volume 1: Temperature, NOAA Atlas NESDIS 68, U.S. Government Printing Office, Washington, D.C., 184 pp.

Locarnini, R. A. and Mishonov, A. V. and Antonov, J. I. and Boyer, T. P. and Garcia, H. E. and Baranova, O. K. and Zweng, M. M. and Paver, C. R. and Reagan, J. R. and Johnson, D. R. and Hamilton, M. and D. Seidov, World Ocean Atlas 2013, Volume 1: Temperature, S. Levitus, Ed., A. Mishonov Technical Ed., NOAA Atlas NESDIS 73, 43pp.

Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith (2000), A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev., 138*, 3187-3199.

Louis, J., and J. Geleyn (1982), A short history of the PBL parameterization at ECMWF. Proc. ECMWF Workshop on Planetary Boundary Layer Parameterization, Reading, United Kingdom, ECMWF, 59–80.

Liu, J., Chen, Z., Francis, J., Song, M., Mote, T., and Hu, Y. (2016). Has Arctic Sea Ice Loss Contributed to Increased Surface Melting of the Greenland Ice Sheet? *Journal of Climate*, 29(9), 3373âĂŞ3386. doi:10.1175/JCLI-D-15-0391.1.

Lynch-Stieglitz, M. (1994). The Development and Validation of a Simple Snow Model for the GISS GCM. *Journal of Climate*, 7(12), 1842âĂŞ1855. doi:10.1175/1520-0442(1994)007<1842:TDAVOA>2.0.CO;2.

Maslanik, J. and J. Stroeve. (1999, updated daily). Near-Real-Time DMSP SSMIS Daily Polar Gridded Sea Ice Concentrations, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi:doi.org/10.5067/U8C09DWVX9LM. January 2018.

Markus, T., and D. J. Cavalieri (2000). An enhancement of the NASA Team sea ice algorithm. *Ieee Transactions on Geoscience and Remote Sensing, 38*(3), 1387-1398, doi:10.1109/36.843033.

McGrath-Spangler, E. L., and A. Molod (2014), Comparison of GEOS-5 AGCM planetary boundary layer depths computed with various definitions, *Atmos. Chem. Phys., 14*, 6717–6727, doi:10.5194/acp-14-6717-2014.

McFarlane, N. A. (1987), The effect of orographically excited gravity-wave drag on the circulation of the lower stratosphere and troposphere, *J. Atmos. Sci., 44*, 1775–1800.

McPhaden, Michael J. and Busalacchi, Antonio J. and Anderson, David L. T. (2010), A TOGA RETROSPECTIVE, *Oceanog.*, **23**, 3, 86-103, doi:10.5670/oceanog.2010.26.

Meyers, G., R. J. Bailey, and A. P. Worby (1995), eostrophic Transport of Indonesian Throughflow, *Deep-Sea Res. Part 1-Oceanographic Research Papers*, 42(7), 1163-1174.

Mesinger, F. and A. Arakawa, (1976). Numerical methods used in atmosphere models. *GARP Publication Series*.

Molod, A. (2012), Constraints on the total water PDF in GCMs from AIRS data and a high resolution model, *J. Climate, 25*, 8341–8352, doi: 10.1175/JCLI-D-11-00412.1.

Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann (2012), The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna, NASA Tech. Rep. Series on Global Modeling and Data Assimilation, NASA/TM-2012-104606, Vol. 28, 117 pp.

Molod, A., M. J. Suarez and G. Partyka (2013), The impact of limiting ocean roughness on GEOS-5 AGCM tropical cyclone forecasts, *Geophys. Res. Lett.*, **40**, 411âĂŞ415, doi:10.1029/2012GL053979.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister (2015), Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, *Geosci. Model Dev., 8*, 1339–1356, doi:10.5194/gmd-8-1339-2015.

Moorthi, S., and M. J. Suarez (1992), Relaxed Arakawa Schubert: A parameterization of moist convection for general circulation models, *Mon. Wea. Rev., 120*, 978–1002.

Mortensen, J., Bendtsen, J., Motyka, R. J., Lennert, K., Truffer, M., Fahnestock, M., and Rysgaard, S. (2013). On the seasonal freshwater stratification in the proximity of fast-flowing tidewater outlet glaciers in a sub-Arctic sill fjord. *Journal of Geophysical Re-*

search: Oceans, 118(3), 1382âĂŞ1395. doi:10.1002/jgrc.20134.

Murray, R., (1996). Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, 126, 251âĂŞ273.

Oke, P. R., G. B. Brassington, D. A. Griffin, and A. Schiller (2010). Ocean data assimilation: a case for ensemble optimal interpolation. *Australian Meteorological and Oceanographic Journal*, 59, 67-76, doi:10.22499/2.5901.008.

Ott, E. and Hunt, B. R. and Szunyogh, I. and Zimin, A. V. and Kostelich, E. J. and Corazza, M. and Kalnay, E. and Patil, D. J. and Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus Series A-Dynamic Meteorology and Oceanography*, **56**, 5 415-428.

Overland, J., Francis, J. A., Hall, R., Hanna, E., Kim, S.-J., and Vihma, T. (2015). The Melting Arctic and Midlatitude Weather Patterns: Are They Connected? *Journal of Climate*, 28(20), 7917-7932, doi:10.1175/JCLI-D-14-00822.1.

Penny, S. G. and Behringer, D. W. and Carton, J. A. and Kalnay, E. (2015). A Hybrid Global Ocean Data Assimilation System at NCEP. *Mon. Wea. Rev.*, **143** 11, 4660-4677, doi:10.1175/mwr-d-14-00376.1.

Penny, S. G. and Kalnay, E. and Carton, J. A. and Hunt, B. R. and Ide, K. and Miyoshi, T. and Chepurin, G. A. (2013). The local ensemble transform Kalman filter and the running-in-place algorithm applied to a global ocean general circulation model. *Nonlinear Processes in Geophysics*, **20**, 6, 1031-1046, doi:10.5194/npg-20-1031-2013.

Picaut, J., and R. Tournier (1991). MONITORING THE 1979-1985 EQUATORIAL PACIFIC CURRENT TRANSPORTS WITH EXPENDABLE BATHYTHERMOGRAPH DATA,. *J. Geophys. Res.*, 96, 3263-3277.

Putman, W. M., and S.-J. Lin (2007), Finite-volume transport on various cubed-sphere grids, *J. Comp. Phys., 227*, 55-78, doi:10.1016/j.jcp.2007.07.022.

Rennermalm, A. K., Smith, L. C., Stroeve, J. C., and Chu, V. W. (2009). Does sea ice influence Greenland ice sheet surface-melt? *Environmental Research Letters*, 4(2), 024011. doi:10.1088/1748-9326/4/2/024011.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Q. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Clim., 15*(13), 1609-1625.

Rienecker, M. M., et al. (2008), The GEOS-5 Data Assimilation System: Documentation of versions 5.0.1 and 5.1.0, and 5.2.0, NASA Tech. Rep. Series on Global Modeling and Data Assimilation, NASA/TM-2008-104606, Vol. 27, 92 pp.

Rienecker, M. M., et al. (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *J. Climate, 24*, 1-25, doi:10.1175/JCLI-D-11-00015.1.

Roemmich, D. and G. C. Johnson, and S. Riser, R. Davis, J. Gilson, W.B. Owens, W. Brechner, S.L. Garzoli, C. Schmid, and M. Ignaszewski (2009), The Argo Program Observing the Global Ocean with Profiling Floats, *Ocean.*, **22**, 2, 34-43, doi:10.5670/oceanog.2009.36.

Saha, S., and coauthors (2014), The NCEP Climate Forecast System Version 2. *J. Climate*, 27, 2185-2208, doi:10.1175/JCLI-D-12-00823.1.

Santoso, A., W. Cai, M. Collins, M. McPhaden, and F.-F. Jin (2015), ENSO EXTREMES AND DIVERSITY Dynamics, Teleconnections, and Impacts, *Bull. Am. Met. Soc, 96*(11), 1969-1972.

Schubert, S. D., H. Wang, R. D. Koster, M. J. Suarez and P. Y. Groisman (2014), Northern Eurasian heat waves and droughts, *J. Climate, 27*, 3169–3207, doi:10.1175/JCLI-D-13-00360.1.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R. (2011). Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research: Oceans*, 116(C5). doi:10.1029/2011JC007084.

Sciascia, R., Straneo, F., Cenedese, C., and Heimbach, P. (2013). Seasonal variability of submarine melt rate and circulation in an East Greenland fjord. *Journal of Geophysical Research: Oceans*, 118(5), 2492âĂŞ2506. doi:10.1002/jgrc.20142.

Screen, J. A., and Simmonds, I. (2010). The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, 464(7293), 1334âĂŞ1337. doi:10.1038/nature09051.

Screen, J. A., Deser, C., Smith, D. M., Zhang, X., Blackport, R., Kushner, P. J., et al. (2018). Consistency and discrepancy in the atmospheric response to Arctic sea-ice loss across climate models. *Nature Geoscience*, 11, 155âĂŞ163. doi:10.1038/s41561-018-0059-y.

Serreze, M. C., and Barry, R. G. (2011). Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change*, 77(1), 85âĂŞ96. doi:10.1016/j.gloplacha.2011.03.004.

Shum, C. K. and Ries, J. C. and Tapley, B. D. (1995), The accuracy and applications of satellite altimetry, *Geophys. J. Int.*, **121**, 2, 321-336, doi:10.1111/j.1365-246X.1995.tb05714.x.

Smith, L. C., Yang, K., Pitcher, L. H., Overstreet, B. T., Chu, V. W., Rennermalm, A. K., âĂę Behar, A. E. (2017). Direct measurements of meltwater runoff on the Greenland ice sheet surface. *Proceedings of the National Academy of Sciences*, 114(50), E10622âĂŞE10631. doi:10.1073/pnas.1707743114.

Sommaruga, R. (2015). When glaciers and ice sheets melt: consequences for planktonic organisms. *Journal of Plankton Research*, 37(3), 509âĂŞ518. doi:10.1093/plankt/fbv027.

Sprintall, J., S. E. Wijffels, R. Molcard, and I. Jaya (2009), Direct estimates of the Indonesian Throughflow entering the Indian Ocean: 2004-2006. *J. Geophys. Res. Oc.*, 114, 19, doi:C07001.

Stieglitz, M., Ducharne, A., Koster, R., and Suarez, M. (2001). The Impact of Detailed Snow Physics on the Simulation of Snow Cover and Subsurface Thermodynamics at Continental Scales. *Journal of Hydrometeorology*, 2(3), 228âĂŞ242. doi:10.1175/1525-7541(2001)002<0228:TIODSP>2.0.CO;2.

Suarez, M., A. Trayanov, C. Hill. P. Schopf, and Y. Vikhliaev (2007), MAPL: A high-level programming paradigm to support more rapid and robust encoding of hierarchical trees of interacting high-performance components, Proceedings of the 2007 Symposium on Component and Framework Technology in High-Performance and Scientific Computing, pp. 11–20, doi:10.1145/1297385.1297388.

Tedesco, M., Fettweis, X., Mote, T., Wahr, J., Alexander, P., Box, J. E., and Wouters, B. (2013). Evidence and analysis of 2012 Greenland records from spaceborne observations, a regional climate model and reanalysis data. *The Cryosphere*, 7(2), 615âĂŞ630. doi:10.5194/tc-7-615-2013.

Tokioka, T., K. Yamazaki, A. Kitoh, and T. Ose (1988), The equatorial 30âĂŞ60 day oscillation and the Arakawa-Schubert penetrative cumulus parameterization. *J. Meteorol. Soc. Japan.*, **66**, 883âĂŞ901.

Vernieres, G. and Rienecker, M. and Kovach, R. and Keppenne, C. (2012). The GEOS-iODAS: Description and Evaluation. NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TMâĂŤ2012-104606.

Wheeler, M. C., and Hendon, H. H. (2004), An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, 132, 1917-1931.

Wijffels, S. E., G. Meyers, and J. S. Godfrey (2008). A 20-yr average of the Indonesian throughflow: Regional currents and the interbasin exchange, *J. Phys. Ocean., 38*(9), 1965-1978, doi:10.1175/2008jpo3987.1.

Xie, P. P., and P. A. Arkin (1997),Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bull. Am. Met. Soc., 78*(11), 2539-2558.

Zhang, S. and Harrison, M. J. and Rosati, A. and Wittenberg, A. (2007), System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 10, 3541-3564, doi:10.1175/mwr3466.1.

Zweng, M.M and Reagan, J.R. and Antonov, J.I. and Locarnini, R.A. and Mishonov, A.V. and Boyer, T.P. and Garcia, H.E. and Baranova, O.K. and Johnson, D.R. and D.Seidov and Biddle, M.M. (2013), World Ocean Atlas 2013, Volume 2: Salinity, S. Levitus, Ed., A. Mishonov Technical Ed., NOAA Atlas NESDIS 75, 39 pp.

Koster, R.D., M.J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000). A catchment-based approach to modeling land surface processes in a GCM, Part 1, Model structure. J. Geophys. Res., 105, 24 809âĂŞ24 822.

Hill, C., C. DeLuca, V. Balaji, M. Suarez, and A. da Silva (2004). Architecture of the Earth System Modeling Framework. Computing in Science and Engineering, Volume 6, Number 1, pp. 18-28.

Price, J.F., C.N.K. Moores and J.C. Van Leer (1978). Observation and Simulation of Storm-Induced Mixed-Layer Deepening. Journal of Physical Oceanography, vol. 8, pp. 582-599.

Brassington GB, Martin MJ, Tolman HL, Akella S, Balmeseda M, Chambers CRS, Chassignet E, Cummings JA, Drillet Y, Jansen PAEM, Laloyaux P, Lea D, Mehra A, Mirouze I, Ritchie H, Samson G, Sandery PA, Smith GC, Suarez M, Todling R. 2015. Progress and challenges in short- to medium-range coupled prediction. J Operational Oceanography. 8: s239âĂŞ s258, doi: 10.1080/1755876X.2015.1049875.

Dee DP, Balmaseda M, Balsamo G, Engelen R, Simmons AJ, TheÌĄpaut JN. 2014. Toward a consistent reanalysis of the climate system. Bull. Amer. Meteor. Soc. 95: 1235âĂŞ1248, doi:http://dx.doi.org/ 10.1175/BAMS-D-13-00043.1

Lea DJ, Mirouze I, Martin MJ, King RR, Hines A, Walters D, Thurlow M. 2015. Assessing a new coupled data assimilation system based on the Met Office coupled atmosphereâĂŞlandâĂŞ oceanâĂŞsea ice model. Mon Weather Rev 143: 4678âĂŞ 4694, doi: http://dx.doi.org/10.1175/MWR-D-15-0174.1.

Laloyaux P, Thèpaut JN, Dee D. 2016b. Impact of scatterometer surface wind data in the ECMWF coupled assimilation system. Mon Weather Rev 144: 1203âĂŞ 1217, doi:http://dx.doi.org/10. 1175/MWR-D-15-0084.1.

Sluka, T. C., S. G. Penny, E. Kalnay, and T. Miyoshi (2016), Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation, Geophys. Res. Lett., 43, 752âĂŞ759, doi:10.1002/2015GL067238.

Wada, A., and M. Kunii (2017), The role of oceanâĂŘatmosphere interaction in Typhoon Sinlaku (2008) using a regional coupled data assimilation system, J. Geophys. Res. Oceans, 122, 3675âĂŞ3695, doi:10.1002/2017JC012750.

Akella, S. , Todling, R. and Suarez, M. (2017), Assimilation for skin SST in the NASA GEOS atmospheric data assimilation system. Q.J.R. Meteorol. Soc, 143: 1032-1046. doi:10.1002/qj.2988

Collow, Allison B. Marquardt, Sarith P. Mahanama, Michael G. Bosilovich, Randal D. Koster, and Siegfried D. Schubert, 2017. An Evaluation of Teleconnections Over the United States in an Ensemble of AMIP Simulations with the MERRA-2 Configuration of the GEOS Atmospheric Model. NASA/TM-2017-104606, Vol. 47, 68 pp. Document (5623 kB).

Reichle, R. H., Q. Liu, R. D. Koster, et al. 2017. "Land Surface Precipitation in MERRA-2." Journal of Climate, 30 (5): 1643-1664 [10.1175/jcli-d-16-0570.1]

Randles, CA, da Silva, Arlindo M and Buchard, Virginie and Colarco, PR and Darmenov, A and Govindaraju, R and Smirnov, A and Holben, B and Ferrare, R and Hair, John et al. 2017. "The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation". Journal of climate, 30 (17): 6823–6850.

Buchard, V and da Silva, AM and Colarco, PR and Darmenov, A and Randles, CA and Govindaraju, R and Torres, O and Campbell, J and Spurr, R. 2015. "Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis". Atmospheric Chemistry and Physics, 15 (10) : 5743. @articlebuchard2016evaluation, title=Evaluation of the surface PM2. 5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States, author=Buchard, V and da Silva, AM and Randles, CA and Colarco, P and Ferrare, R and Hair, J and Hostetler, C and Tackett, J and Winker, D, journal=Atmospheric Environment, volume=125, pages=100–111, year=2016, publisher=Elsevier @articleholben1998aeronet, title=AERONETÃćÂĂÂŤA federated instrument network and data archive for aerosol characterization, author=Holben, Brent N and Eck, T

FÃłÂŐÂň and Slutsker, IÃłÂŐÂň and Tanre, D and Buis, JP and Setzer, A and Vermote, E and Reagan, JA and Kaufman, YJ and Nakajima, T and others, journal=Remote sensing of environment, volume=66, number=1, pages=1–16, year=1998, publisher=Elsevier

@articlekahn2005multiangle, title=Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations, author=Kahn, Ralph A and Gaitley, Barbara J and Martonchik, John V and Diner, David J and Crean, Kathleen A and Holben, Brent, journal=Journal of Geophysical Research: Atmospheres, volume=110, number=D10, year=2005, publisher=Wiley Online Library

@articlemarticorena1995modeling, title=Modeling the atmospheric dust cycle: 1. Design of a soil-derived dust emission scheme, author=Marticorena, B and Bergametti, G, journal=Journal of Geophysical Research: Atmospheres, volume=100, number=D8, pages=16415–16430, year=1995, publisher=Wiley Online Library

@articlegong2003parameterization, title=A parameterization of sea-salt aerosol source function for sub-and super-micron particles, author=Gong, SL, journal=Global biogeochemical cycles, volume=17, number=4, year=2003, publisher=Wiley Online Library

@articlejaegle2011global, title=Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations, author=Jaeglé, L and Quinn, PK and Bates, TS and Alexander, B and Lin, J-T, journal=Atmospheric Chemistry and Physics, volume=11, number=7, pages=3137–3157, year=2011, publisher=Copernicus GmbH

@articledarmenov2015quick, title=The Quick Fire Emissions Dataset (QFED): Documentation of versions 2.1, 2.2 and 2.4, author=Darmenov, A and da Silva, A, journal=NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TM-2015-104606, volume=38, pages=212, year=2015

@articletextor2006analysis, title=Analysis and quantification of the diversities of aerosol life cycles within AeroCom, author=Textor, C and Schulz, M and Guibert, S and Kinne, Stefan and Balkanski, Y and Bauer, S and Berntsen, T and Berglen, T and Boucher, Olivier and Chin, M and others, journal=Atmospheric Chemistry and Physics, volume=6, number=7, pages=1777–1813, year=2006, publisher=Copernicus GmbH

@articlecolarco2010online, title=Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, author=Colarco, Peter and da Silva, Arlindo and Chin, Mian and Diehl, Thomas, journal=Journal of Geophysical Research: Atmospheres, volume=115, number=D14, year=2010, publisher=Wiley Online Library

@articlerandles2013direct, title=Direct and semi-direct aerosol effects in the NASA GEOS-5 AGCM: aerosol-climate interactions due to prognostic versus prescribed aerosols, author=Randles, CA and Colarco, PR and Da Silva, A, journal=Journal of Geophysical Research: Atmospheres, volume=118, number=1, pages=149–169, year=2013, publisher=Wiley Online Library

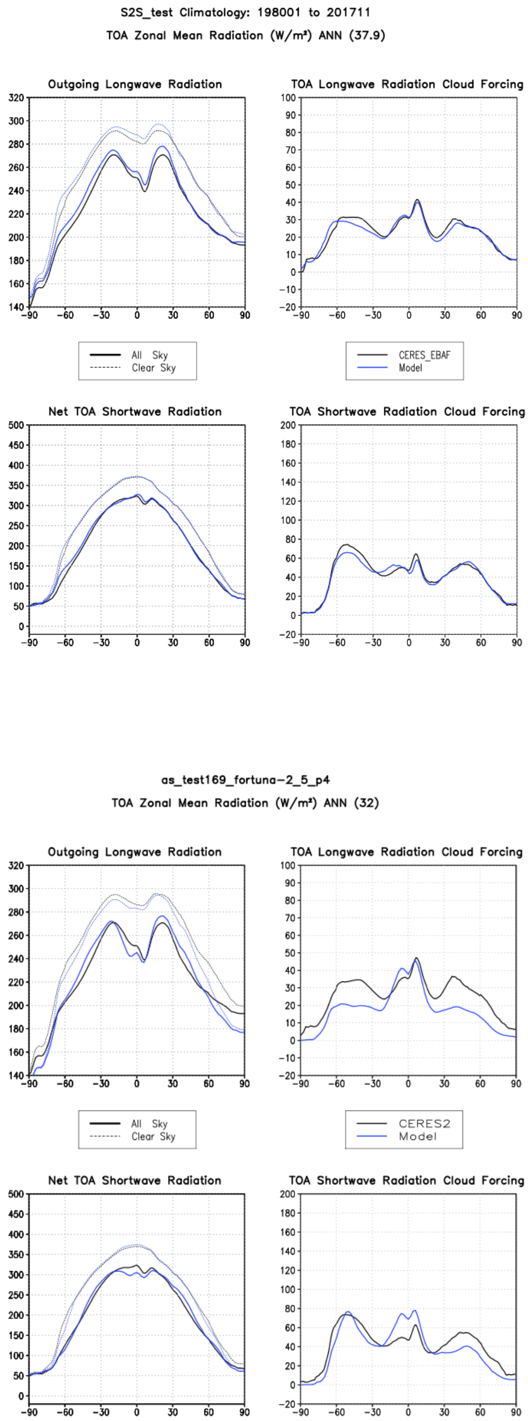

S2S_test Climatology: 198001 to 201711
TOA Zonal Mean Radiation (W/m²) ANN (37.9)
Outgoing Longwave Radiation
TOA Longwave Radiation Cloud Forcing
All Sky
Clear Sky
CERES_EBAF
Model
Net TOA Shortwave Radiation
TOA Shortwave Radiation Cloud Forcing
as_test169_fortuna-2_5_p4
TOA Zonal Mean Radiation (W/m²) ANN (32)
Outgoing Longwave Radiation
TOA Longwave Radiation Cloud Forcing
All Sky
Clear Sky
CERES2
Model
Net TOA Shortwave Radiation
TOA Shortwave Radiation Cloud Forcing

Figure 4.

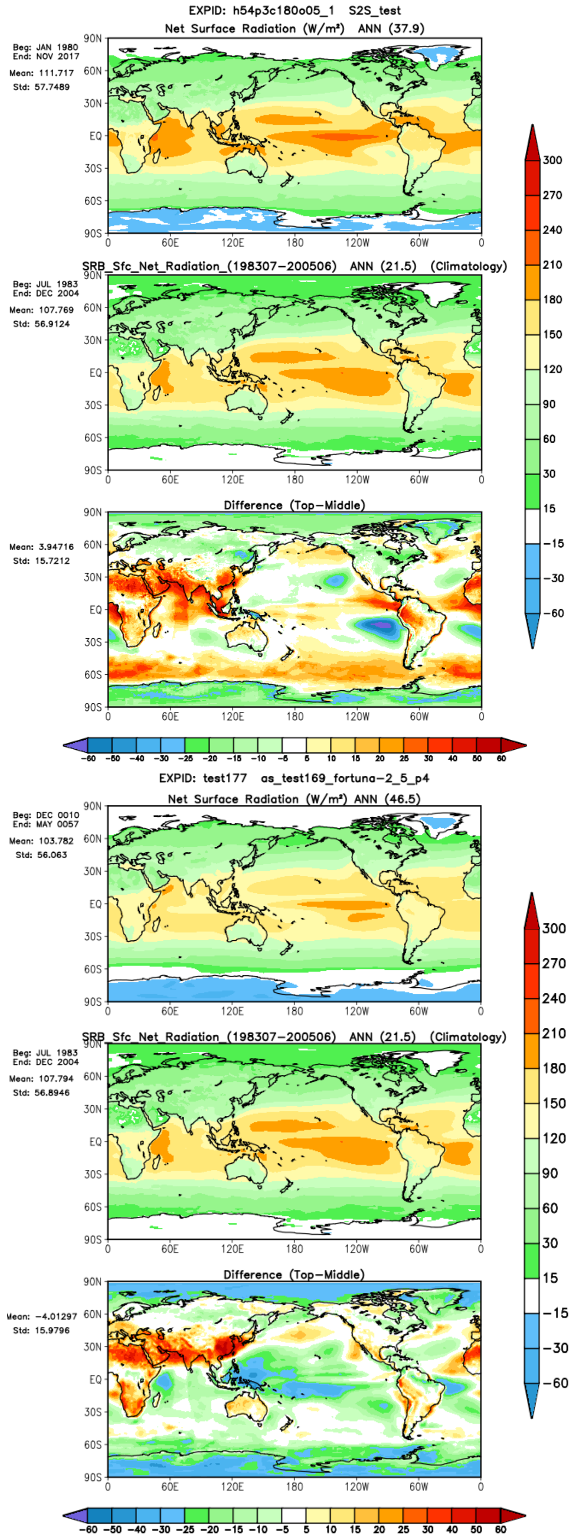
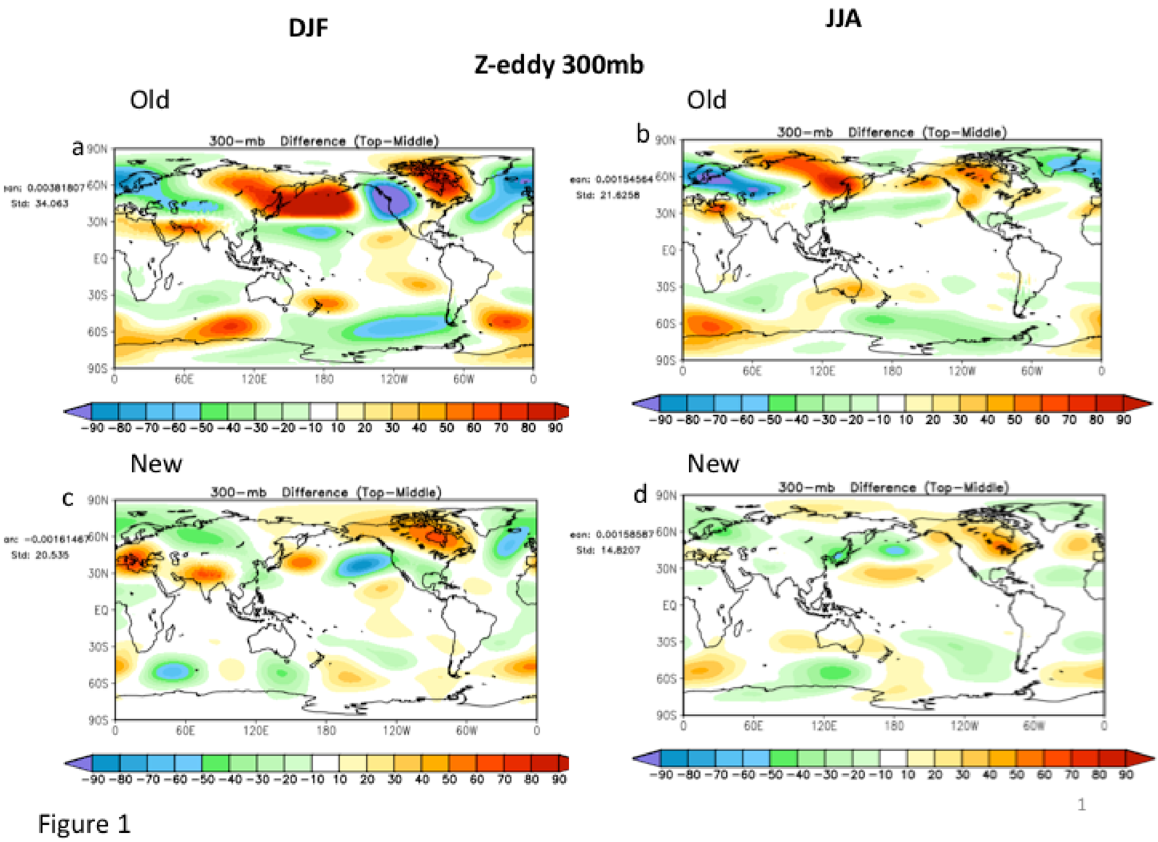

EXPID: h54p3c180o05_1 S2S_test
Net Surface Radiation (W/m²) ANN (37.9)
Beg: JAN 1980
End: NOV 2017
Mean: 111.717
Std: 57.7489
SRB_Sfc_Net_Radiation_(198307-200506) ANN (21.5) (Climatology)
Beg: JUL 1983
End: DEC 2004
Mean: 107.769
Std: 56.9124
Difference (Top-Middle)
Mean: 3.94716
Std: 15.7212
EXPID: test177 as_test169_fortuna-2_5_p4
Net Surface Radiation (W/m²) ANN (46.5)
Beg: DEC 0010
End: MAY 0057
Mean: 103.782
Std: 56.063
SRB_Sfc_Net_Radiation_(198307-200506) ANN (21.5) (Climatology)
Beg: JUL 1983
End: DEC 2004
Mean: 107.794
Std: 56.8946
Difference (Top-Middle)
Mean: -4.01297
Std: 15.9796
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
300
270
240
210
180
150
120
90
60
30
15
-15
-30
-60
-60 -50 -40 -30 -25 -20 -15 -10 -5 5 10 15 20 25 30 40 50 60

Figure 5.



483 **Figure 6.** Seasonal mean Eddy Height difference from MERRA-2 at 300 mb in m. a)December-January-
484 February mean for S2S-1_0, b) June-July-August mean for S2S-1_0, c) December-January-February mean for
485 S2S-2_1, d) June-July-August mean for S2S-2.

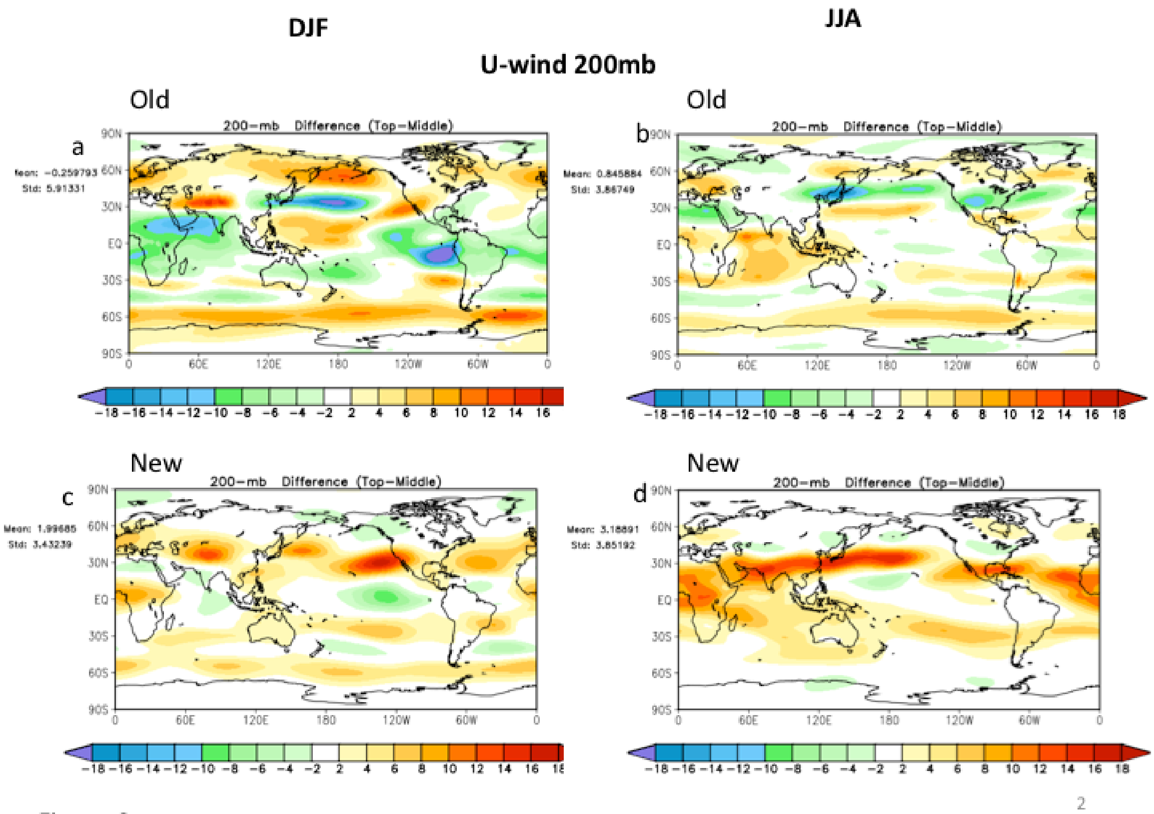
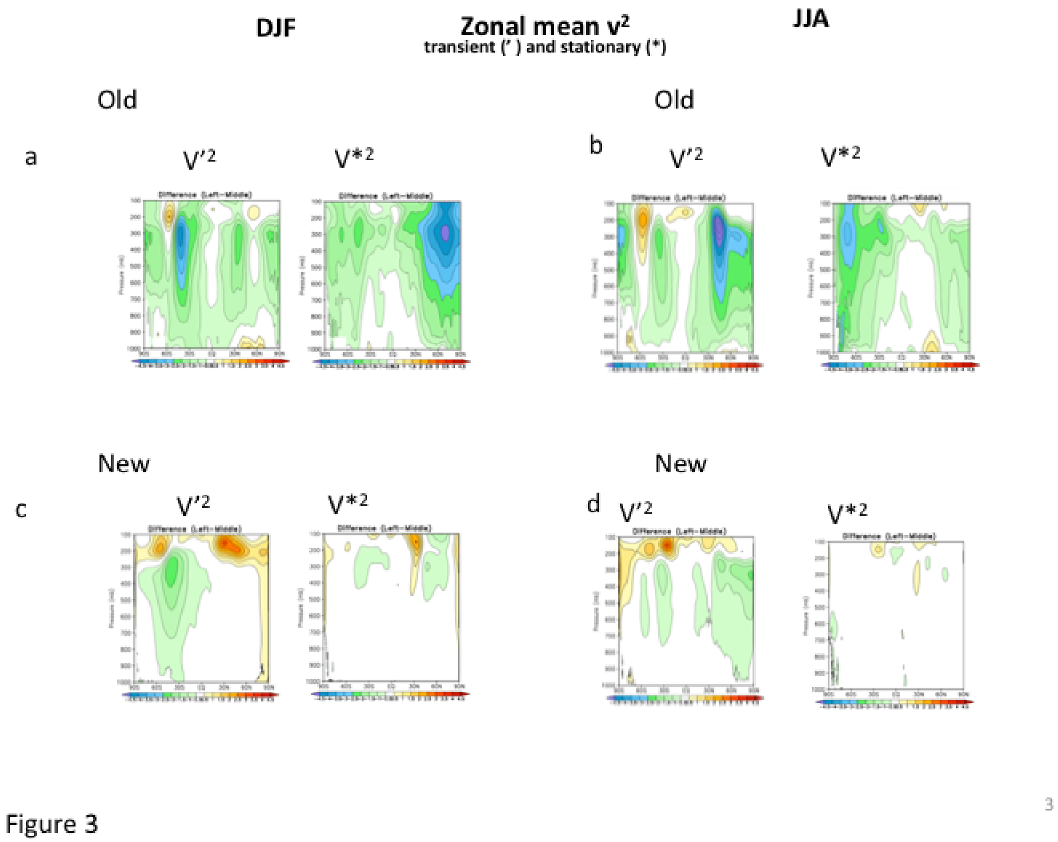


Figure 2

Figure 7. Seasonal mean Zonal wind difference from MERRA-2 at 200 mb in ms^{-1}. a) December-January-February mean for S2S-1, b) June-July-August mean for S2S-1, c) December-January-February mean for S2S-2, d) June-July-August mean for S2S-2.



489 **Figure 8.** Seasonal mean difference from MERRA-2 of the stationary and transient components of the variance of meridional wind in $m^2 s^{-2}$. a)December-January- February mean for S2S-1_0, b) June-July-August mean for S2S-1_0, c) December-January-February mean for S2S-2_1, d) June-July-August mean for S2S-2_1.

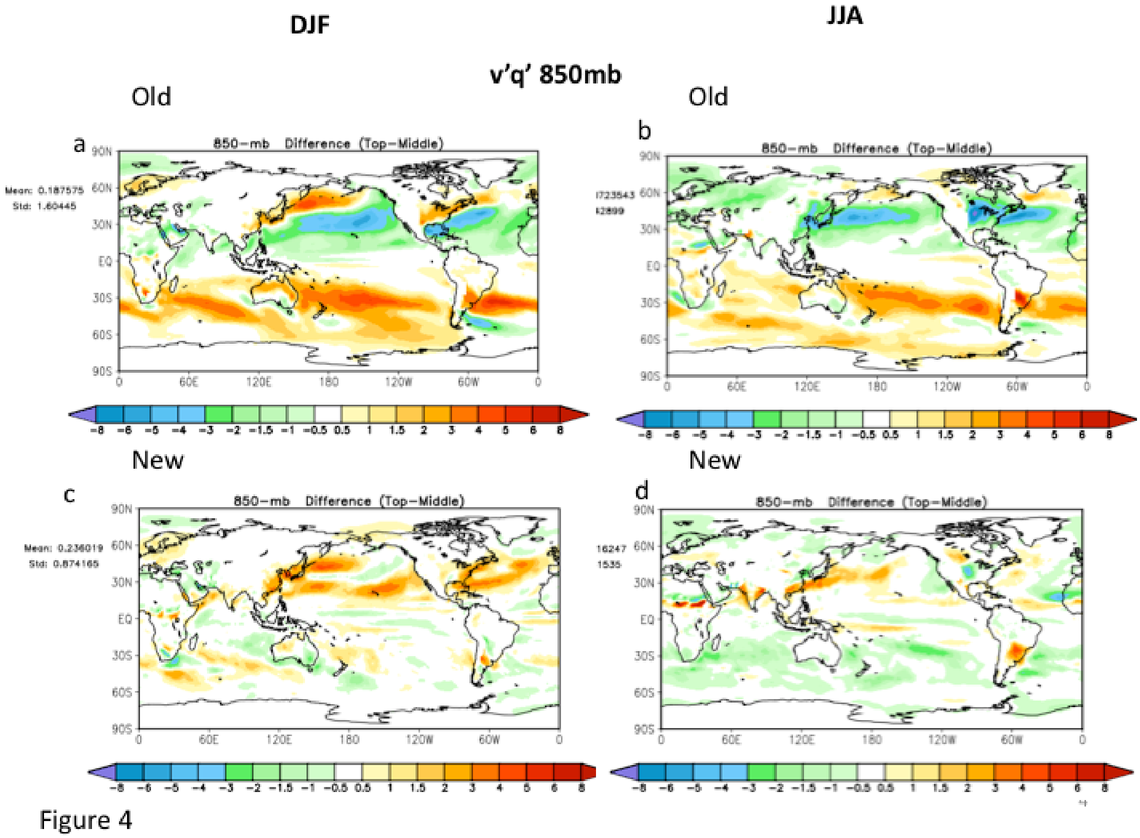


Figure 9. Seasonal mean difference from MERRA-2 of the transient meridional transport of moisture at 850 mb in $gkg^{-1}ms^{-1}$. a)December-January- February mean for S2S-1_0, b) June-July-August mean for S2S-1_0, c) December-January-February mean for S2S-2_1, d) June-July-August mean for S2S-2_1.

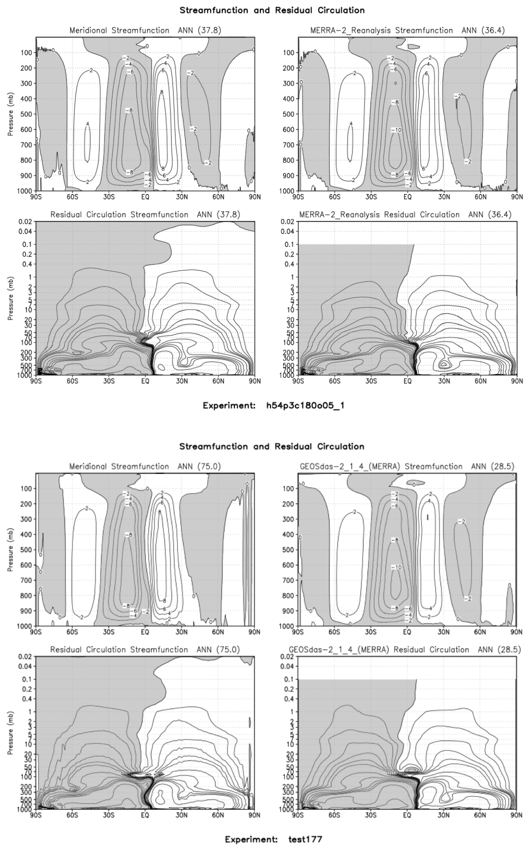
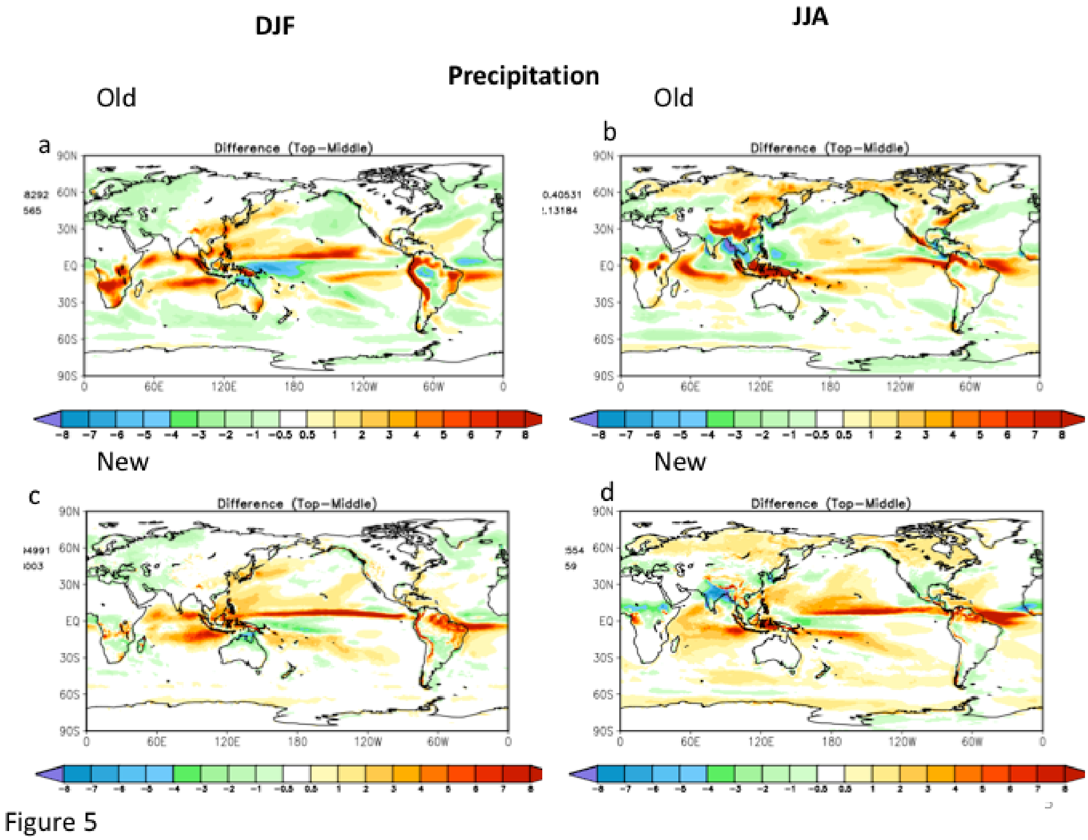
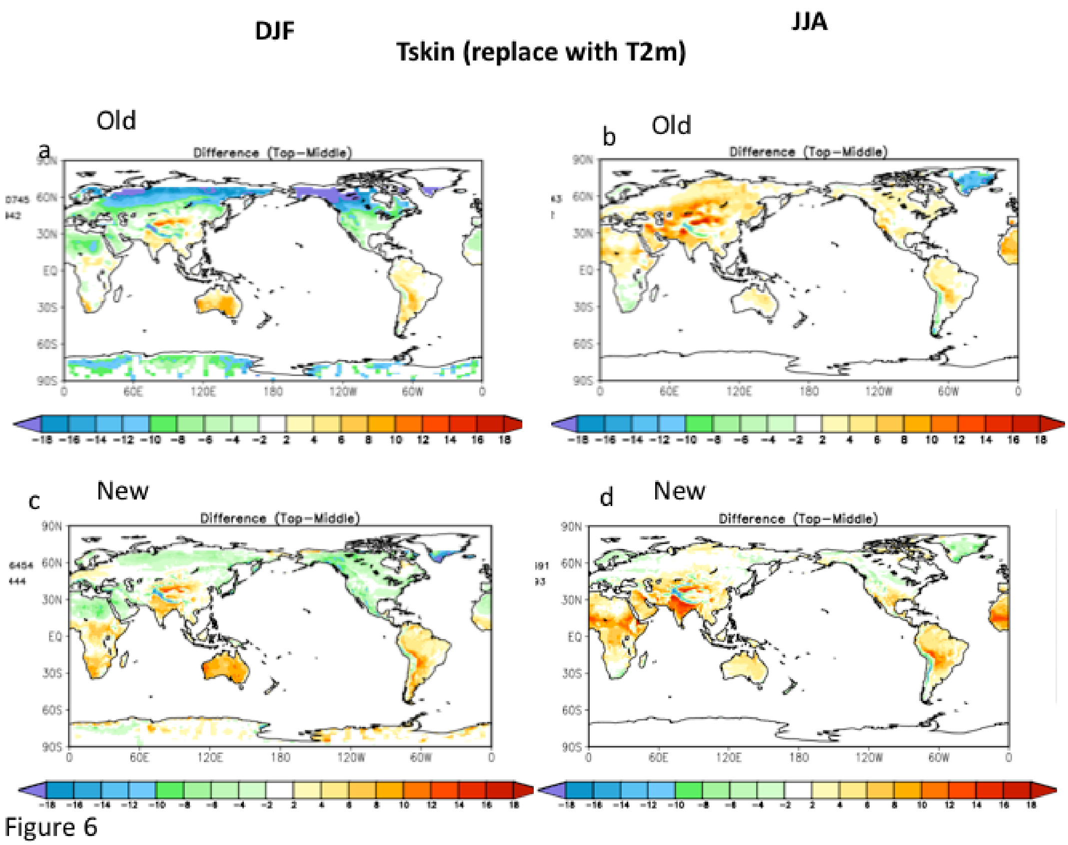


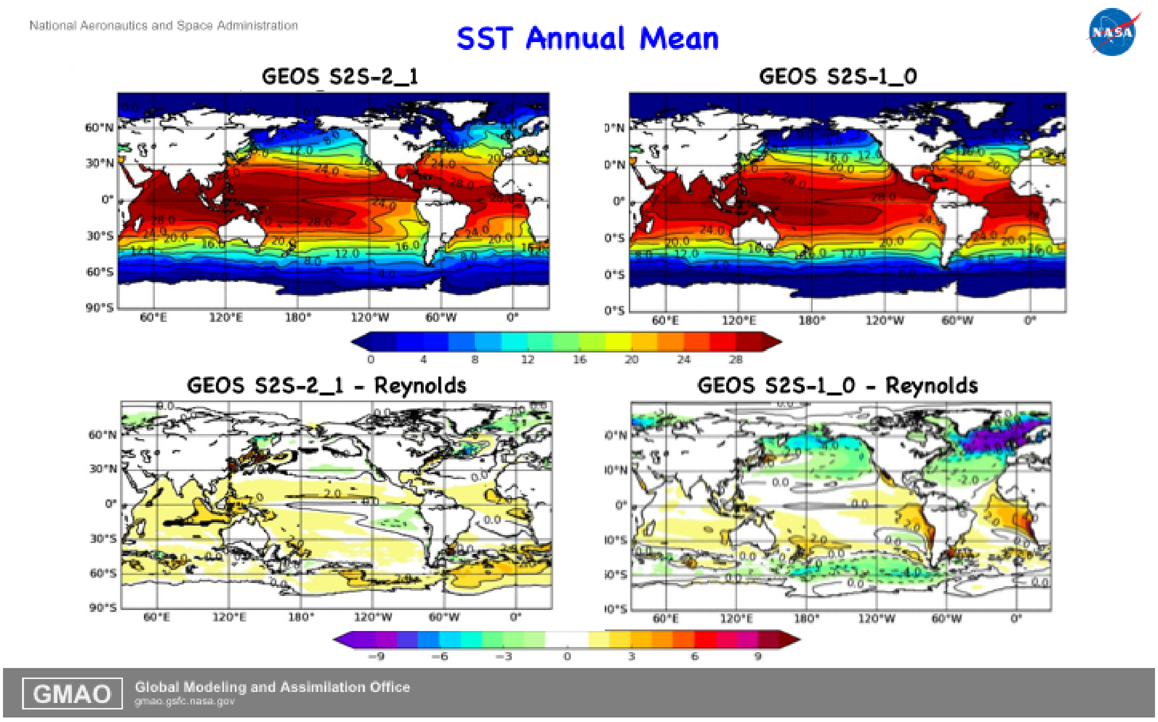
Figure 10. Mean Meridional Circulation....



496 **Figure 11.**



497 **Figure 12.**



509 **Figure 13.** GEOS S2S-2_1 and GEOS S2S-2_0 Sea surface temperature differenced from Reynolds analy-
510 sis

Equatorial Annual Cycle

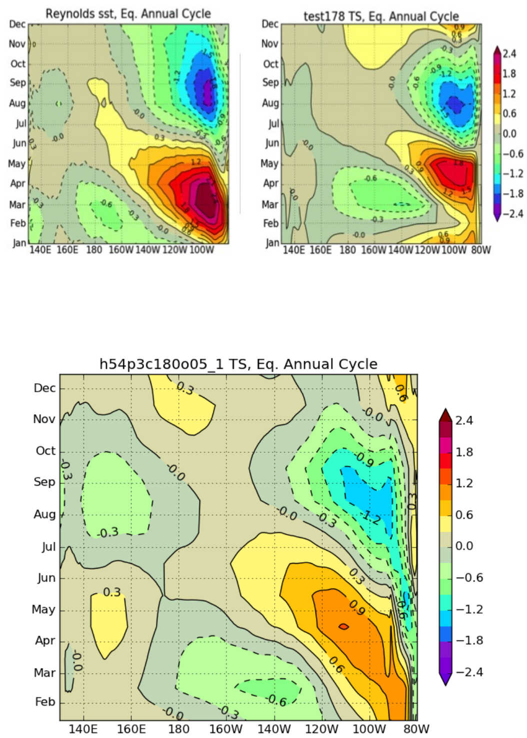
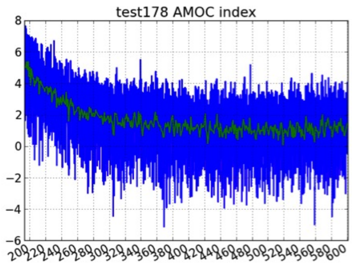


Figure 14.


test178 AMOC index
8
6
4
2
0
−2
−4
−6

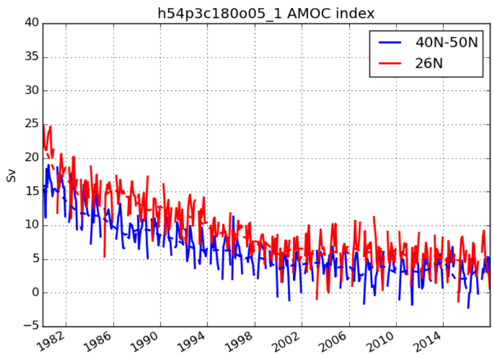

h54p3c180o05_1 AMOC index
40N-50N
26N
Sv
40
35
30
25
20
15
10
5
0
−5
1982
1986
1990
1994
1998
2002
2006
2010
2014

Figure 15.

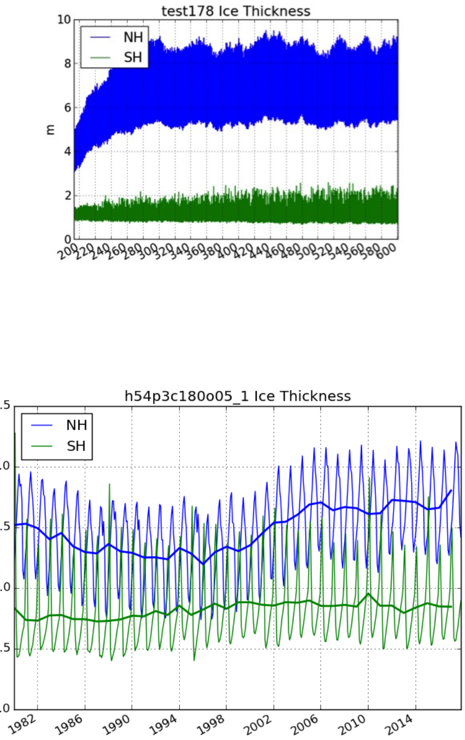
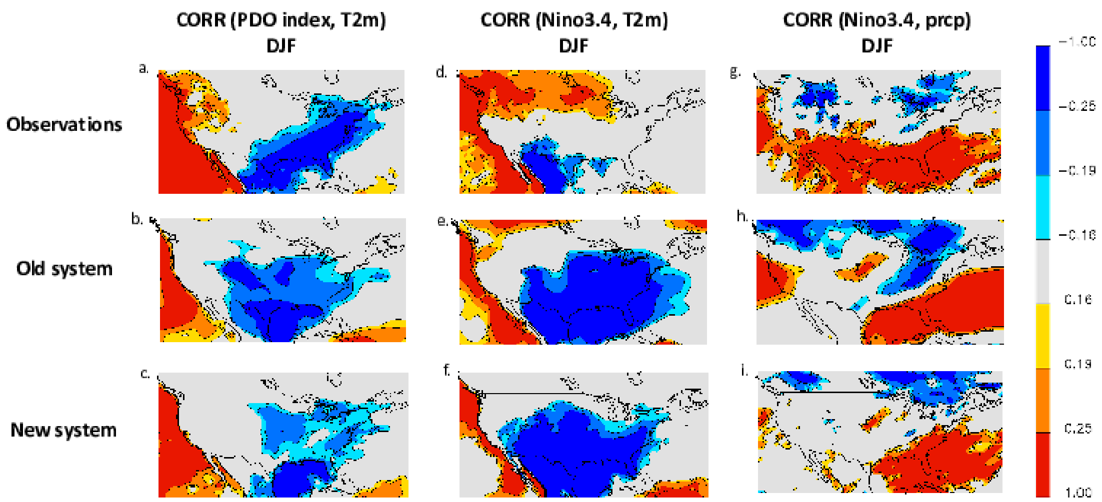
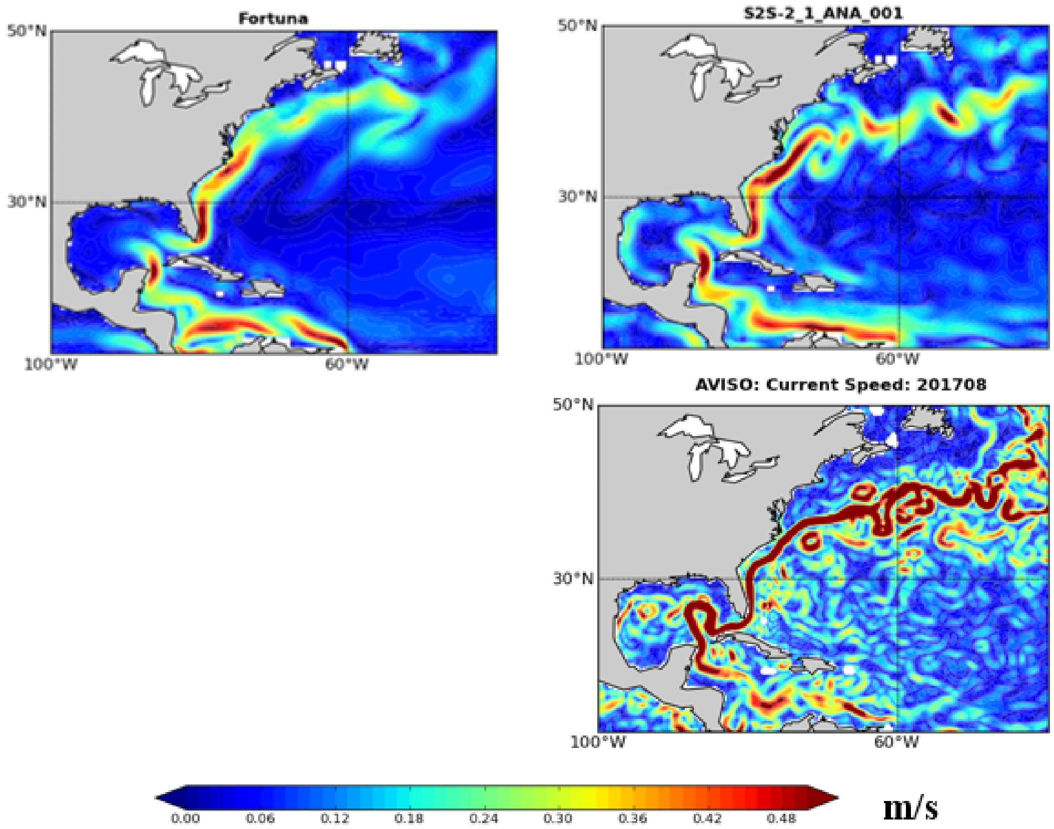

test178 Ice Thickness
NH
SH
m
h54p3c180o05_1 Ice Thickness
NH
SH
m
1982
1986
1990
1994
1998
2002
2006
2010
2014

Figure 16.



553 **Figure 17.** Schematic of the replay process to apply the ODAS. Horizontal blue (red) arrow indicates the
554 ocean forecast (analysis) ocean model execution. Vertical arrows represent observers taken every 6 hours.
555 Steps in the process are indicated by green circled numbers and defined in the key.



584 **Figure 18.** Current speed for western boundary currents for August 2017. Top left panel is the S2S-v1.0,
585 top right is for S2S-v2.1, and the bottom right is geostrophic currents from the AVISO multi-satellite product.

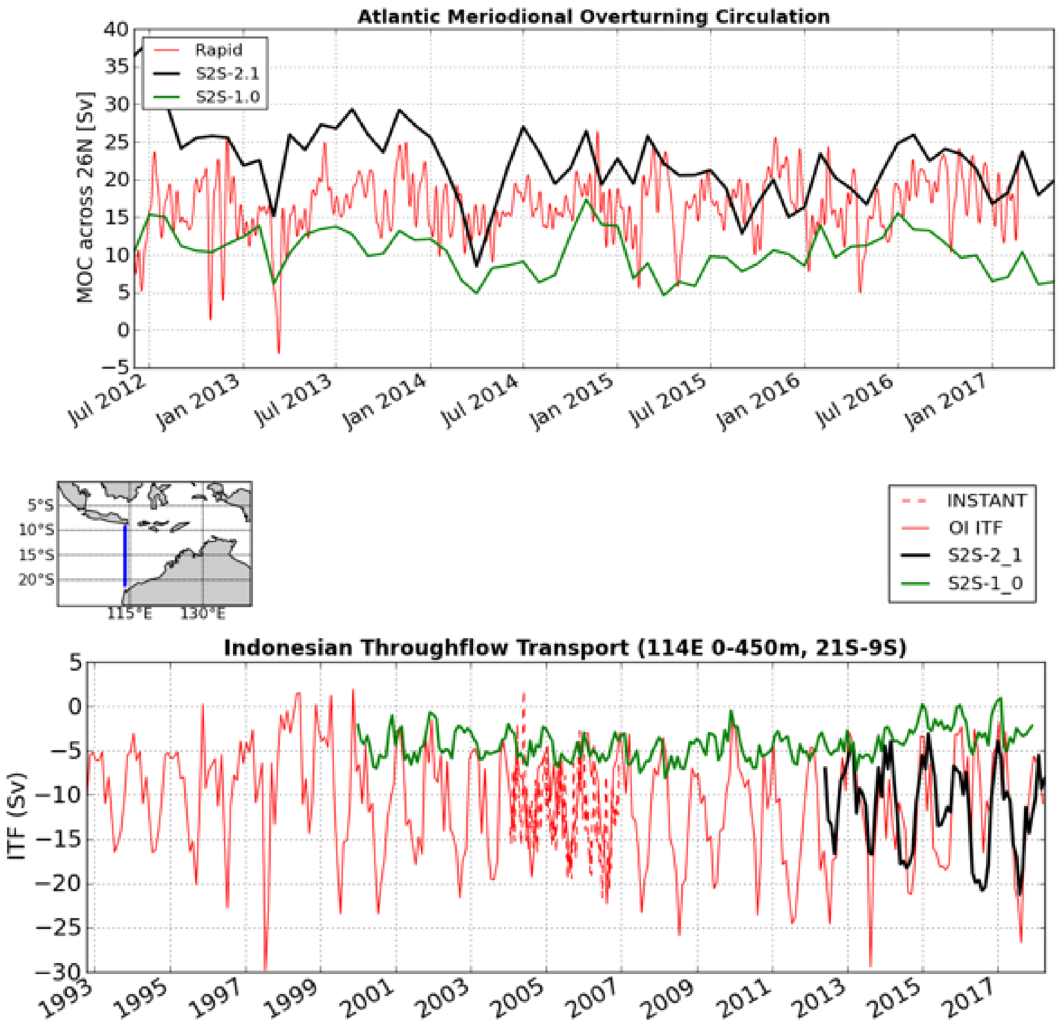


Figure 19. Major global heat conveyer belt indices, top) the Atlantic Meridional Overturning circulation (AMOC) is measured by in situ observations of the RAPID array (red) across 26.5oN in the Atlantic. S2S-v1.0 (green) and S2S-v2.1 (black) are compared over July 2012 until July 2017. The bottom panel shows indices of the Indonesian Throughflow (see inset for location). Geostrophic transport calculated using an optimal interpolation (*Carton* [1989]) of all available in situ temperature and salinity observations (solid red) compares well with in measurements from INSTANT moorings *Sprintall et al.* [2009] (dashed red). S2S-v1.0 (green line) clearly underestimates the ITF transport whereas S2S-v2.1 (black line0 corresponds well with observations.

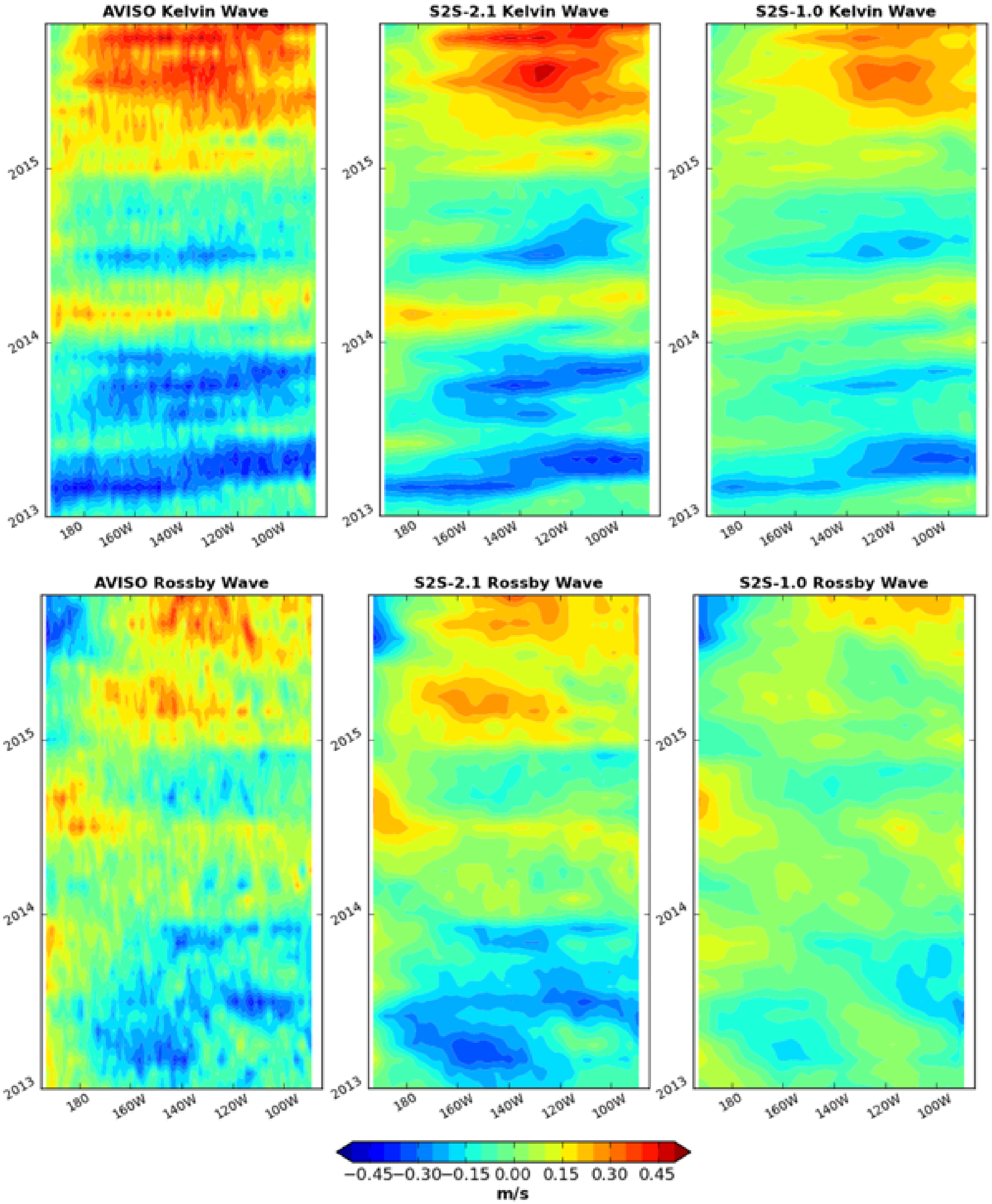
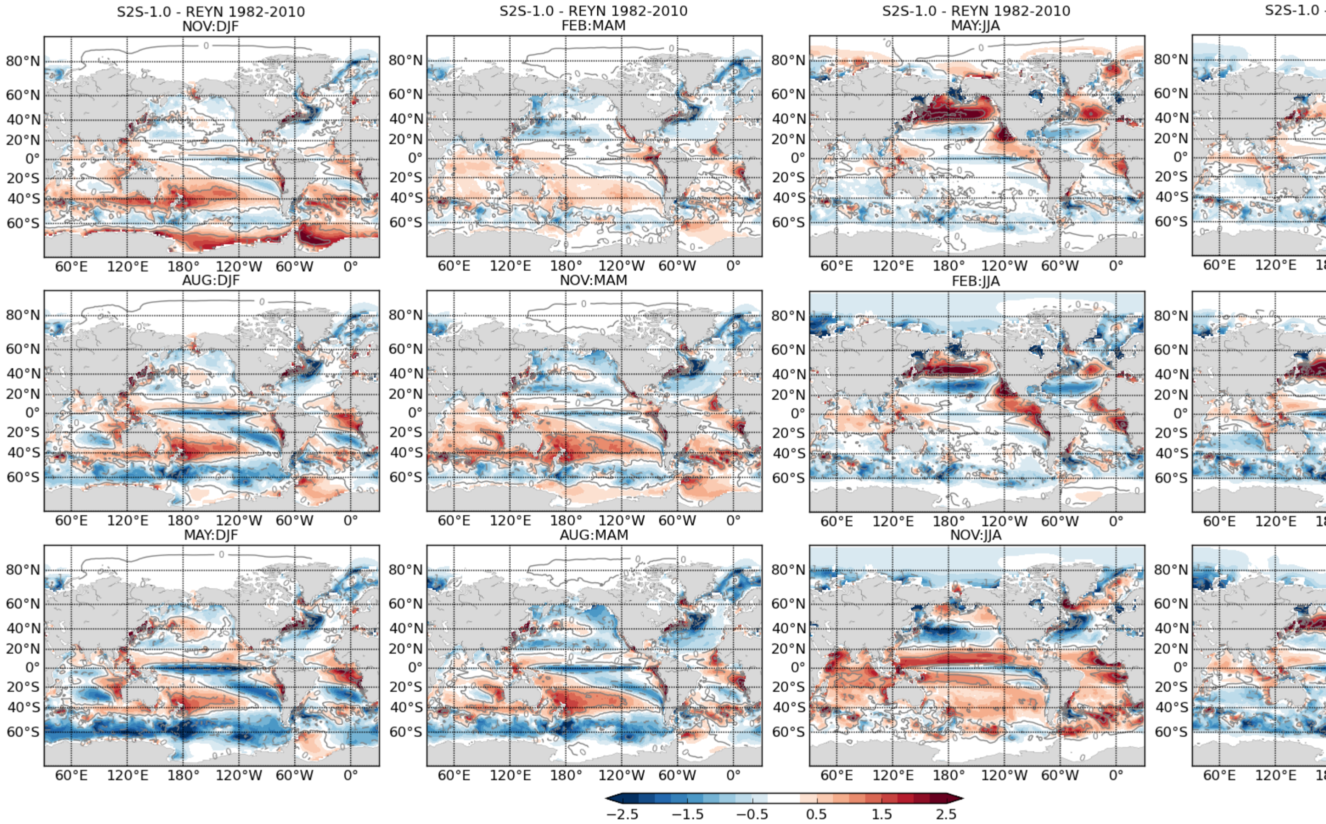
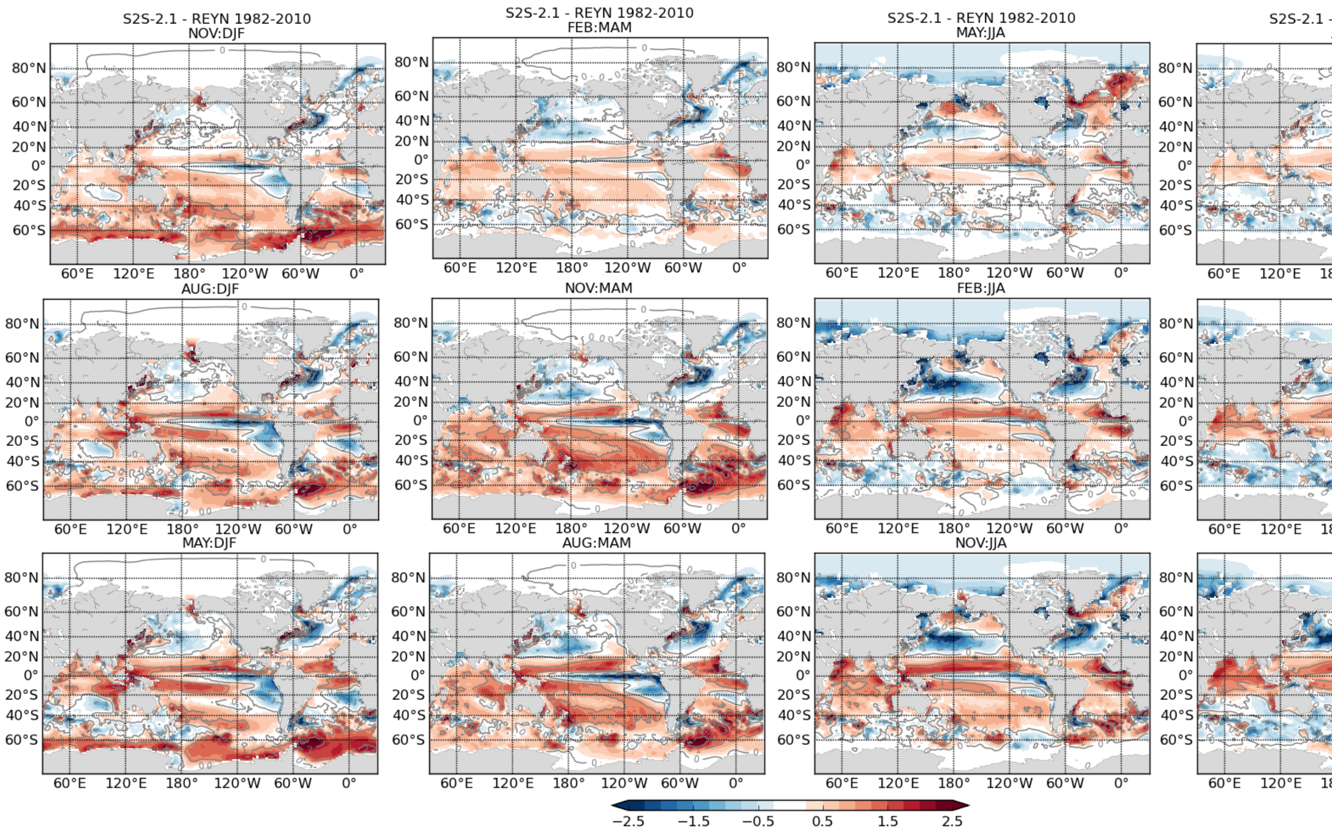


Figure 20. Longitude versus time distribution of the equatorial (top) Kelvin and (bottom) the first meridional mode of equatorial Rossby waves through their signature in zonal surface current deduced from the observed AVISO multi-satellite altimetry *AVISO* [2013] (left), S2S-v1.0 (middle) and S2S-v2.1 (right). Kelvin waves travel west-to-east and take about 3 months to transit the Pacific and Rossby waves travel from east-to-west and take about 8 months.



686 **Figure 21.** S2S-1.0 seasonal mean SST drift at 1, 3, 6 months leads.



687 **Figure 22.** S2S-2.1 seasonal mean SST drift at 1, 3, 6 months leads.

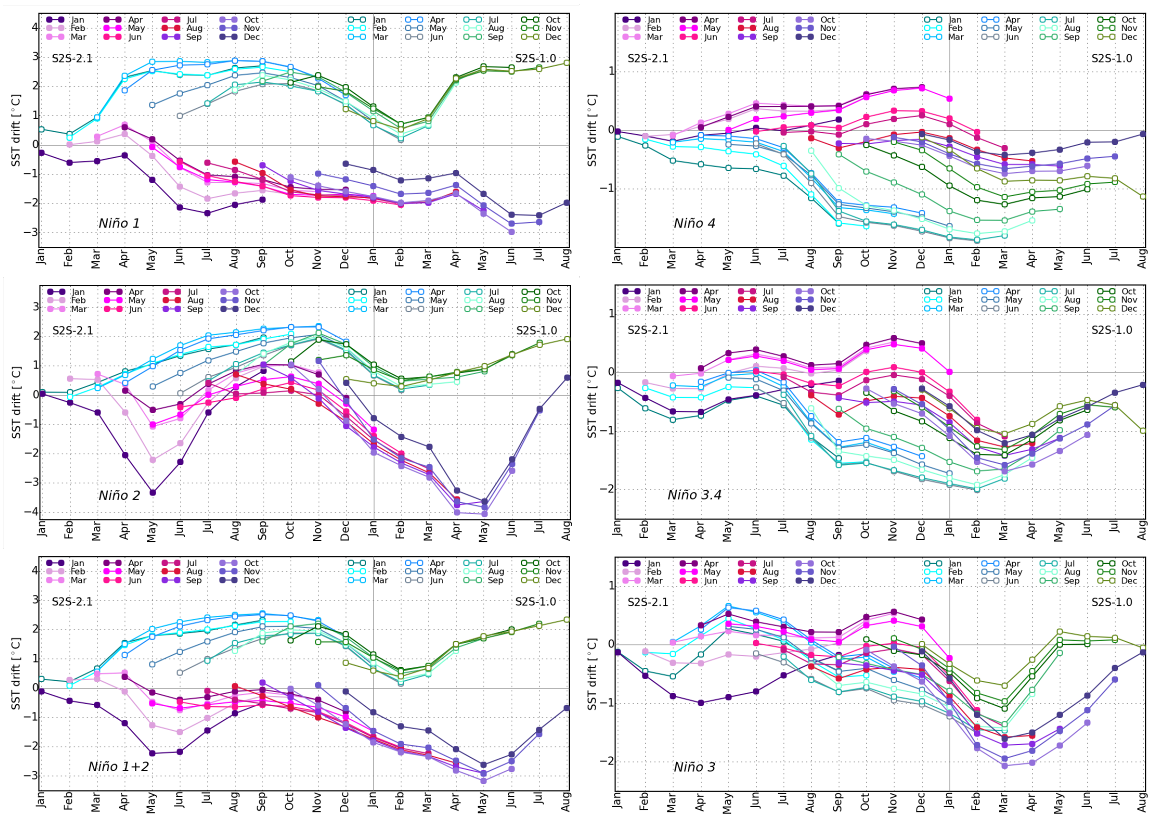


Figure 23. S2S-1.0 and S2S-2.1 monthly mean drift for ENSO indices with respect to Reynolds. Filled circles correspond to S2S-2.1, empty circles to S2S-1.0. The forecasts are color-coded by their initialization month with pink/purple colors for S2S_2.1 and green/blue for S2S_1.0.

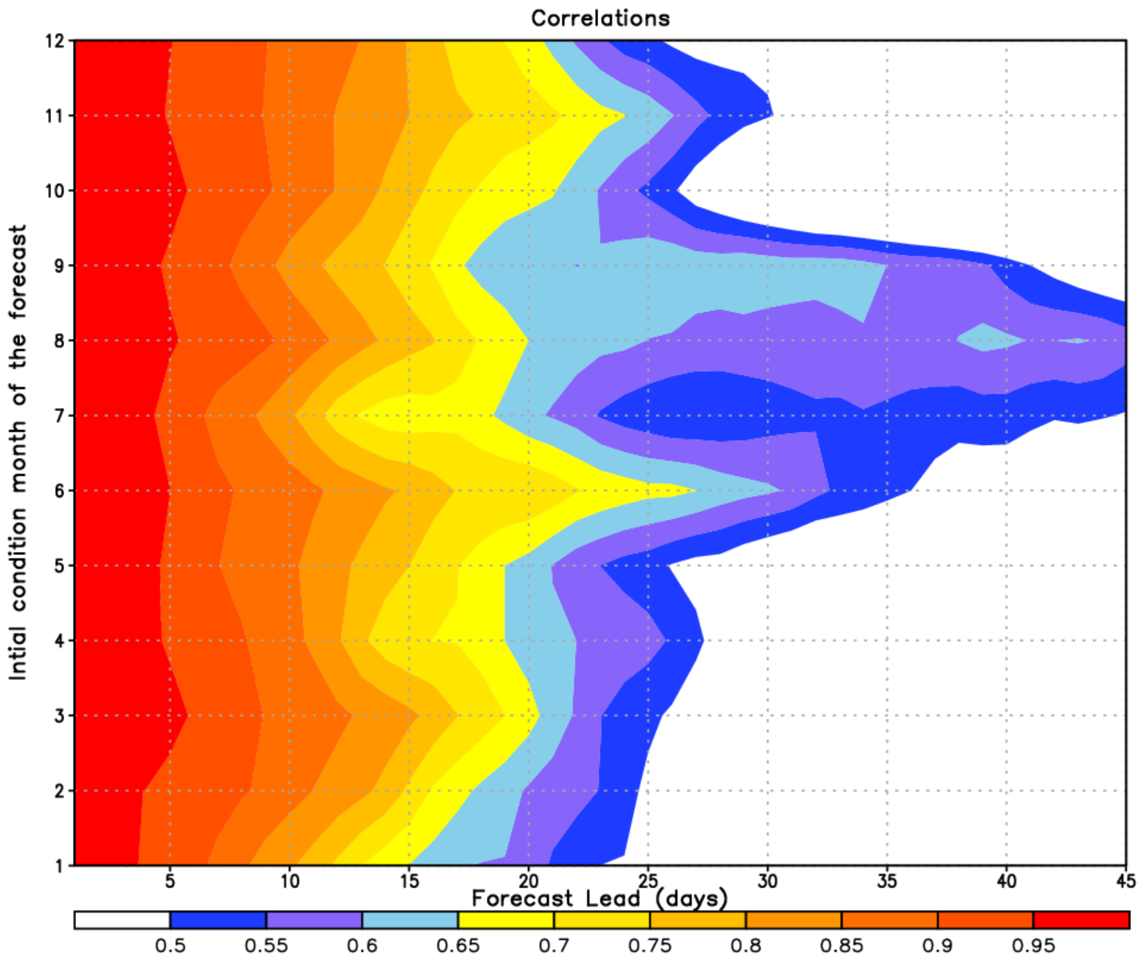


Figure 24. Prediction skill of the MJO in terms of the bivariate anomaly correlation of the MJO indices. They are calculated following [*Wheeler and Hendon*, 2004] and [*Gottschalck et al.*, 2010]. The retrospective forecasts are carried out for the period 1999-2016, with initial dates at every five days forecasting for 45 days. Prediction skill is assessed for the 45 day forecast periods, when the observed MJO is present. X-axis denotes the forecast lead day while y-axis is the initial condition month of the forecast.

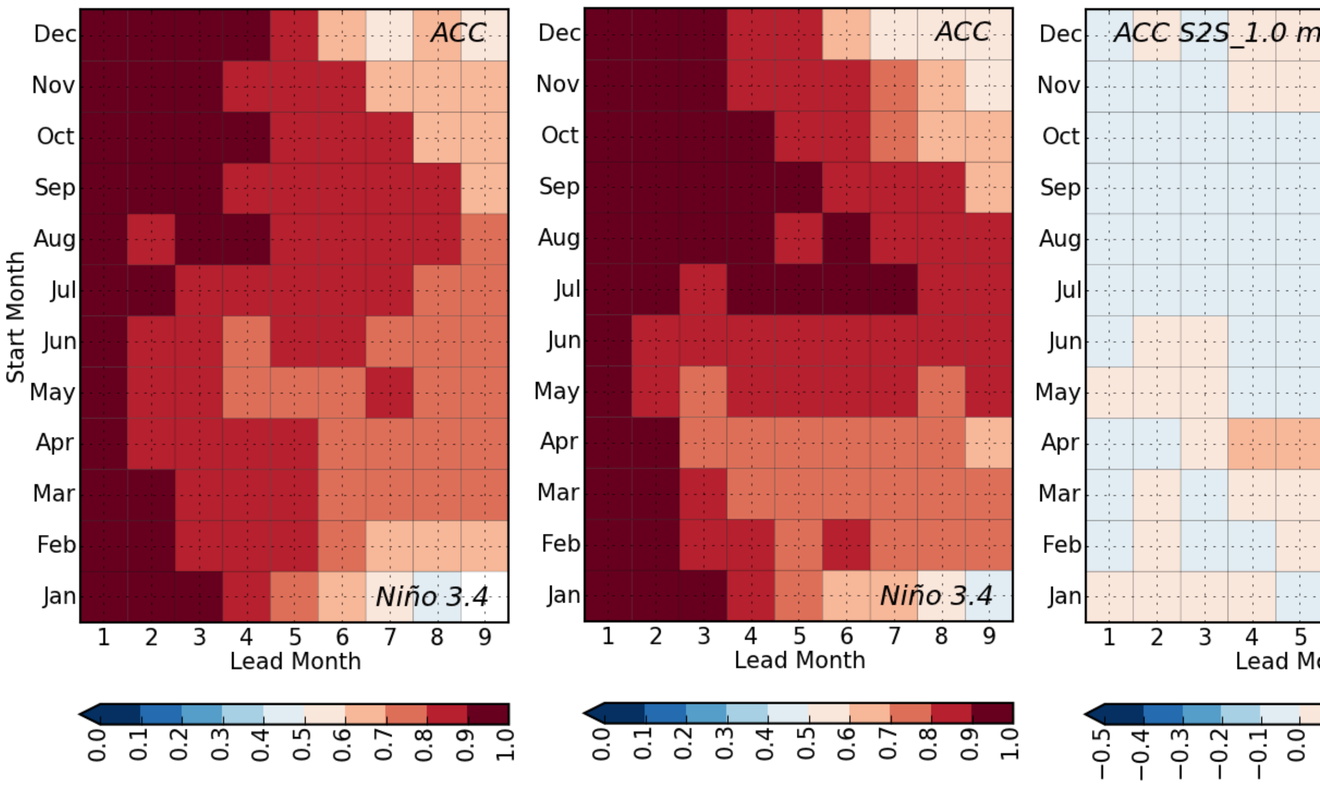
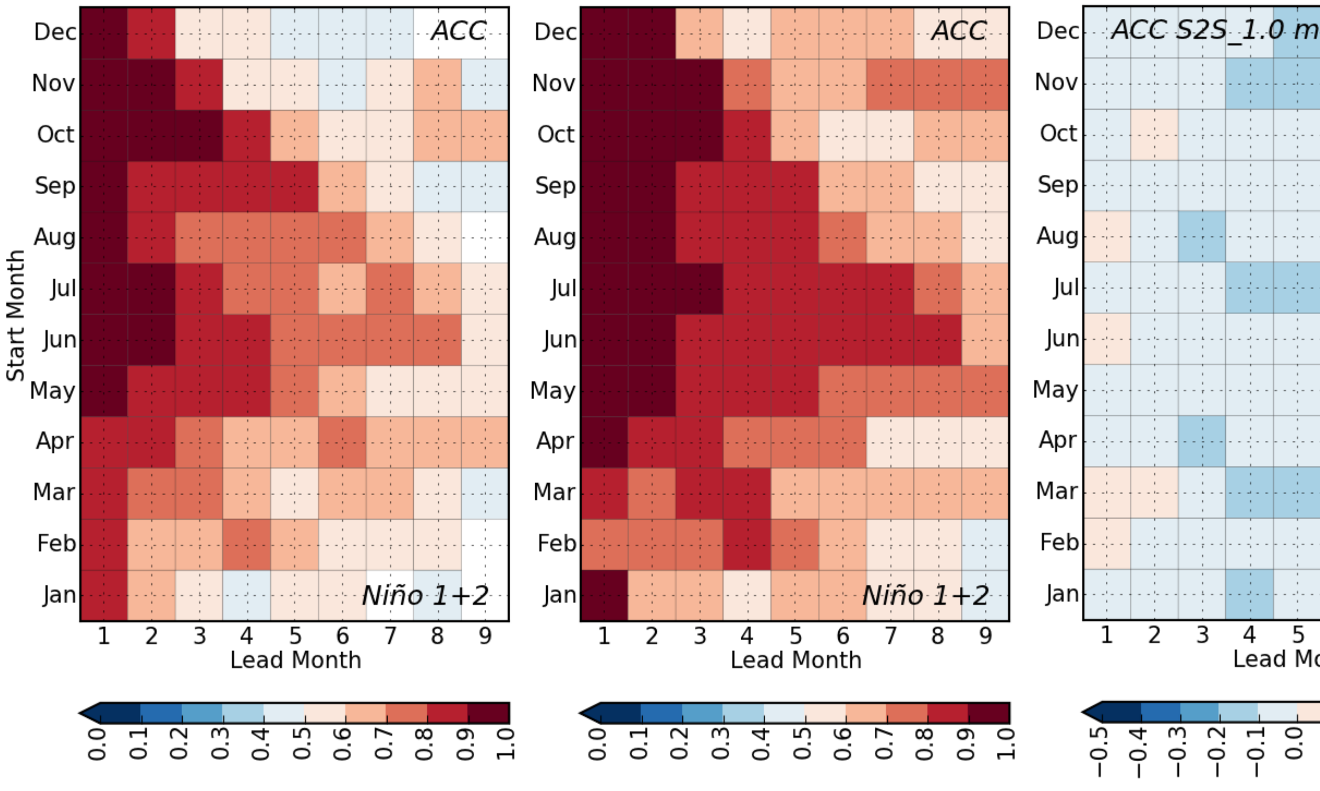
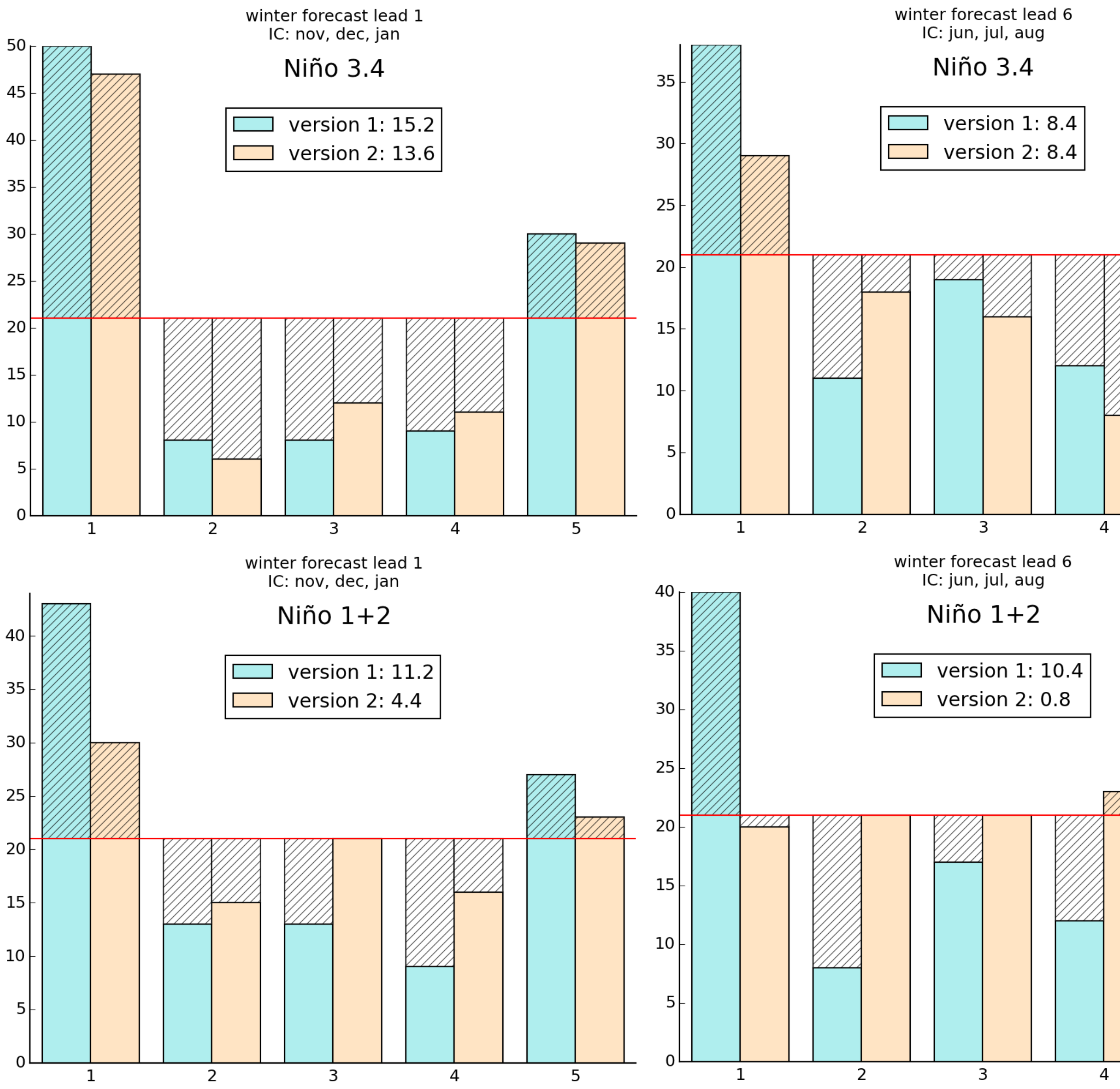


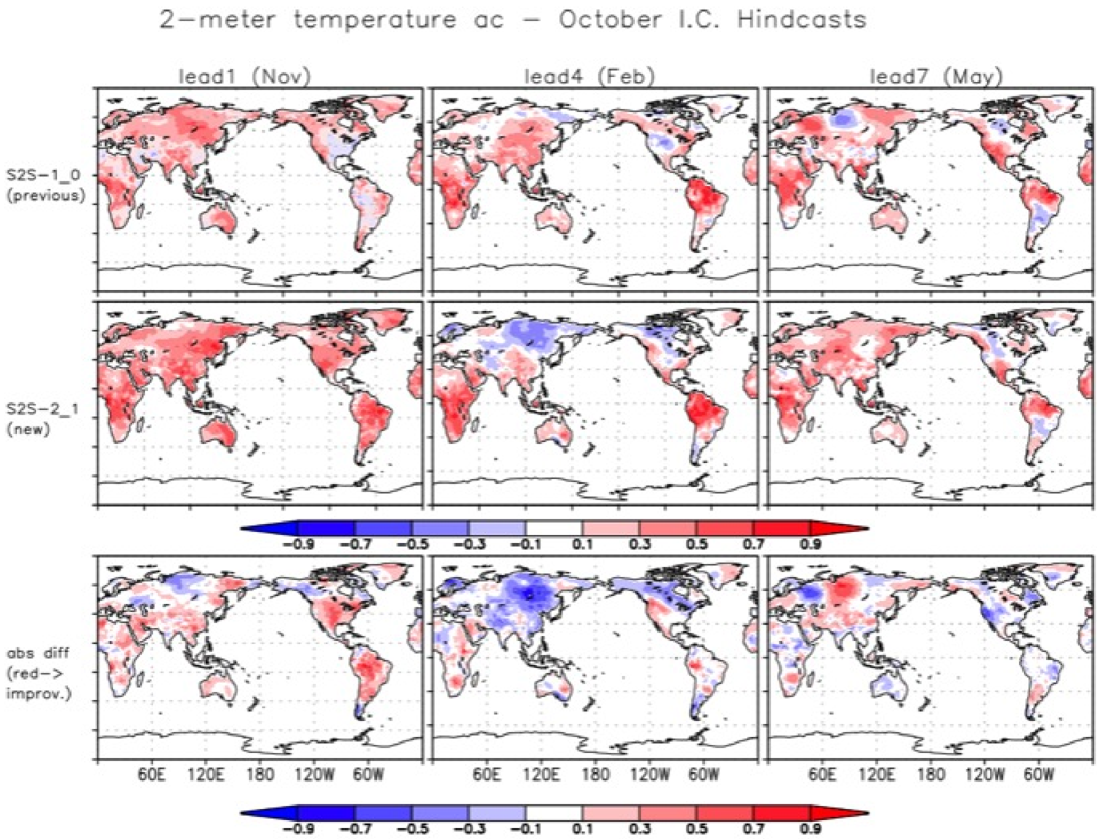
Figure 25. Anomaly correlation for Niño 3.4 SST index. Reynolds SST as observations. First forecast month is on the y-axis, lead time on the x-axis. Left panel S2S-1_0, middle S2S-2_1, right S2S-1_0 minus S2S-2_1.



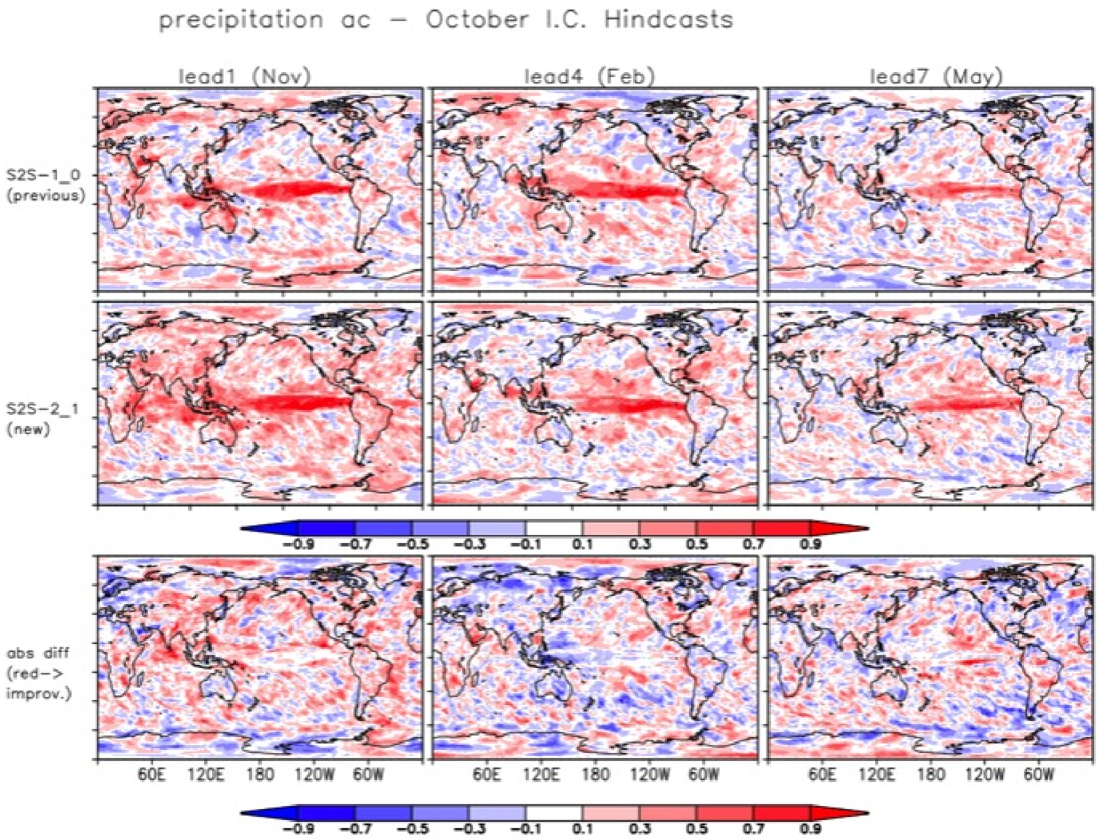
731 **Figure 26.** Same as 25, but for Nino1+2 index.



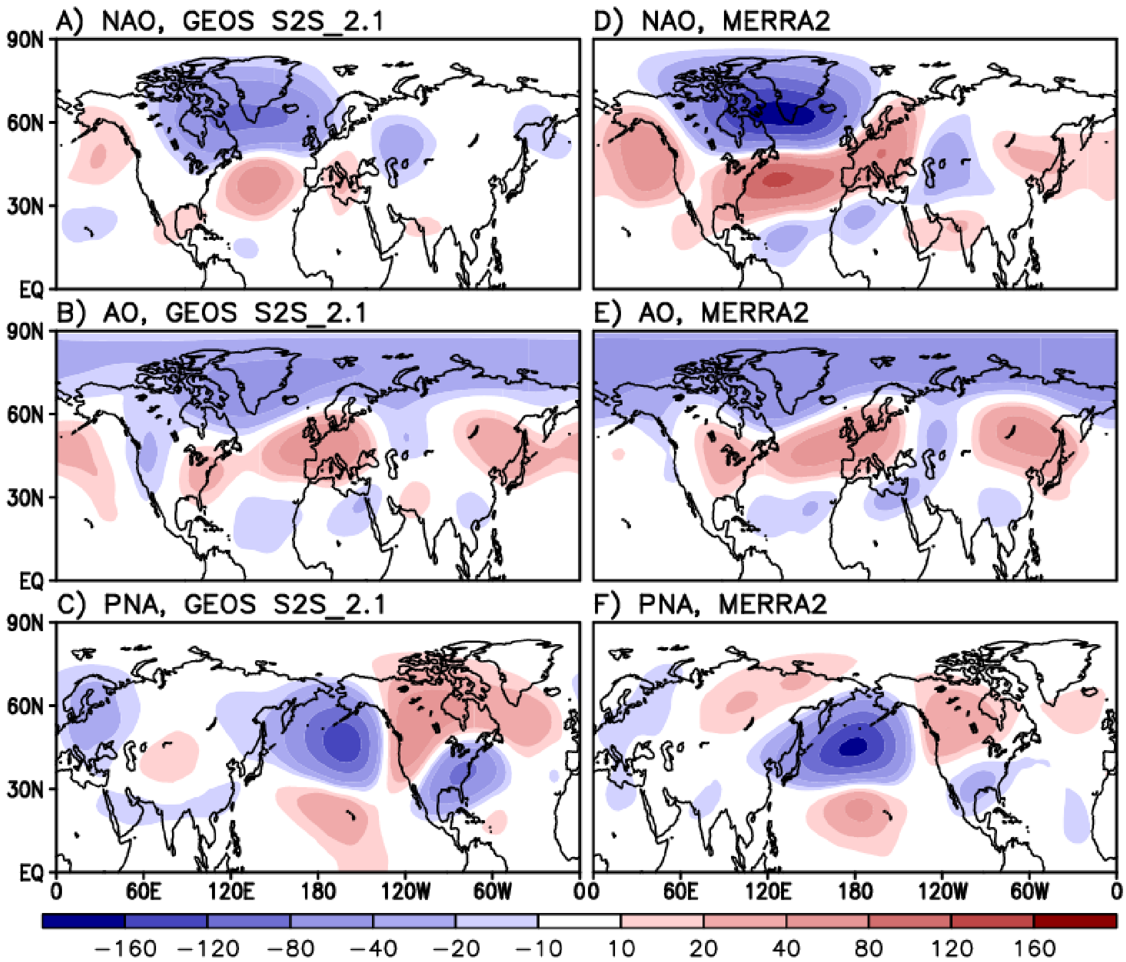
752 **Figure 27.** Rank histogram for winter (DJF) forecasts at lead 1 (IC:NDJ) and lead 6 (IC:JJA). S2S_1-0
753 shown in aqua, S2S_2-1 in tan color.



755 **Figure 28.** T2m skill in reforecasts



756 **Figure 29.** Prec skill in reforecasts



761 **Figure 30.** Major teleconnection patterns captured by GEOS S2S-2.1 and MERRA2

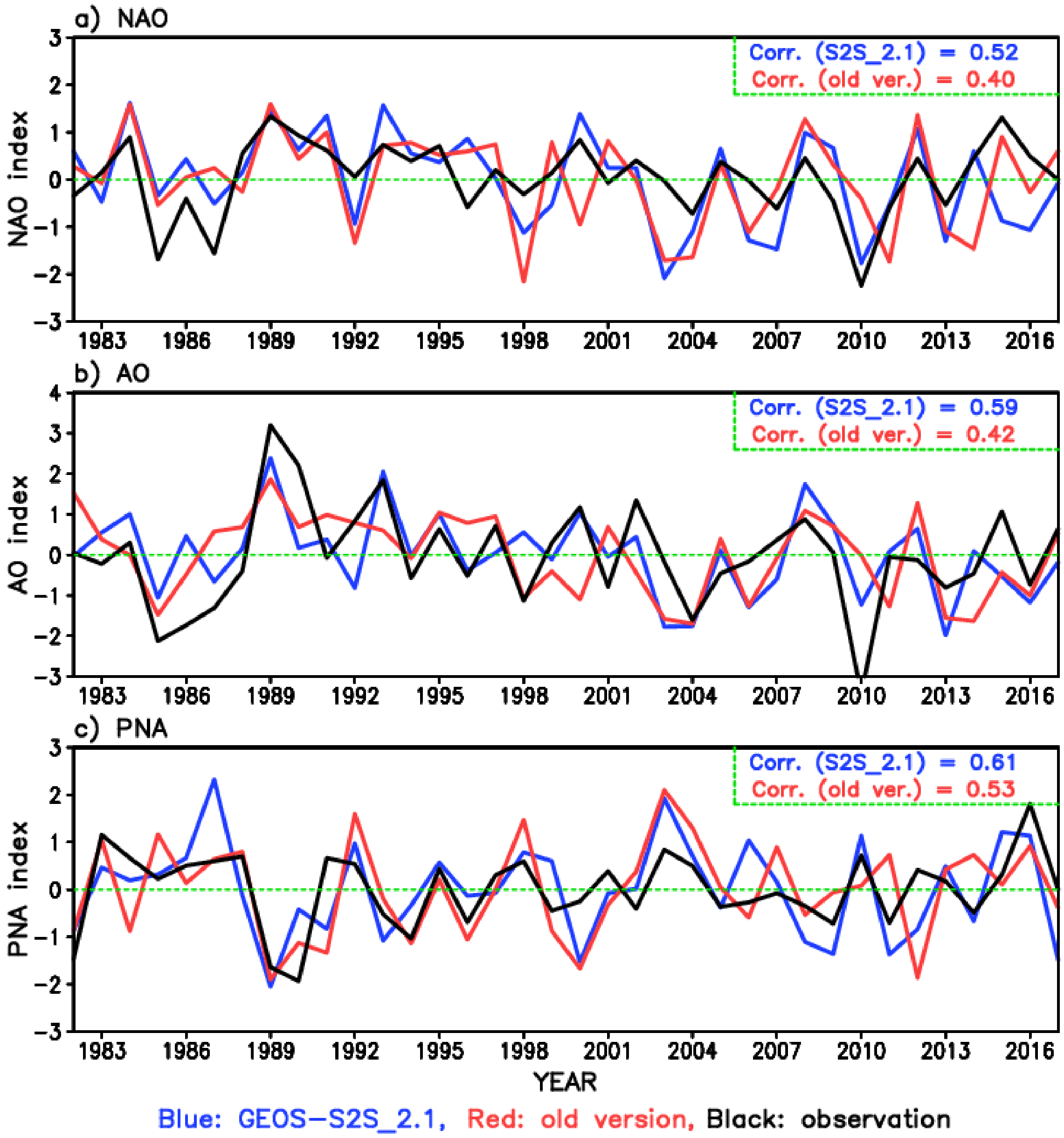


Figure 31. January/February mean NAO (top), AO (middle), and PNA (bottom) teleconnection indices predicted by GEOS S2S_2.1 (blue) and old model version (red) initialized on 27 December. Black line represents the observed teleconnection indices. Correlations between observation and GEOS S2S_2.1 (blue), and between observation and old model version (red), are, respectively, given on the upper-right corner of the each panel.

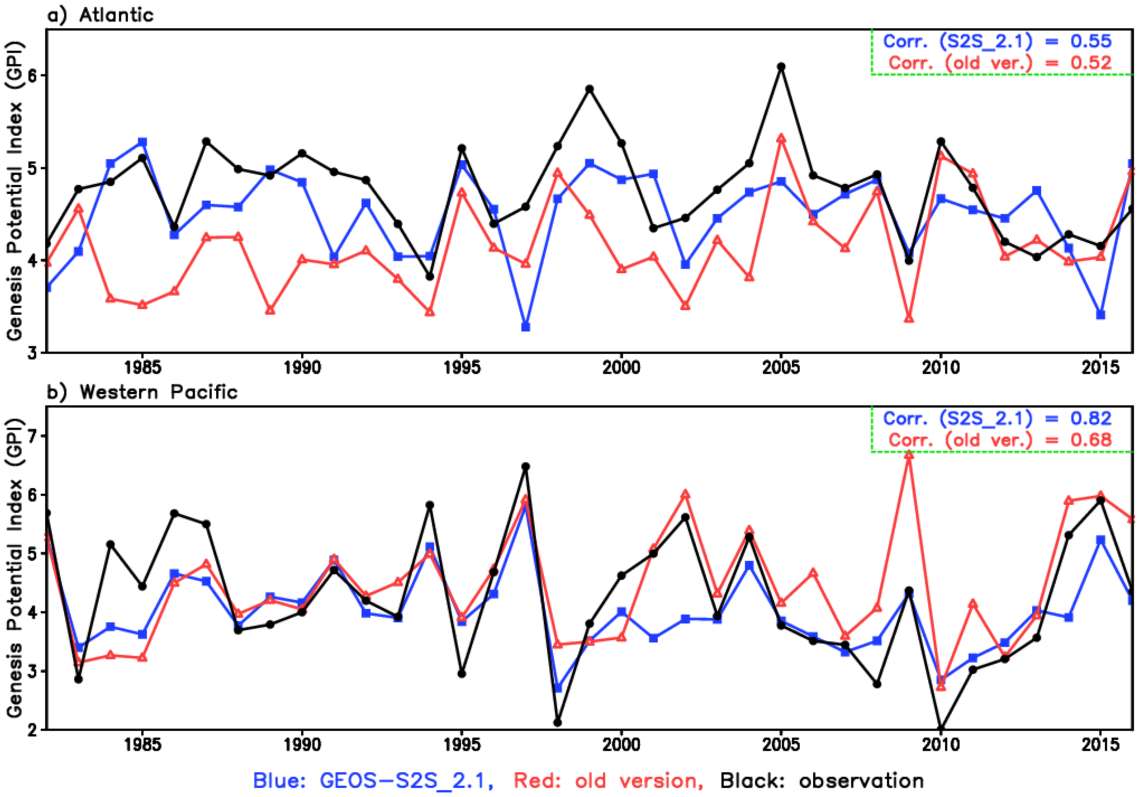


Figure 32. Predicted tropical cyclone activity in terms of the genesis potential index (GPI) initialized on 31 May. The predicted GPIs averaged over June/July/August/September (JJAS) each year are examined for the North Atlantic (upper) and the Western Pacific (bottom) region, respectively. Blue, red, and black solid lines denote the results from the S2S_2.1, old model version, and MERRA-2. Correlations between MERRA-2 and S2S_2.1 (blue), and between MERRA-2 and old model version are, respectively, given on the upper-right corner of the each panel.

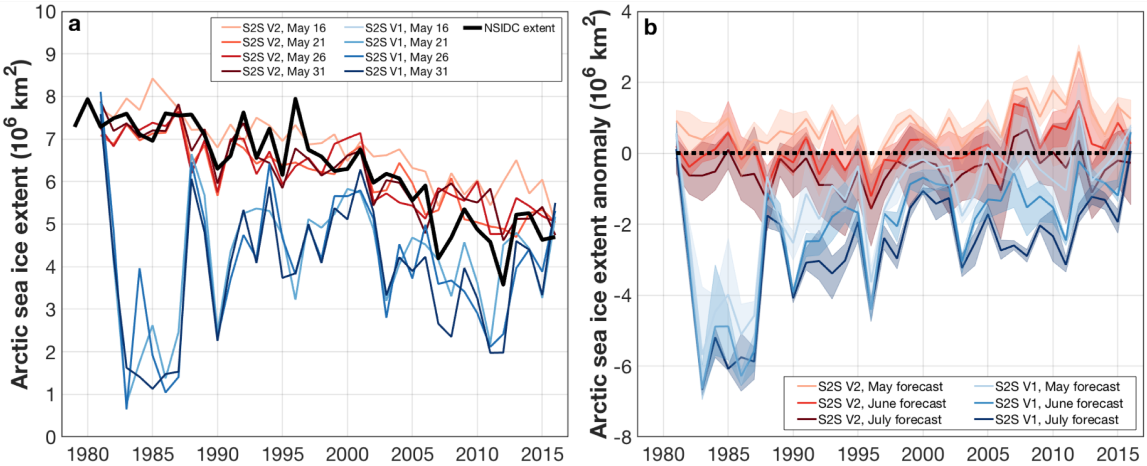


Figure 33. a. June retrospective forecast ensemble variability of September Northern Hemisphere sea ice extent for GEOS S2S Version 1 (blues) and Version 2 (reds) compared to sea ice concentrations derived from satellite brightness temperature (black, *Cavalieri et al.* [1996]). b. May, June, and July retrospective forecast anomalies of GEOS S2S Version 1 and Version 2 (as differenced from *Cavalieri et al.* [1996]) for September Northern Hemisphere sea ice extent. The ensemble spread for each forecast is indicated with the shaded bars.

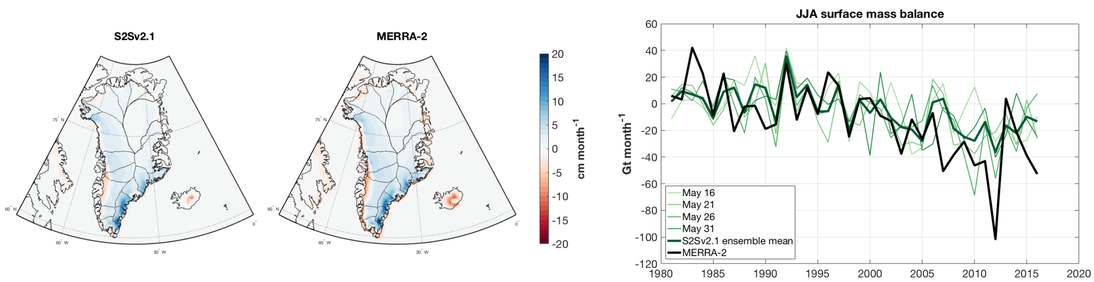


Figure 34. a. S2Sv2.1 mean one month lead ice sheet surface mass balance over the retrospective forecast period (1981-2016) b. Mean ice sheet surface mass balance (1981-2016) c. Greenland ice sheet summer (JJA) surface mass balance predicted from forecasts initialized in May (thin green lines), the ensemble mean (thick green line) and MERRA-2 (thick black line).

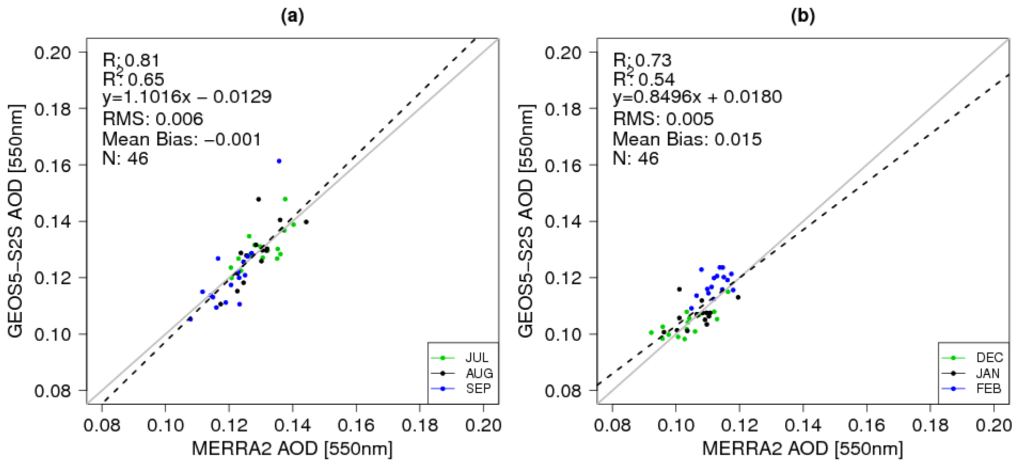
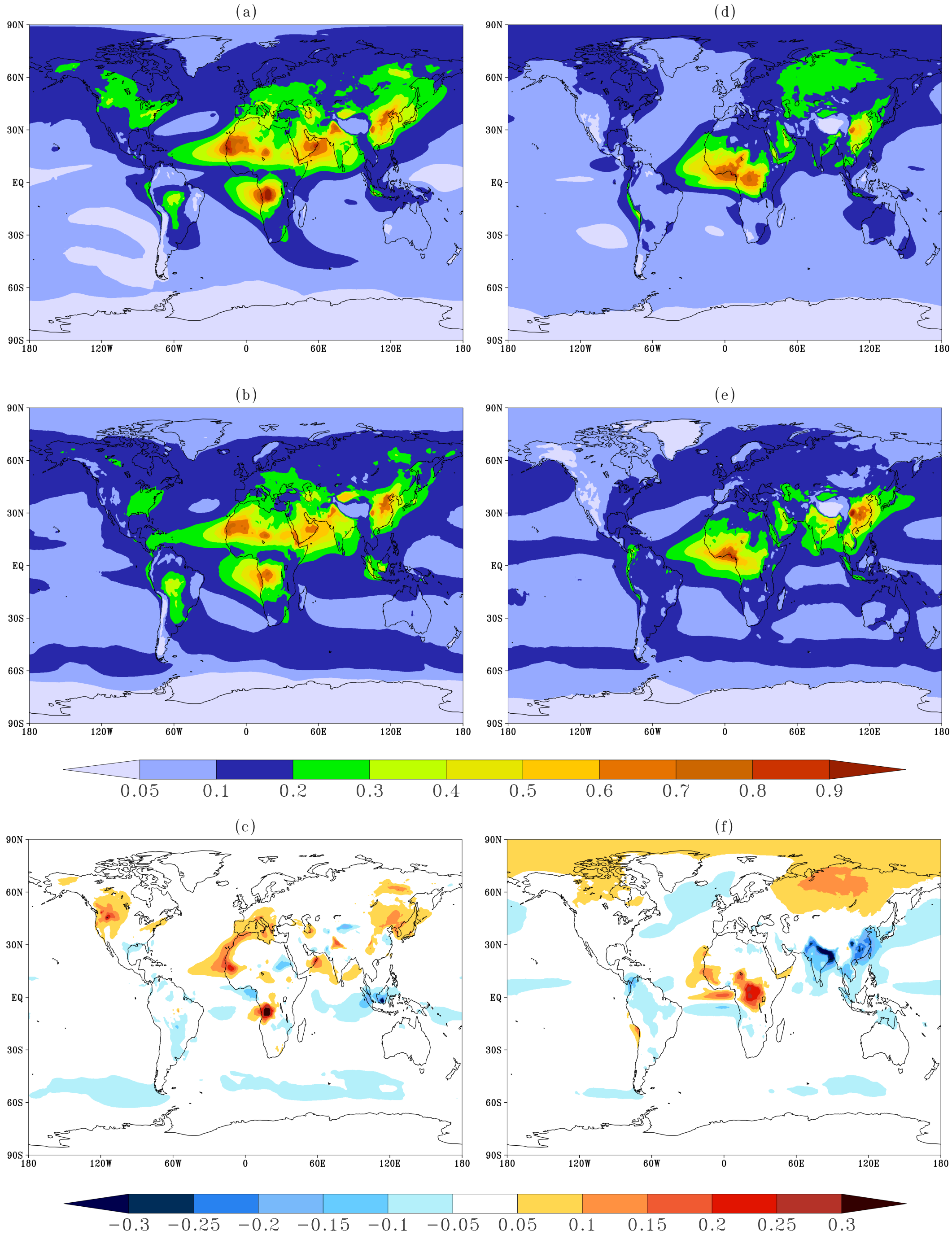


Figure 35. Monthly mean AOD globally averaged from GEOS5-S2S ensemble mean compared to MERRA2 for the period 2000- 2015 for (a) July-August-September, and (b) December-January-February. The colors indicate the months, the solid gray the 1-1 line, and the dashed regression linear line. Statistics are computed in natural log-transformed AOD space.



972 **Figure 36.** Spatial distribution of aerosol optical depth at 550nm in July-August-September (on the left)
973 and December-January-February (on the right) averaged for the period 2000-2015 from GEOS5-S2S ensem-
974 ble mean (top), reanalysis from MERRA2 (middle), and the GEOS5-S2S mean bias relative to MERRA2 (on
975 the bottom).

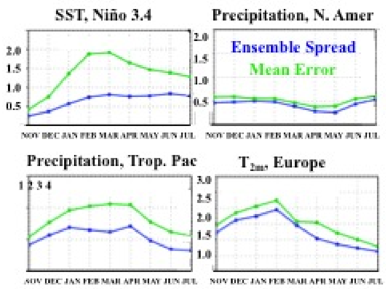

SST, Niño 3.4
Precipitation, N. Amer
Ensemble Spread
Mean Error
Precipitation, Trop. Pac
T2m, Europe
NOV DEC JAN FEB MAR APR MAY JUN JUL

Figure 37.