U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5830 University Research Court
College Park, MD 20740-3818

Office Note 510
https://doi.org/10.25923/knxm-kz26

Description and Results from UFS Coupled Prototypes for Future Global, Ensemble and Seasonal Forecasts at NCEP

Lydia Stefanova[1], Jessica Meixner[2], Jiande Wang[1], Sulagna Ray[3], Avichal Mehra[2], Michael Barlage[2], Lisa Bengtsson[4], Partha S. Bhattacharjee[1], Rainer Bleck[5], Arun Chawla[2], Benjamin W. Green[5,7], Jongil Han[2], Wei Li[1], Xu Li[1], Raffaele Montuoro[2], Shrinivas Moorthi[2], Cristiana Stan[6], Shan Sun[5], Denise Worthen[1], Fanglin Yang[2], Weizhong Zheng[1]

[1] IMSG at NOAA/NWS/NCEP/EMC, College Park, MD
[2] NOAA/NWS/NCEP/EMC, College Park, MD
[3] SRG at NOAA/NWS/NCEP/EMC, College Park, MD
[4] NOAA/ESRL/PSL, Boulder, CO
[5] NOAA/ESRL/GSL, Boulder, CO
[6] George Mason University, Fairfax, VA
[7] CIRES, University of Colorado, Boulder, CO

July 2022

Revised August 2022

E-mail: lydia.b.stefanova@noaa.gov

Table of Contents

Abstract

NOAA's Unified Forecast System (UFS) is a community-based modeling system developed to provide a common framework for all NCEP Production Suite operational forecast models. One of the UFS targets is a coupled model for global predictions of weather to seasonal time scales. Once completed, the coupled UFS will be a fully coupled atmosphere/land/ocean/sea ice/waves/aerosols system. The coupled UFS is built incrementally, by gradually adding necessary features and updating components. During the development process, discrete system prototypes are defined and evaluated in a standardized setting with consistent metrics. This note describes the sequence of coupled UFS prototypes developed to date and presents results from their evaluation.

Executive Summary

The next generation NWS/NCEP operational GFS and GEFS systems, GFSv17 (for medium-range deterministic predictions) and GEFSv13 (for medium-range to subseasonal ensemble predictions), as well as the future seasonal forecasting system, SFSv1, are being developed as fully coupled atmosphere/land/ocean/sea-ice/wave/aerosols UFS applications. The final coupled system will consist of the FV3 dynamical core and a CCPP-based atmospheric physics package, Noah MP land model, MOM6 ocean model, CICE6 sea ice model, WAVEWATCH III wave model, and GOCART aerosol model.

To date, seven coupled UFS prototypes (P1–P7) with increasing complexity, number of components, and component refinement have been constructed towards that target. Once a system prototype is finalized, it is tested to ensure that the configuration is stable. Deterministic forecast skill and biases are evaluated, and deficiencies needing developers' attention in subsequent iterations are identified for the aforementioned operational targets.

Some notable achievements of this new global coupled system demonstrated by these prototypes include:

- All benchmark RAW (*i.e.*, without bias correction) AC scores in weeks 1 and 2 are an improvement over CFSv2 in all prototypes with a few exceptions. Results are similar in nature for the SEC (*i.e.*, with bias correction) AC scores, showing improvement over CFSv2 in all prototypes (Section 12.2).

- In P2 and subsequent prototypes, OLR biases have been significantly reduced compared to P1 in some geographical regions, like the high latitudes (Section 12.3). The tropical pattern of biases remains in place and needs further consideration.

- For MJO skill, the gain in forecast skill lead time in all prototypes is on the order of 3–5 days for RMM1 and RMM2, and 4–7 days for the combined RMM when compared with operational CFSv2 (Section 12.4).

- In the latest full prototype (P7), the Arctic ice area between September and April is closer to NOAA/NSIDC-CDR while from May to August, the Arctic ice area is closer to the values from NASA-Team (Section 12.5).

- For the later prototypes, SST skill for Weeks 3&4 forecasts is the highest in the ENSO active region of tropical Pacific but the zonal wind stress is generally weak in all prototypes compared to ERA5 (Section 12.6).

- Upon adoption of the Noah-MP land surface scheme in P7, basin-scale water budget analysis over ten large river basins with diverse climate regimes shows improvement in eight of the ten basins (Section 12.7).

- In P7's sub-version P7.2 (the only configuration which included aerosols), aerosol optical depths (AOD) forecasts initialized in four different seasons of 2012 and 2013 show promising results when validated against MERRA-2 (Section 12.8).

But significant challenges remain—for example, obtaining skillful Weeks 3&4 forecasts for CONUS precipitation remains elusive (Section 12.2)—which must be addressed by future prototypes/configurations. Considerable modifications to the atmospheric physics and land surface parameterization in the most recent prototype (P7, Section 11.1) have brought about some skill reduction and exacerbation of model biases. Since P7 is the first instance of bringing these modifications together into a single system, this is not an unexpected initial result. Analysis

of the biases in P7 provides useful guidance for necessary ongoing parameterization adjustments in the next prototype, P8.

The progression from prototype 1 to prototype 7 is described in this technical report. Throughout the development process, numerous technical and science challenges have been met. The outcome of this undertaking is a six-way global coupled system for weather to sub-seasonal scale predictions which is a first for NOAA/NWS, and is the first development of such a system under UFS.

1. Background

Over the last few years, NOAA has begun transitioning its suite of operational applications to the UFS framework (Tallapragada, 2018). The UFS (https://ufscommunity.org/) is a community-based coupled comprehensive Earth modeling system with the capability of configuring a set of common components (models for atmosphere, land, ocean, sea ice, waves, aerosols, *etc.*) into different UFS applications targeting predictions that span local to global domains, and forecast timescales ranging from sub-hourly to seasonal.

At present (August 2022), the NWS/NCEP Environmental Modeling Center (EMC) uses the Global Forecast System (GFS) for medium-range (MWR) global high-resolution deterministic weather prediction; the Global Ensemble Forecast System (GEFS) for weather to subseasonal (out to 35 days) global ensemble prediction; and the Climate Forecast System (CFS) for subseasonal to seasonal (S2S) and seasonal global ensemble prediction.

The current GFS (GFSv16; NOAA, 2021) was implemented in operations in March 2021 and replaced the previous version, GFSv15 (NOAA, 2019). GFSv15 was the first of NOAA's GFS forecasting systems to use FV3 as its dynamical core, as a move towards the UFS framework. The current GEFS is GEFSv12 (NOAA, 2020; Zhou *et al.*, 2022), implemented in operations in September 2020. GEFSv12 is the first version of the NOAAs GEFS to be built within the UFS framework, albeit without coupling to ocean or sea ice components. The ocean and sea ice boundary conditions in GEFSv12 are derived from the current Climate Forecast System version 2 (CFSv2; Saha *et al*. 2014), which has been in NCEP operations since March 2011.

Neither GFSv16 nor GEFSv12 contain ocean or sea ice model components. Both are run with one-way coupling from atmosphere to waves, and one member of the GEFSv12 ensemble is run with one-way coupling from atmosphere to aerosols. However, since the coupling is one-way, no feedback from waves or aerosols is returned to the atmosphere. CFSv2, on the other hand, is a fully coupled atmosphere-ocean-ice system.

Coupled atmosphere-ocean-ice models have long been the norm for seasonal forecasting and climate prediction, but not necessarily for shorter time scales. However, more recent research and applications in a few leading international NWP centers have shown that even on the short- to medium-range weather timescale, the coupled interactions between the atmosphere, ocean, wave and sea ice can provide additional predictive skill (Brassington *et al.,* 2015; Mogensen *et*

al., 2017; Smith *et al.*, 2018; Velinga *et al.* 2020; Gentile *et al.,* 2021).

The next generation GFS and GEFS systems, GFSv17 and GEFSv13, as well as the future seasonal forecasting system (SFSv1), are envisioned as fully coupled atmosphere/land/ocean/sea-ice/wave/aerosols applications of the UFS, using the FV3 dynamical core and a CCPP-based atmospheric physics package with Noah-MP land model, MOM6 ocean model, CICE6 sea ice model, WAVEWATCH III wave model, and GOCART aerosol model.

In addition to developing the coupled model itself, the operational systems will need a data assimilation (DA) system to determine the initial conditions, and a long-term (minimum 30 years) reanalysis and reforecast climatology to provide the means for post-processing calibration and validation necessary for probabilistic and deterministic products at longer timescales. The specific targets for GFSv17, GEFSv13, and SFSv1 are listed in Table 1.1.

Table 1.1 Targets for the next-generation operational systems GFSv17, GEFSv13, and SFSv1

System (Target Implementation Fiscal Year)	**Coupled Model Components**	**Physics**	**DA**	**Reanalysis**	**Reforecast**
GFSv17 (FY 2024)	FV3 MOM6 CICE6 WW3 NOAH-MP GOCART	Advanced Physics	Weakly Coupled DA	N/A	N/A
GEFSv13 (FY 2024)				Yes	Subseasonal Ensemble Reforecast
SFSv1 (FY 2026)		Physics for Seasonal Scales	Coupled DA		Seasonal Ensemble Reforecast

The development of these forecasting systems is coordinated with the NOAA-supported collaborative multi-institutional Research to Operations (UFS-R2O) project. Further information on UFS-R2O is available at https://vlab.noaa.gov/web/ufs-r2o.

This note describes the effort to date in configuring the coupled UFS towards GFSv17, GEFSv13, and SFSv1 via constructing and running a series of incremental prototypes and evaluating forecast quality and skill throughout the process. The structure of the note is as follows. The general prototype definition is briefly described in Section 2. The common testing setup for running the prototypes is described in Section 3. The methodology used for evaluation of the prototype runs are described in Section 4. Each of Sections 5–11 is dedicated to a separate

system prototype and describes its configuration and evaluation. The results from the evaluation of all prototypes to date are summarized in Section 12. Finally, Section 13 provides conclusions and future directions.

2. Global Coupled UFS Prototypes: General Description

The development of the UFS coupled modeling system is an evolving process. Over the course of the system development, individual components are upgraded, changed, and/or added. The coupled UFS prototypes represent discrete snapshots of the system. To date, seven major prototypes have been completed and tested. Each prototype is specified by the set of modeling components and the coupling infrastructure, parameter settings, physics choices and initial conditions. The most recent iteration of the system used here employs the seasonal to subseasonal with waves and aerosols (S2SWA) application of the "ufs-weather-model" (see https://ufs-weather-model.readthedocs.io), coupling an atmospheric, ocean, ice, wave, and aerosol component models, and can be found on GitHub at https://github.com/ufs-community/ufs-weather-model.

The initial five prototypes (P1, P2, P3.1, P4, and P5) were focused on engineering aspects of developing the coupled system (changes to the modeling components and coupling infrastructure) and modifications to the initial conditions. Physics settings were kept unchanged as much as possible, except when modifications were necessitated by code updates and bug fixes, or to ensure model stability. Starting with P6, and continuing with P7, substantive changes in model physics began being introduced in addition to ongoing code updates and bug fixes.

The components comprising each prototype are summarized in Table 2.1. The initial conditions for each prototype's components are specified in Table 2.2. Note that in some earlier documentation the source of FV3 initial conditions for P1–P6, CICE5 initial conditions for P1–P2, and the MOM6 initial conditions for P1 has been listed as CFSR. In this table, it is specified more accurately as CDAS1, an extension of CFSR past 2011.

In addition to the prototypes listed in these tables, a sub-version of P7 (P7.2) was configured with the addition of a one-way coupled atmospheric chemistry and aerosols component, and tested with a limited number of simulations.

In all coupled UFS prototypes, the land component is an internal physical parameterization of the atmospheric model. There is no active lake ice model in any of the coupled UFS prototypes; the specification of lake ice values is summarized in Table 2.3.

Table 2.1 Model components in coupled UFS prototypes P1–P7. Additional details provided in each prototype's dedicated section.

Prototype	Atmospheric Model C384 (~0.25 degree) horizontal resolution			Ocean Model Tripolar ~0.25 degree horizontal resolution	Wave Model Regular lat/lon 0.5 degree grid	Ice Model Tripolar ~0.25 degree horizontal resolution	Mediator
	Dynamical Model **Vertical Resolution** **Fractional Grid**	**Physics Settings** **Physics Driver**	**Land Model**				
P1	FV3 64 layers Nonfractional grid	GFSv15.2 (see Section 5.1) IPD driver	Noah LSM	MOM6	*None*	CICE5	NEMS
P2							
P3.1							
P4		GFSv15.2 CCPP driver			WW3		
P5						CICE6 (Mushy TD not turned on)	CMEPS
P6	FV3 127 layers Fractional grid	GFSv16 (see Section 10.1.2) CCPP driver					
P7		Additional modification to GFSv16 (see Section 11.1.2) CCPP driver	Noah-MP LSM			CICE6 (Mushy TD turned on)	

Table 2.2 Initial conditions specification by component for coupled UFS prototypes P1–P7.

<table>
<tr><th rowspan="3">Prototype</th><th colspan="6">Initial Conditions Source</th></tr>
<tr><th colspan="3">FV3 GFS</th><th rowspan="2">MOM6</th><th rowspan="2">CICE</th><th rowspan="2">WW3</th></tr>
<tr><th>Atm</th><th>Soil</th><th>Snow</th></tr>
<tr><td>P1</td><td rowspan="5">CDAS1</td><td rowspan="6">CDAS1</td><td rowspan="6">CDAS1</td><td>CDAS1/MOM4</td><td>CDAS1/SIS1</td><td rowspan="2">n/a</td></tr>
<tr><td>P2</td><td rowspan="5">CPC 3Dvar</td><td rowspan="2"></td></tr>
<tr><td>P3.1</td><td rowspan="4">CPC-CSIS</td></tr>
<tr><td>P4</td><td rowspan="3">Generated with CFSv2 forcings</td></tr>
<tr><td>P5</td></tr>
<tr><td>P6</td><td>CDAS1 (now interpolated to 127 layers)</td></tr>
<tr><td>P7</td><td>GEFSv12 reanalysis</td><td>Spin up of Noah-MP with GDAS forcing</td><td>GEFSv12 reanalysis</td><td></td><td></td><td>Generated with GEFSv12 forcings</td></tr>
</table>

Table 2.3 Lake ice definitions for coupled UFS prototypes P1–P7.

Prototype	Lake Ice
P1–P3.1	Filled by nearest neighbor values from the ocean model. Changes accordingly over the course of integration
P4	Orography data updated to include a lake mask. Initial lake ice determined from CFSR monthly 2.5x2x5 degree climatology. Lake ice does not change over the course of integration.
P5	Lake ice = 0 (bug discovered during evaluation, and corrected in P6)
P6–P7	Initial lake ice determined from CFSR monthly 2.5x2x5 degree climatology. Lake ice does not change over the course of integration.

Additional details are provided within each prototype's dedicated section (Sections 5–11), describing the notable changes to components, and physics and parameter settings, along with any updates to the infrastructure and initial conditions.

3. Description of Model Runs

All full prototypes (P1–P7) of the coupled UFS are evaluated based on a collection of model runs (hereafter referred to as 'benchmark runs') made with the same predefined set of initialization dates to ensure a common ground for intercomparison between the various prototypes. The number of initialization dates in this setup represents a compromise between, on one hand, the need for a large enough sample to encompass various meteorological scenarios and provide sufficient robustness to bias and skill calculations, and, on the other hand, the need for judicious use of computational and storage resources.

The benchmark runs' initialization dates are the first and 15th of each month, starting with April 1, 2011 and ending with March 15, 2018. This period covers periods of both La Niña (2011, 2017) and El Niño (2015) events, as well as a period of record low Arctic sea ice (2012).

For each initialization date, a free run out to 35 days is conducted. Thus, for each full prototype there are a total of 168 hindcasts (in the remainder of this note, referred to as 'forecasts'), which allow for evaluations of bias and skill for subseasonal predictions out to week 5.

The sub-version of P7 with addition of one-way coupled atmospheric chemistry and aerosols (P7.2) was initialized on only eight dates: the first of the month in January, April, July, and October of 2012 and 2013.

The output from P1–P4 is archived at NOAA High Performance Storage System (HPSS). The output from P5–P7 is publicly accessible from https://registry.opendata.aws/noaa-ufs-s2s/, made available through the UFS Research to Operations (UFS-R2O) project that is sponsored by the NWS Office of Science and Technology Integration (OSTI) Modeling Program Division and the NOAA Oceanic and Atmospheric Research (OAR) Weather Program Office (WPO).

4. Evaluation Methodology

The datasets used in prototype evaluations as either sources of observations or proxy for observations are listed in Table A1.1. The reader is referred to this table for the meaning of all reference dataset abbreviations used in the descriptions below.

4.1 Benchmark Anomaly Correlation Scores

The collection of 168 35-day forecasts for each UFS prototype is evaluated with a consistent set of metrics to allow tracking of changes in forecast skill across prototypes. The methodology for skill evaluation follows a similar approach to what was used in the development and evaluation of CFSv2 (van den Dool and Saha, 2008), and uses the same Fortran codes.

First, smoothed climatologies for the model and for the reference (either gridded observation-based estimates or reanalysis) are calculated at each lead time. The smoothed model/reference

climatologies are calculated as the annual mean plus first four Fourier harmonics; the calculations are performed separately for each lead time, grid point and variable. The motivation for this approach and further details of the procedure are described in Johansson *et al.* (2007).

Raw (RAW) anomalies are defined as the difference between the model forecast and the smoothed observed climatology, and systematic-error-corrected (SEC) anomalies are defined as the difference between the model forecast and the smoothed model climatology. As in van den Dool and Saha (2008) and Becker *et al.* (2014), anomaly correlation (AC) scores are calculated by aggregating the squared residuals in both space and time, rather than as spatial averages of temporal correlations or temporal averages of spatial correlations.

The RAW and SEC AC scores calculated as described above constitute the benchmark AC scores for each prototype. The benchmark AC scores are reported for global tropical (20°S to 20°N) and Niño 3.4 sea surface temperatures (SST) with respect to OSTIA, CONUS precipitation and average daily 2-m air temperature (calculated, in this case only, as the average of daily minimum and maximum 2-m temperature) with respect to CPC global gridded data sets, and Niño 3.4 region precipitation and 500mb heights (Z500) for the northern (20°N to 80°N; NH) and southern (20°S to 80°S; SH) hemisphere with respect to CFSR. The AC scores for upper air variables are calculated at 1x1 degree resolution. For surface variables, AC scores are calculated at the grid-space projection of CFSv2's spectral T126 resolution (roughly 100km).

The prototype results are compared to those from the operational CFSv2 runs using the same initial dates, and subjected to identical processing. CFSv2 was selected as the basis of model comparison because of its availability of 35-day forecasts initialized on the same dates as the benchmark runs (GFSv16 forecasts are only available up to 16 days, and GEFSv12 35-day forecasts are initialized every Wednesday, instead of the 1st/15th of the month). Note that after the prototype development stage is completed, and the coupled UFS MRW and S2S applications finalized, prior to operational implementation their forecast skill will be compared with that of the corresponding current operational system.

4.2 Annual Mean Biases

Global maps of the annual mean biases are presented for select variables: SST with respect to OSTIA, 2-m air temperatures (T2m, calculated as a true daily mean, not from averaging the daily minimum and maximum) with respect to CFSR, daily minimum and maximum 2-m temperature (T2min, T2max) with respect to CPC global gridded dataset, precipitation with respect to IMERG, and outgoing longwave radiation (OLR) with respect to NOAA-CDR, and compared to the annual mean biases of the preceding prototype (or, in the case of the first coupled UFS prototype, to CFSv2). The biases are presented at 1x1 resolution, for the average Week 1, Week 2, and Weeks 3&4 simulations.

4.3 MJO

The Madden Julian Oscillation (MJO) is one of the major modes of global tropical variability on weekly to monthly timescales, manifesting as an eastward propagation of anomalies in precipitation, cloudiness and zonal winds, with a time scale of roughly 30–60 days. MJO variability impacts weather events in both tropics and extratropics, and is a major source of predictability on subseasonal time scales.

The MJO is commonly defined in terms of the multivariate index, RMM (Real-time Multivariate MJO series, Wheeler and Hendon, 2004). The RMM consists of two components, RMM1 and RMM2, obtained as the first two principal components from an EOF decomposition of near-equatorial zonal winds at 850mb (U850) and 200mb (U200), and outgoing longwave radiation (OLR) anomalies. The EOF associated with RMM1 represents a structure of enhanced convection over the Maritime Continent and suppressed convection over the Western Hemisphere and Africa. The EOF associated with RMM2 represents a structure of enhanced convection over the Western Pacific and suppressed convection over the Indian Ocean.

The prototypes' MJO predictions are primarily assessed through the AC and RMSE of the individual modes and individual near-equatorial components (U850, U200, and OLR), and the bivariate correlation/RMSE of the combined index. Also assessed in later prototypes are the errors in MJO amplitude and phase. The MJO amplitude is defined as

$$AMP = \sqrt{RMM1^2 + RMM2^2},$$

and its error as

$$E(AMP) = \frac{1}{N}\sum_1^N [AMP(f) - AMP(o)].$$

The MJO phase is defined as

$$phase = tan^{-1}\left(\frac{RMM2}{RMM1}\right)$$

and its error as

$$E(\text{phase}) = \frac{1}{N}\sum_1^N tan^{-1}\left(\frac{RMM1(o) * RMM2(f) - RMM2(o) * RMM1(f)}{RMM1(o) * RMM2(o) + RMM1(f) * RMM2(f)}\right)$$

The methods for calculating the MJO indices are based on Wheeler and Hendon (2004) and Lin *et al.* (2008). The observations are the daily OLR from the CPC archives and U200 & U850 from CDAS2 from 01 Jan 2010 through 30 Sep 2018, using CPC codes/scripts developed by Mingyue Chen and adapted for this study.

4.4 Ocean

4.4.1 Evaluation of the Ocean Initial Conditions for P2–P7

The ocean initial conditions used in P2 through P7 are evaluated using monthly fields from ocean reanalyses ECMWF-ORAS5 and SPEAR-ECDA for temperature, salinity, mixed layer depth (MLD), upper ocean (300m) heat content, and meridional transport. Given that the initial conditions are spaced approximately 15 days apart, a monthly mean for the dataset is approximated by the average of the first day of integration (1st and 15th of each month) of one of the prototypes (P5). Due to the minimal change in ocean state over the course of a single day, it is equally valid for all prototypes and reflects the state of the initial conditions.

Global and temporal averages of temperature and salinity profiles, and spatial maps of temperature, salinity and MLD are used to assess the mean biases in the initial conditions for the period 1 April 2011–15 March 2018. Correlations of zonally and meridionally averaged temperature and salinity are used to evaluate the variability of the respective fields with reanalysis-based estimates. The agreement between the two reanalyses, ECMWF-ORAS5 and SPEAR-ECDA, is highlighted. Spatial maps of MLD (calculated uniformly using a density criterion of 0.125 kg/m^3 across all the datasets) correlations with the two reanalyses are also shown. Individual basin-averaged upper ocean heat content is assessed to capture the upper ocean temporal variation. The mean meridional overturning circulation stream function is also calculated for the period 2011–2018 from reanalyses to assess the same in the initial conditions. For these evaluation results please see Appendix 2.

4.4.2 Evaluation of the Ocean Forecasts in earlier prototypes (P3.1 and P4)

The ocean forecast in the first two prototypes, P1 and P2 was only evaluated in terms of SST as part of the annual mean bias calculations described earlier.

In prototypes P3.1 and P4, the evaluation consists of comparisons of the simulated mean ocean states averaged over the first month of forecasts initialized on the first of the month for January, April, July, and October of all years of the benchmark period with the corresponding mean states from observations. These comparisons encompass: (a) sea surface temperature (SST) versus OSTIA; (b) mixed layer depth (MLD) versus climatological monthly means from ARGO; (c) sea surface height (SSH) versus AVISO; (d) surface zonal currents versus OSCAR; (e) equatorial undercurrent versus GODAS; (f) sea surface salinity (SSS) versus SMOS; and (g) equatorial vertical temperature structure and thermocline depth (D20) versus EN4. Also presented are the biases of the ocean surface radiation fluxes (net short- and long-wave) with respect to CERES-EBAF, heat fluxes (sensible and latent) with respect to ERA5, and zonal wind stress (also with respect to ERA5).

4.4.3 Evaluation of the Ocean Forecasts in later prototypes (P5, P6, P7)

In more recent prototypes, P5 through P7, the ocean evaluation focuses on the Weeks 3&4 forecasts from all initialization dates. Due to long memory in the ocean, the subsurface is expected to largely retain a strong influence of the initial conditions, although appreciable departures are already present by Weeks 3&4. Model skill is assessed through anomaly correlation maps of SST and zonal wind stress, and time series of the Niño 3.4 Index and the Dipole Mode Index (DMI). The Weeks 3&4 forecasts are also aggregated into seasonal means and the biases for OND and MJJ forecasts are evaluated separately. Seasonal biases are assessed for SST (with respect to OSTIA), net surface radiation (with respect to CERES-SYN1°), and zonal wind stress (with respect to ERA5). The biases are calculated as the difference between the model seasonal mean and similarly aggregated based on valid-on time corresponding observational estimates. Also assessed are the changes between consequent prototypes in the seasonal means of MLD, thermocline depth (D20), surface currents, and equatorially averaged cross-sections of temperature and zonal currents.

4.5 Sea Ice

The sea ice evaluation for P1–P7 consists of a qualitative comparison of the daily Arctic sea ice area in all individual forecasts from 2012–2017. The prototype Arctic ice area is compared with the CFSv2 forecast and with the estimated sea ice area derived from the ice concentrations in NASA-Team and NOAA/NSIDC-CDR (v3).

4.6 Land

Additional evaluations targeting the impact of land model parameterizations are performed for P4 (illustrating the consequences from an unintended parameterization change) and P7 (illustrating the impact of introducing Noah-MP). For P4, this evaluation consists of comparing Week 1 averaged differences in forecast temperatures (T2m and land skin temperature, Tskin) resulting from only changing the unintended parameterization with the differences between P4 and the preceding P3.1.

The land evaluations for P7 include an analysis of 35-day averaged diurnal cycles of Tskin from forecasts initialized on 01 Jan, Apr, Jul, and Oct for each of the seven years at three representative CONUS stations with SURFRAD observations, matched to the nearest model grid point. In-situ measurements at the South Pole Observatory, obtained from the public ftp file archive of the ESRL Global Monitoring Laboratory (https://gml.noaa.gov/dv/data/?site=spo) are used for verification of surface fields over Antarctica. Additionally, basin-scale water budgets are calculated over ten large river basins (Amazon, Yangtze, Congo, Mississippi, Danube, Nile, Ob, Murray, Lena, and Yukon) representing diverse climate regimes. The mean P7 annual precipitation for each basin is compared to the preceding prototype (P6) and with that calculated from the gridded CPC global 0.5 degree unified rain gauge data. Similarly compared is the basins'

mean annual evapotranspiration (ET), using observations obtained from the GLEAM (v3) global 0.25° monthly ET product.

4.7 Aerosols

Predicted aerosol optical depths (AOD) in a sub-version of P7 (P7.2), the only configuration to include an atmospheric chemistry and aerosols component, are evaluated against the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) analysis, as well as against data from the NASA's Moderate resolution Imaging Spectrometer (MODIS). The spatial distribution of forecast AOD mean difference (model minus either MERRA-2 or MODIS) and RMSE is calculated for Weeks 1, 2, and 3&4 combined. Additionally, time-series of the global RMSE are calculated with 6-hourly output over the first 4 weeks of forecast. Results are shown for the total AOD, and for the AOD of individual species.

5. Prototype 1 (P1)

5.1 Configuration and Settings

The coupled model for P1 consists of an atmosphere, ocean and ice model. The NOAA Environmental Modeling System (NEMS) which encompasses a model driver and a mediator for coupling fields between the components is employed. This coupling infrastructure relies on the National Unified Operational Prediction Capability (NUOPC) Layer of the Earth System Modeling Framework, ESMF (Theurich *et al.*, 2016). Within NUOPC, each component model communicates with the mediator through a software layer, referred to as that component's cap.

In this coupled model configuration, the atmosphere/ice fluxes are computed by the ice model and the atmosphere/ocean fluxes are computed by the atmospheric model. The land/sea masks of the ocean/ice model and the atmospheric model do not match. Therefore the sea surface temperature was interpolated from the ocean grid to the atmospheric model grid using a conservative method followed by a nearest neighbor fill. The lake values are filled by nearest neighbor values from the ocean model.

The atmospheric model in P1 and all other prototypes employs the GFDL's Finite-Volume Cubed-Sphere Dynamical Core, FV3 (Harris *et al.*, 2021). The atmospheric model uses the C384 cubed-sphere grid (corresponding to horizontal resolution of ~25km) and has 64 hybrid vertical layers, with model top at about 0.2hPa (54km). P1 uses the Interoperable Physics Driver (IPD) with GFSv15.2 physics, which includes Scale-aware Simplified Arakawa-Schubert Scheme for convection (Han and Pan, 2011), Hybrid Eddy-Diffusivity Mass-Flux (EDMF) Boundary Layer Parameterization (Han *et al.*, 2016), Rapid Radiative Transfer Model for GCMs (RRTMG) with Monte-Carlo Independent Column Approximation (McICA) for radiation scheme, GFDL microphysics with five prognostics cloud species, and Noah LSM (Mitchell *et al.*, 2005).

The ocean model is the Modular Ocean Model, MOM6, following the OM4.0 setup (Adcroft *et al.*, 2019) on a ¼ degree tripolar grid with 75 hybrid depth-isopycnal coordinates.

Sea ice is modeled by the Los Alamos Sea Ice Model, CICE5, with the same ¼ degree tripolar grid as the ocean. There are five ice thickness categories, and Mushy thermodynamics is not used.

In P1 the time step is 450 seconds for FV3 and 900 seconds for CICE5 and for the MOM6 dynamical time step. The ocean model thermal time step is 3600 seconds. The coupling time step, or the frequency at which the components communicate with each other, for all three models in P1 is 1800 seconds. A portion of the benchmark runs failed with these settings due to numerical instabilities, and required rerunning with reduced timesteps. The time step specifications were adjusted in the next prototypes to improve model stability.

Note that in this initial prototype it was discovered that the surface temperature in the atmospheric model was not correctly reflecting the sea ice surface temperature; this issue was subsequently fixed for the next prototype.

5.2 Initial Conditions

All component's initial conditions for P1 are derived by interpolation from the CDAS1 system which uses: (a) the GFS atmospheric model on a spectral T384L64 Eulerian grid; (b) the MOM4 ocean model with Z-coordinates on a tripolar grid, with global ½ deg resolution, increasing to ¼ deg in the tropics; and (c) the SIS1 sea ice model on the same tripolar grid as MOM4. The ocean initial conditions for P1 only contain temperature and salinity values; at the start of integrations, the ocean is at rest and with a flat surface.

The CICE5 ice enthalpy for 7 ice layers is estimated by linearly interpolating the CDAS1/SIS1 2-layer ice temperatures and specifying an idealized ice salinity profile. For some fields (e.g. downward shortwave radiation by bands), there is no comparable field available from the SIS1 file and these fields are initialized as zero. Also, while both SIS1 and CICE5 are configured with 5-categories of ice thickness, the thickness distributions used by the two models are different. This difference is ignored in creating the CICE5 ICs, so that the SIS1 category fields are mapped 1:1 to the CICE5 category fields. Finally, the required fields are mapped from the SIS1 ½ deg MOM6 tripole grid to the CICE5 ¼ deg MOM6 tripole grid using ESMF nearest-source-to-destination mapping weights.

Lake ice is derived from the nearest ocean grid point and changes accordingly over the course of integration.

5.3 Evaluation

5.3.1 Benchmark AC scores

The SEC approach used to estimate and remove the model bias appears to be functioning correctly despite the relatively short dataset of the benchmark runs, as evidenced by the SEC AC scores being higher than their raw counterparts. Both raw and SEC AC scores for P1 show improvement over CFSv2 for the majority of benchmark scores and lead times (Fig. 5.1). This is the case for global tropical SST, Niño 3.4 SST (except for Week 1), CONUS T2m, CONUS precipitation (except for Weeks 3&4), Z500 (except for Week 3 in SH), and for the Week 1 forecasts of Niño 3.4 precipitation. The CONUS AC precipitation scores after Week 2 are very low (below 10%) in both P1 and CFSv2. The spatial patterns of CONUS T2m (Fig. 5.2) and precipitation (Fig. 5.3) ACs are qualitatively similar between the two systems. For subsequent prototypes the AC maps remain qualitatively similar, and are not shown.

5.3.2 Annual Mean Biases

The annual SST bias from 50°S to 50°N in P1 is warm along the equator and most of the tropics (Fig. 5.4). The warmest tropical biases are found in the western Indian Ocean/eastern Pacific, and along the western coasts of South America and Africa. Cool SST biases are present in the tropical eastern South Pacific, near the South American coast, over a region extending from the equator to 30°S, and from 120°W to about 80°W, and in the southern equatorial Atlantic. The spatial structure of the SST biases does not change between Week 1 and Weeks 3&4, but their magnitude increases with lead time, i.e., the warm biases grow warmer and the cool biases grow cooler. By Weeks 3&4, the warm bias over the central equatorial Pacific is +0.2 to +0.5K, and along the western South American and African coasts, +1 to +2K; the cool bias in the eastern South Pacific reaches −0.2 to −1K. Compared to CFSv2, the SST in P1 is warmer in most tropical and subtropical oceans, but the cool bias regions of P1 are cooler than CFSv2.

Surface temperatures over areas poleward of 50°S and 50°N are not compared to OSTIA observations due to the presence of sea ice at high latitudes. The comparison of the high latitude surface temperatures between CFSv2 and P1 reveals much warmer icy oceans in P1. Upon further investigation, it was found that this was due to the model setting the atmosphere's bottom boundary surface temperature over sea ice to the freezing temperature of water, instead of the correct (lower) value calculated by the ice model. As a consequence, the Arctic annual mean T2m (Fig. 5.5) show a strong warm bias (up to +10K) in all forecast weeks. This error was corrected in P2 and subsequent prototypes

Over land, at all weeks the annual mean T2m (Fig. 5.5) in P1 have a warm bias over Africa (up to +5K for Sahara and the Arabian peninsula), Australia, South America, southern Asia, and the western coast of North America. The remainder of North America, and most of northern Europe and northern Asia have a cool bias (up to −5K in Siberia by Weeks 3&4). Both the warm and cool biases grow in magnitude with lead time. Compared to CFSv2, the warm biases in P1 tend to be warmer, and the cool biases—cooler.

For all forecast weeks, the annual mean T2max (Fig. 5.6) in P1 has a pronounced cool bias over most of Europe, Asia, and North America, and a warm bias over equatorial Africa, South America north of 30°S, and the Indian subcontinent. Compared to CFSv2, T2max is warmer over most land areas between 60°S and 60°N, but cooler over the highest-latitude land areas of the northern hemisphere and over Antarctica.

The bias of the annual mean T2min (Fig. 5.7) for most—but not all—regions tends to be of opposite sign to the bias of T2max. For most of North America, northern Africa, and northeast and southwest Asia, T2min has a warm bias, while for equatorial Africa and South America T2min has a cool bias. Most of the warm biases of T2min are also present in CFSv2.

The precipitation bias (Fig. 5.8) in P1 is generally wet over most ocean areas, and of mixed sign over land. Compared to CFSv2, P1 shows a reduction of the CONUS wet bias and the tropical South America dry bias. The wet bias over tropical oceans in P1 is shifted southward. The previously discussed areas of prominent cool SST bias in the tropical SST in P1 are characterized by increased wet bias.

The annual mean bias of OLR (Fig. 5.9) in P1 is overwhelmingly positive in sign, with a global mean of roughly +11 W/m^2 in all forecast weeks. The largest biases are in the tropics, consistent with the geographical distribution of largest climatological precipitation totals, and over high latitude oceans. Compared to CFSv2, this represents a marked increase in global OLR bias (the global mean bias in CFSv2 is roughly +2 W/m^2). The only regions for which the positive OLR bias in P1 is reduced relative to CFSv2 are in the southern hemisphere, off the west coasts of South America and Africa (coinciding with the region of cold SST bias) and Antarctica.

5.3.3 MJO

The AC of MJO RMM1, RMM2 and combined RMM in P1 shows a clear improvement over the operational CFSv2 at all lead times (Fig. 5.10, top). In CFSv2, the lead time until the AC decreases to 0.6 is approximately 10.5 days for RMM1 and 15 days for RMM2. By comparison, in P1 these numbers are approximately 13.5 and 18 days respectively. The combined RMM AC shows similar improvement, increasing from approximately 12 days in CFSv2 to approximately 17 days in P1. All three equatorial filtered components (U850, U200, and OLR—Fig. 5.10, bottom) from which the MJO indices are calculated show a clear increase of skill in P1 compared to CFSv2, indicating that the gain in lead time of useful MJO AC reflects an improved representation of the large-scale circulation and convection in the tropics compared to CFSv2.

5.3.4 Sea Ice

The Arctic ice area derived from NASA-Team ice concentrations is systematically lower than that from NOAA/NSIDC-CDR v3 for all months and years (Fig. 5.11). The CFSv2 Arctic ice area forecasts tend to be close to the NOAA/NSIDC-CDR estimate, occasionally exceeding it in forecasts initialized in summer months. Compared to both NASA-Team and NOAA/NSIDC-CDR,

the ice area in CFSv2 grows too rapidly during the ice formation season, and decreases too slowly during ice melt season.

In most instances, P1 predicts a smaller Arctic ice area than CFSv2. The P1 ice area forecast is systematically lower than NOAA/NSIDC-CDR v3 in all months. For forecasts initialized in winter, the P1 forecasts tend to be lower than NASA-Team as well. The P1 forecasts for winter months show rapid reduction of ice area in the first several days after initialization, likely partially due to the previously mentioned bug in the surface temperatures seen by the atmosphere over ice covered oceans. For non-winter months, the P1 forecast Arctic ice area lies between the two observational estimates.

6. Prototype 2 (P2)

6.1 Configuration and settings

In P2, several issues in the coupled model are fixed. The error in sea ice surface temperature mentioned in the discussion of P1 is corrected. The solar zenith angle calculation is improved by centering in time. Additionally, more accurate representation of hydrometeors in radiation is achieved by setting parameters "*lgfdlmprad*" and "*effr*" to true in the gfs_physics_nml namelist. With these changes, radiation directly uses the microphysics (and convection) determined parameters such as suspended liquid water and ice, rain, snow and graupel and their effective radii, instead of estimating them within the radiation itself, as done previously. A less-diffusive option for the advection scheme in the FV3 dynamical core was selected by changing the fv_core_nml namelist parameter "*hord*" (previously set to 6) to 5.

The model time steps are modified from their P1 values to improve model stability. In P2, the "fast coupling time step" is set to 450 seconds, whereas P1 used the same coupling time step (1800 seconds) for both fast and slow coupling. The ice model time step in P2 is reduced to 450 seconds (compared with 900 seconds in P1) to allow the atmospheric and ice models to run on the same time steps within the "fast coupling" loop. The ocean model thermal time step is reduced to 1800 seconds compared with 3600 seconds in P1. The ocean model communicates to the ice and atmospheric models on the "slow coupling" time step, which is 1800 seconds as it was in P1.

Starting with P2, a unified MOM6 NUOPC cap (*i.e.*, the software layer providing communication between MOM6 and the mediator) jointly developed by EMC, NCAR, and GFDL replaces the in-house EMC cap used for P1.

6.2 Initial conditions

The main feature differentiating P2 from P1 is the ocean initial conditions. The ocean initial conditions for P2 are obtained from the CPC-3DVar ocean data assimilation product which is based on GFDL MOM6-examples OM-025 setting (Adcroft *et al.*, 2019; for more details see

https://github.com/NOAA-GFDL/MOM6-examples/blob/dev/gfdl/ice_ocean_SIS2/OM4_025/MOM_input). Unlike the ocean initial conditions for P1, which were derived from CDAS1/MOM4 and consisted only of temperature and salinity, the CPC-3DVar ocean initialization provides temperature, salinity, currents, and sea surface heights. The CPC-3DVar ocean initial conditions are also used for all subsequent prototypes. An assessment comparing the CPC-3DVar initial conditions with existing ocean reanalyses is presented in Appendix 2.

The atmospheric and sea ice initial conditions for P2 are identical to those in P1. As in preceding prototypes, lake ice in P2 is derived from the nearest ocean grid point and changes accordingly over the course of integration.

6.3 Evaluation

6.3.1 Benchmark AC Scores

The comparison of P2 and P1 benchmark AC scores (Fig. 6.1) shows clearly increased scores for global tropical SST and Niño 3.4 SST at all lead times. This improvement is on the order of 2 (4) percentage points in the RAW (SEC) scores for global tropical SST, and on the order of 1 (2) percentage points in the RAW (SEC) scores for Niño 3.4 SST. Also increased by a few percentage points are the AC scores for Niño 3.4 precipitation at all lead times. The impact is mixed for the remainder of benchmark scores at shorter lead times, but at Weeks 3&4 all benchmark AC scores for P2 match or exceed those of P1, and exceed those of CFSv2.

As Fig. 6.1 demonstrates, the model forecast skill for CONUS precipitation beyond Week 2 is particularly low. Applying a data adaptive method known as the multi-channel singular spectral analysis (MSSA), Krishnamurthy *et al.* (2021) showed that in observations, in addition to the canonical ENSO mode of variability and long-term climate trends that have been known to affect the surface weather over CONUS during the boreal summer (defined in their study as JJAS), another mode, which manifests as an intraseasonal oscillation (ISO) with a period of about 50 days, influences the daily variability of precipitation and T2m. This mode has a strong signature in the lower troposphere large-scale circulation over 180°–15°W in the latitudinal band between 20°S–41°N, and projects well onto the T2m and precipitation. Their evaluation of 7 years of forecasts produced by P2 showed an overall good agreement between the climate modes simulated by the model and their observed counterparts. Some regional disagreements between the model and forecasts still exist and show from Week 1 through week 4 forecast leads. The forecast skill of the climate modes signal in JJAS precipitation over CONUS, decomposed with the methodology in Krishnamurthy *et al.* (2021), demonstrates that beyond Week 2 the ISO mode is the leading source of predictability (Fig. 6.2). While the anomaly correlation based on the daily anomaly is almost zero beyond Week 2, the correlation for the ISO mode is slightly higher than 0.3.

6.3.2 Annual Mean Biases

The annual mean SST bias (Fig. 6.3) in P2 has a similar spatial pattern as P1, but is warmer. Analysis of a limited set of experiments (not shown) indicated that the new ocean initial conditions, and not the changes in atmospheric physics parameterization, were the main reason for this warming. As in P1, the warm bias increases with lead time. As expected, correcting the bug in surface temperature over sea ice resulted in significantly lower annual mean surface temperatures in regions with sea ice (Arctic and Antarctic).

The annual mean T2m bias (Fig. 6.4) retains a similar spatial structure, but is noticeably warmer in P2 than in P1, especially over land and at higher latitudes, and increasingly so with growing lead time. The T2m bias over the Arctic is reduced in magnitude, but still of positive sign. Both T2max (Fig. 6.5) and T2min (Fig. 6.6) temperatures are warmer in P2 than in P1, and increasingly so at longer lead times.

The spatial pattern of precipitation bias is similar between the two prototypes (Fig. 6.7). P2 shows a slight reduction in the wet bias over tropical oceans. An exception to this are the surroundings of the Maritime Continent, where wet bias is increased.

The OLR bias in P2 (Fig. 6.8) is still overwhelmingly positive (global mean bias roughly +5 W/m^2), but its magnitude is much reduced compared to P1 (global mean bias roughly +11 W/m^2). The spatial pattern of the bias is similar between the two prototypes, with maximum values along the equatorial convection regions.

6.3.3 MJO

P2 performs better than both P1 (at most lead times) and the operational CFSv2 (at all lead times) for RMM1, RMM2, and the combined RMM (Fig. 5.10, top). The RMM1 AC for both prototypes remains above 0.6 until forecast day 13.5, compared to forecast day 10.5 for CFSv2. The AC for RMM2 in P2 is larger than P1, increasing the lead time of useful skill by about 2 days (from 18 to 20 days). The combined RMM AC score shows P2 gains approximately 2 days compared to P1, and 7 days compared to CFSv2 in MJO forecast skill. The filtered equatorial components (Fig. 5.10, bottom) show a modest improvement in the AC for U200 and OLR between P1 and P2.

6.3.4 Sea Ice

The Arctic sea ice area (Fig. 6.9) in P2 is larger than in P1, consistent with the colder surface and near-surface temperature in the P2 Arctic. This increase in ice area is particularly prominent for forecasts initialized in the January–June period. Over these months, the P2 forecast moves closer to the observational estimate derived from NASA-Team. For the remainder of months, the P2 forecast lies between NASA-Team and NOAA/NSIDC-CDR.

7. Prototype 3.1 (P3.1)

7.1 Configuration and Settings

The initial prototype configured following P2 (P3) did not contain all necessary updates and describing it is beyond the scope of this document. Instead, described here is the following prototype, P3.1.

There were numerous updates to FV3 between P2 and P3.1, primarily consisting of bug fixes to various parts of the source code. In addition, there were many updates made as part of the GFSv16 physics improvements, however most of these options are turned off by default. The following are a few updates to note. The ice fraction imported from the ice model into the atmospheric model is now explicitly limited between zero and one to eliminate any instances of physically impossible values. This fixes one of the stability issues previously encountered. Although P3.1, as all preceding prototypes, uses non-fractional grid (*i.e.*, a given atmospheric grid cell is considered to be either 100% over land or 100% over ocean, depending on the dominant surface), the model underwent many updates related to the treatment of fractional grid in preparation for its subsequent implementation (for more details on the fractional grid, see Section 10.1.1). Some of these updates were related to the radiation treatment at the bottom boundary. The radiation calculation now recognizes fractional sea ice coverage, *i.e.*, accounts for the fact that water points can be partially ice covered and albedoes over ice and water are very different. Also included are updates to the GFDL microphysics package to reduce the cold bias in the lower troposphere and near the surface found in the uncoupled runs. Two changes in the microphysics that might have affected the OLR and could explain the reduction of OLR bias seen in P3.1 (described in Section 7.3.2 below) should be noted: (a) a new scheme for the effective radius of suspended ice was added following Wyser (1998), and (b) the terminal fall velocity of ice was reduced to 90% of its original value (Ruiyu Sun, NOAA/EMC, internal communication).

7.2 Initial Conditions

One of the main features differentiating P3.1 from P2 is the initialization of sea ice. The sea ice initial conditions for P3.1, and for all subsequent prototypes, are derived from the CPC Sea Ice Initialization System (CPC-CSIS; described in Collow *et al.*, 2019 and Liu *et al.*, 2019) which assimilates NASA-Team ice concentrations. The CICE5 initial conditions are generated using the CPC-CSIS restart files using a procedure similar to the one employed in P1.

The atmospheric and oceanic initial conditions for P3.1 are identical to those in P2. As in preceding prototypes, lake ice in P3.1 is derived from the nearest ocean grid point and changes accordingly over the course of integration.

7.3 Evaluation

7.3.1 Benchmark AC Scores

Most AC scores for P3.1 are similar to those in P2 (Fig. 7.1) at all lead times, with the exception of Z500 (both NH and SH), which decay faster with lead time and are clearly lower than the P2 scores by Weeks 3&4.

7.3.2 Annual Mean Biases

The SST bias in P3.1 is very similar to P2 (Fig. 7.2). As in preceding prototypes, there is a warm bias in the tropics, most pronounced in the Indian Ocean, western Equatorial Pacific, and along the western coasts of South America and Africa, and a cool bias region in the eastern South Pacific offshore from South America. Both types of biases increase with lead time. The annual mean surface temperatures over areas with sea ice have decreased in the Arctic, and increased in the Antarctic.

The annually averaged T2m in P3.1 are, as in P2, displaying a warm bias compared to CFSR (Fig. 7.3). The spatial pattern of bias is very similar to P2, with warmest biases found over tropical land. With lead time, P3.1 is progressively warmer than P2 in a region extending from Alaska to the Great Lakes, and over much of extratropical Asia. The T2m is lower over the Arctic and sub-Arctic land, resulting in a reduced warm bias there.

The bias of annually averaged T2max (Fig. 7.4) and T2min (Fig. 7.5) in P3.1 compared to CPC's gridded dataset also shows a similar spatial pattern to that of P2. The most pronounced feature of the T2max is a warm bias over the humid regions of equatorial Africa, the Indian subcontinent, the Amazon region, and northern Siberia, and cool bias over much of Greenland, North America and east Asia. The pattern of differences in T2max between P3.1 and P2 is similar to that of T2m. For T2min, again, the spatial pattern of bias is very similar in both prototypes, with predominantly warm bias in most regions with the exception of Greenland, Southeast Asia and the Amazon basin. The spatial pattern of the differences between the two prototypes for T2min is the same as for T2max.

The precipitation bias (Fig. 7.6) in P3.1 is also similar to P2, with predominantly wet bias in the southern Indian ocean, central and western tropical Pacific and in the tropical Atlantic. Dry precipitation bias is predominant between 60°E and 180°E along and northward of the equator. Dry bias is also present in the northeastern Atlantic and in the Pacific south of 30°S.

The OLR bias (Fig. 7.7) is clearly reduced in P3.1 compared to P2, especially at high latitudes. For the Arctic, the bias is reduced by 10 to 20 W/m^2 and essentially eliminated. In the tropics, the reduction is modest, and positive bias remains pronounced.

7.3.3 MJO

The AC of RMM1 decreases to 0.6 in forecast day 13.5 for P2, whereas for P3.1 that is forecast day 15.5 (Fig. 7.8). The improved RMM1 AC score indicates improved forecast of convection over the Maritime Continent in P3.1 compared to P2. However, P3.1 has a slightly lower AC for RMM2 compared to that of P2, indicating a slightly degraded convection over the Pacific Ocean. The combined RMM AC score in P2 is slightly better than P3.1 (AC score >0.6 through forecast day 20 in P2, compared to day 19 in P3.1). However, both prototypes' MJO skill is an improvement over the operational CFSv2.

7.3.4 Sea Ice

The Arctic sea ice area (Fig. 7.9a) in P3.1 forecasts is generally lower than that of P2, and closer to the NASA-Team estimate. Since the initial conditions for CPC-CSIS rely on data assimilation using NASA-Team ice concentrations, which has markedly lower sea ice concentration than CDR (Fig. 7.9b), this is not surprising. The P2 forecasts of the Arctic sea ice area are particularly close to the NASA-Team values for forecasts initialized between June and September. For forecasts initialized between October and May, P3.1 tends to produce a brief initial rapid increase of ice area, lasting a couple of days, followed by more gradual evolution.

7.3.5 Ocean Mean State

In this section major oceans mean state features are compared with observations. Overall, the deepening and shallowing of mixed layer depth (MLD) in the boreal winter and summer season is well represented in P3.1 (Fig. 7.10 and Fig. 7.11). MLD is further compared with Argo dataset by its probability distribution function. From Fig. 7.12 it is clear there is a slightly higher percentage of low (0–40m) MLD values in most seasons in P3.1 compared with Argo, and a lower percentage in the 40–80m range. P3.1 and Argo show similar percentages for high MLD ranges (>200m).

For sea surface height (Fig. 7.13 and Fig. 7.14), the twin peaks of SSH in the subtropical Pacific and Atlantic Oceans are well represented in the model. Ocean surface currents are compared with OSCAR (Fig. 7.15 and Fig. 7.16). A clear stratification of North Equatorial Current, Equatorial Countercurrent, and the South Equatorial Current, along with large scale subtropical gyres can be seen in P3.1, although currents in the model tend to be slightly stronger than OSCAR (Fig. 7.17). Strong currents in several key regions (Gulf Stream, Kuroshio and Loop currents) are well represented (Fig. 7.18 through 7.23). Sea surface salinity in P3.1 in general matches well with SMOS (Fig. 7.24) but tends to be saltier in coastal regions and fresher in the South Polar region. The vertical structure of temperature and currents in equatorial region shows similar 20-degree isothermal depths between P3.1 and EN4 (Fig. 7.25), and the seasonality of Equatorial Undercurrents is well captured in the model although for a slightly overshooting in magnitude (Fig. 7.26 and Fig. 7.27).

Next, SST and air-sea fluxes are compared with observations. Overall, SST shows slightly warm bias in ITCZ and cold bias in cold tongue regions (Fig. 7.28). The SST bias patterns correspond well to the net shortwave radiation difference between P3.1 and ERA5 (Fig. 7.29) which indicates

that the net shortwave bias is likely a major contributor to the SST bias. The surface net longwave radiation (Fig. 7.30) in P3.1 is somewhat lower than ERA5. Latent heat (Fig. 7.31) has a minor positive bias along ITCZ which matches the pattern of SST bias. Sensible heat (Fig. 7.32) has a minor positive bias in North Hemisphere Western coastal regions. Wind stress (Fig. 7.33) shows good agreement with ERA5.

8. Prototype 4 (P4)

8.1 Configuration and Settings

In P4, a wave component (WAVEWATCH III) is added to the coupled system. The wave model uses a regular 0.5 degree grid. The directional spacing is 10 degrees and the first direction is offset by 0.5 degrees. The spectral domain is discretized into 36 directions (10 degree directional spacing) and 36 frequency bins starting at 0.035 Hz and a frequency increment factor of 1.10. The 'ST4' physics package (Ardhuin *et al.*, 2010), which is used operationally at NCEP for both the global deterministic and ensemble forecasts, is employed.

The wave model receives the 10-meter wind components from the atmospheric model, the surface currents from the ocean model, and the ice concentration from the ice model for its forcing. The wave model sends the Stokes drift to the ocean model for calculation of sea-state dependent Langmuir mixing instead of a wind-speed based parameterization. The wave model sends the z_0 roughness length to the atmosphere model. The land/sea mask between the atmosphere and the wave model currently do not match, which requires determining the values of the surface roughness where there are not valid values from the wave model. For this initial prototype with waves, these values are set to zero (an update implemented in the following prototype, P5, ensures that the wave model does not send surface roughness values over ice covered points and that the atmospheric model would perform its own calculation of surface roughness when a valid value is not given from the wave model). A schematic of the entire coupled system is shown in Fig. 8.1.

There were various updates to the atmospheric model. Most notably, the Common Community Physics Package (CCPP) replaces the IPD physics driver. A new surface thermal roughness length parameterization based on Trier *et al*. (2011) is inadvertently invoked in place of the previous parameterization based on Zheng *et al.* (2012), leading to systematic changes in near-surface temperatures. This parameterization was reverted in the next prototype.

The orography data are updated to include a lake mask, which allows for the lake temperature to be set by climatology instead of the nearest ocean point. After this change, lake ice in P4 forecasts is fixed to its value at initialization, and is not changing with integration.

The atmospheric model is updated to export ocean-only fluxes to the mediator for coupling with the ocean and the NEMS mediator was updated to remove any flux calculations.

8.2 Initial conditions

The ocean, ice, and atmospheric initial conditions in P4 are the same as in the preceding prototype. The wave initial conditions were created by running multiple streams of spin-up runs. Each initial condition was generated by a minimum of one-month spin-up to ensure that the swell was sufficiently propagated after starting from rest. The spin-up was forced using CFSv2 10m wind speeds and ice concentrations.

8.3 Evaluation

8.3.1 Benchmark AC scores

The AC scores for P4 (Fig. 8.2) at leads up to two weeks are similar to those of P3.1. After Week 2, the tropical SST scores are marginally lower, but the scores for Z500—which had degraded slightly in P3.1 compared to P2—are somewhat improved. The scores for CONUS T2m scores in Weeks 3&4 are clearly higher in P4, primarily due to improvements of the Week 3 forecast.

8.3.2 Annual Mean Biases

The SST bias in P4 is very similar to P3.1 (Fig. 8.3). As in all preceding prototypes, there is a warm bias in the tropics, most pronounced in the Indian Ocean, western Equatorial Pacific, and along the western coasts of South America and Africa, and a cool bias region in the eastern South Pacific offshore from South America. Both types of biases increase with lead time.

The difference between the two prototypes in terms of annual mean surface temperatures over areas with sea ice is mixed in the Arctic (cooler at shorter lead, warmer at longer lead). Over sea ice regions in the Southern hemisphere, P4 is progressively cooler than P3.1.

The large-scale spatial patterns of annual mean bias for T2m, T2max, and T2min (Figs 8.4, 8.5, 8.6) remain very similar between P4 and P3.1. Both minimum and maximum temperatures are somewhat warmer over the Antarctic and over high latitudes in the northern hemisphere. Elsewhere, maximum temperatures tend to be slightly cooler in P4, and minimum temperatures—generally warmer, with the exception of most tropical and coniferous forest regions.

The annual mean bias of precipitation rate in P4 is similar to P3.1 (Fig. 8.7). At Week 1, the precipitation over most land regions, with the exception of the Amazon basin, is slightly lower in P4. This underlying pattern of differences persists to some degree at longer lead times. The Weeks 3&4 wet precipitation bias is somewhat reduced in the central equatorial Pacific (150°E–120°W) and the dry precipitation bias around the Maritime Continent is somewhat ameliorated.

The annual mean biases of OLR (Fig. 8.8) show little change between the two prototypes at all lead times.

8.3.2 MJO

The combined RMM AC remains very similar between P3.1 and P4 (Fig. 8.9, top) over the first three weeks of forecast. The RMM1 AC in P3.1 remains above 0.6 through day 14, which is 1.5 days earlier than P4; however the number of days for which the AC of RMM2 and combined RMM remain above 0.6 is virtually the same for the two prototypes. The AC of the equatorial component variables (U850, U200, and OLR; Fig. 8.9, bottom) changes very little between P3.1 and P4.

8.3.4 Sea Ice

The Arctic sea ice area (Fig. 8.10) in P4 is overall quite similar to that of P3.1. Where differences exist, they tend to be small, and in the direction of lowering the forecast ice area by slowing down ice area increase during the ice growth season. As in all preceding prototypes, the forecast area remains systematically below the estimate derived from NOAA/NSIDC-CDR, and close to that from NASA-Team.

8.3.5 Ocean Mean State

The ocean mean state results for P4 are qualitatively similar to those for P3.1, and are not shown here.

8.3.6 Land

As mentioned in Section 8.3.2, the annual mean T2m in P4 is warmer than in P3.1 resulting in a further increase of warm biases (Fig. 8.4). Further analysis of the T2m difference (Fig. 8.11) between P4 and P3.1 in four different seasons of 2017 indicates that P4 is warmer over some regions, such as CONUS and Eurasia in July. More evidence can be seen from the surface skin temperatures (Fig. 8.12) that P4 is substantially warmer than P3.1, mostly over arid and semiarid regions, as well as over tall vegetation regions during the warm seasons. Zheng *et al.* (2012) indicated that the surface skin temperature calculation would be closely related to the land and atmosphere interaction processes. More specifically, compared to the formula of the thermal roughness length in P3.1 (Zheng *et al.*, 2012), the thermal roughness length parametrized with Trier *et al.* (2011) in P4 would be too small and result in insufficient surface heat flux between land and atmosphere for a given surface-air temperature difference.

An experiment (P4 EXP) was performed with the same P4 setup, except for replacing the thermal roughness length parameterizations with those used in P3.1. The results (Fig. 8.13) illustrate that reverting this parameterization indeed essentially eliminates the large differences in 2-m temperature and skin temperature seen between P3.1 and P4.

Note that there still remains a significant cooling over the winter hemisphere's ice-covered ocean (Arctic: Fig. 8.13, top; Antarctic: Fig. 8.13, bottom) and warming over the Great Lakes in winter (Fig. 8.13, top). The former is likely related to the introduction of wave coupling, while the latter is likely due to the lake ice and lake temperature in P4 being specified by the atmospheric model instead of the nearest ocean grid point.

9. Prototype 5 (P5)

9.1 Configuration and Settings

9.1.1 Mediator

Beginning with P5, the Community Mediator for Earth Prediction Systems (CMEPS) replaces the NEMS mediator component in the coupled system. CMEPS is used for coupling ATM-OCN-ICE while NUOPC connectors are used for coupling ATM-WAV, OCN-WAV and ICE-WAV. Fields exchanged between ATM and either OCN or ICE are mapped conservatively with nearest-neighbor mapping. This accounts for grid cells with land-sea mask differences between the ATM and OCN or ICE components. Fields exchanged between WAV and other components are mapped with bilinear mapping.

9.1.2 Sea Ice Model

Beginning with P5, CICE6 replaces CICE5 as the sea ice component in the coupled system. The number of ice categories (5), layers (7) and snow layers (1) for CICE6 are the same as those used previously for CICE5.The sea ice thermal conductivity parameterization is changed from option 'bubbly' to 'MU71'.

9.1.3. Bug Fixes

A bug-fix for the freezing point of sea-water is made in the ice model, replacing a constant freezing point temperature with one varying linearly with the surface salinity imported from the OCN, and a bug-fix is made to z_0 (the roughness length) passed between components so that ATM sets an internally consistent value for z_0 when z_0 is not available from the wave model (a situation that arises at some coastal grid points due to the coarser resolution of the wave model compared to the ocean mode; and over areas with sea ice exceeding 70%). The thermal roughness parameterization based on Trier *et al.* (2011) that was inadvertently invoked in the preceding P4 was reverted to the parameterization based on Zheng *et al.* (2012) used by P1–P3.1.

9.2 Initial conditions

All initial conditions used for P5 are identical to the initial conditions for P4.

9.3 Evaluation

9.3.1 Benchmark AC Scores

The AC scores for P5 for Week 1 and Week 2 are similar to those of P4 (Fig. 9.1). At longer leads, a noticeable degradation is apparent in the NH Z500 and CONUS T2m. A closer examination of the reasons for the degradation of CONUS T2m scores suggests that the loss of annually

aggregated skill is primarily due to the loss of skill in winter months. It is hypothesized that this is related to the absence of lake ice in P5, a shortcoming that was uncovered during evaluation and corrected in subsequent prototypes.

Krishnamurthy and Stan (2022) evaluated the extreme precipitation and temperature events produced by P5. They found that the spatial structures of the extreme events in precipitation are reasonably well predicted but with higher values and frequency. The predictions of temperature extremes are closer to observation over central and eastern parts of the U.S. whereas the model fails to generate the extreme events over large regions in the western U.S. When captured, the extreme temperature events are predicted with a higher accuracy than the precipitation extremes.

9.3.2 Annual Mean Biases

The annual mean sea surface temperature bias in P5 is virtually identical to that of P4 (Fig. 9.2). Over the Arctic region, surface temperatures are warmer than in P4, and the warm difference increases with lead time. A detailed description of P5 SST biases is provided in Section 9.3.5 further below.

The biases of the annual mean T2m, T2max, and T2min are similar (Figs 9.3, 9.4, 9.5). The large-scale pattern of the differences in these fields between P5 and P4 closely corresponds to the inverse of the differences between P4 and P3.1, consistent with the reversal of the thermal roughness parameterization inadvertently modified in P4. In addition, the Great Lakes are much warmer in P5 than in P4—a consequence of the inadvertent elimination of lake ice in P5.

The annual mean biases of precipitation rate in P5 are similar to P4 (Fig. 9.6). At Week 1, the precipitation over most land regions, with the exception of the Amazon basin, is slightly higher in P4, and it too closely corresponds to the inverse of the differences between P4 and P3.1. The Weeks 3&4 wet precipitation bias is somewhat increased in the central equatorial Pacific (150°E–120°W), as is the dry precipitation bias around the Maritime Continent.

The annual mean biases of OLR (Fig. 9.7) show little change between the two prototypes at all lead times.

9.3.3 MJO

The comparison of P4 and P5 AC for MJO RMM1, RMM2, RMM, and for the filtered equatorial components is shown in Fig. 9.8 (top).

The RMM1 AC remains similar between P4 and P5 over the first two weeks of forecast, decreasing to 0.6 around forecast day 14. After Week 2, the RMM1 AC of P5 is lower than P4. The AC values for RMM2 are similar between the two prototypes, reaching 0.6 around forecast day 19. The combined RMM loses 2 days of useful forecast in P5 compared to that in P4 (day 16 in P5 compared to day 18 in P4). In terms of component variables, the AC for U200 is slightly lower in P5 compared to that in P4 (Fig. 9.8, bottom).

Consistent with the AC, The MJO RMM RMSE in P5 is larger than P3.1 and P4 (Fig. 9.9) after the first two weeks, due to the increased RMSE in both RMM1 (enhanced convection over the Maritime Continent) and RMM2 (enhanced convection over west Pacific). The larger RMSE in RMMs is primarily contributed by U200 (Fig. 9.10). The error in OLR also has some contribution but it is likely to be smaller than that from the two wind components due to the smaller weight given to OLR in calculating the RMMs.

The larger RMSE in MJO RMMs in P5 is consistent with the larger bias in phase speed (Fig. 9.11, bottom). Overall, MJO propagates too slowly in all prototypes, and this is also the case for P5. In addition to the slower propagation, P3.1 and P4 tend to have an overly strong MJO, with amplitude bias increasing with lead time, while P5 shows relatively smaller amplitude bias than previous two versions (Fig. 9.11, top).

9.3.4 Sea Ice

The Arctic sea ice area (Fig. 9.12) in P5 is overall quite similar to that of P4. The most notable differences are found during the low ice season, when P5 is lower than P4, and even closer to the estimate derived from NASA-Team.

9.3.5 Ocean Weeks 3&4

SST skill (anomaly correlation to OSTIA) for Weeks 3&4 forecasts are the highest (~0.9°C) in the ENSO active region of central-eastern tropical Pacific extending along the US west coast in the northern and southeastward in the southern hemisphere (Fig. 9.13, top). Correlations are comparatively weaker in western boundary regions and moving poleward. Compared to the Pacific, SST skill in the Indian Ocean and the Atlantic is lower. Zonal wind stress ($\boldsymbol{\tau}_x$) forecasts are most skillful in the central-western equatorial Pacific, just west of the region of highest SST skill (Fig. 9.13, bottom). The Weeks 3&4 forecasts of Niño 3.4 index (ENSO index) and the Dipole Mode Index (IOD index) agree closely with OSTIA (Fig. 9.14). The Niño 3.4 SST anomaly in P5 is too strong for the 2015/16 El Niño, and 2017/18 La Niña events.

Overall, warm SST bias mostly dominates the tropics, with prominent biases in the Indo-Pacific warm pool region, the ITCZ, coastal South America in the Pacific, and the eastern coastal Atlantic. A warm (cold) SST bias is generally associated with a positive (negative) net radiation bias, and is in regions with shallow MLD. The MLD is in general shallower in the eastern and western tropical boundaries, and in the tropical Indian Ocean, however it deepens moving poleward into the extra-tropics (Figs. 9.19, 9.20).

The overlying $\boldsymbol{\tau}_x$ is generally weak in P5, with overly weak easterlies in the tropical Pacific, tropical Atlantic, and southern tropical Indian Ocean, and overly weak westerlies in the equatorial and northern Indian Ocean (Figs. 9.22, 9.23). The weaker $\boldsymbol{\tau}_x$ could in turn result in a weaker off-

equatorial wind stress curl, which reduces the Ekman upwelling at the equator and the poleward Sverdrup transport, and relaxes the zonal slope in equatorial thermocline (D20).

The changes in equatorial D20 are prominent in regions where it is shallow. In general, D20 is shallower in the eastern boundary regions, deepening westward, and outcropping around 40°N/S. The equatorial zonal slope in D20 is prominent in the Pacific, much reduced in the Atlantic, and minimal in the Indian Ocean. Due to sparse daily MLD and D20 observational coverage spanning 2011–2018, assessment of biases in Weeks 3&4 forecasts of MLD and D20 in this analysis was not possible.

The SST bias is more pronounced in the summer hemispheres, *i.e.*, OND for the southern hemisphere (Fig. 9.15) and MJJ for the northern hemisphere (Fig. 9.16). The warmer SST biases are consistent with larger positive net radiation biases in the summer hemisphere (Figs 9.17, 9.18) over regions with shallow MLD. In OND (Fig. 9.15), the warm bias affects most of the southern tropics, and the coasts of the eastern Pacific and eastern Atlantic. In MJJ (Fig. 9.16), the warm bias is seen along the US west coast and in the western equatorial Pacific warm pool, where convection is prominent in observations, and underestimated in the model (Fig. 9.7).

9.3.5.1 Pacific Ocean

In the Pacific the cold tongue SST bias in the eastern equatorial region—a region of persistent climatological bias across GCMs being a primary obstacle to simulating ENSO (Guilyardi *et al.*, 2009; Vecchi and Wittenberg, 2010) — is seasonally enhanced in OND, and associated with an increase in negative radiation bias, indicating a likely increase in excess cloud cover reducing the incoming radiative heating of the surface (Figs. 9.15, 9.17). The $\boldsymbol{\tau}_x$ bias in OND is much smaller than in MJJ (Figs. 9.22, 9.23), particularly at the equator, indicating the less-likely role of the equatorial D20 (which is strongly tied to the equatorial and off-equatorial wind stress) in contributing to the enhanced cold tongue bias in OND.

Moving further eastward along the eastern coastal regions of the equatorial Pacific east of 80°W, a warm SST bias persists across seasons, and also extends southward along the coastline (Figs. 9.15, 9.16). The strong diurnal variations in this region of shallow MLD is a challenge to simulate in the ocean model, which affects the SST simulations. Thus, the warm SST bias could as well be from the ocean initial conditions (CPC-3DVar), which is known to have biases in the MLD of this region.

Moving slightly north, the warm bias along the ITCZ increases from OND to MJJ (Figs. 9.15, 9.16), inconsistent with the decrease in net radiation bias (Figs. 9.17, 9.18), indicating the possible role of latent and sensible heat flux in the SST bias of the region.

The decrease in negative radiation bias from OND to MJJ in the southeastern Pacific is not consistent with an increase in cold SST bias. Although the excess stratus cloud deck bias in the region likely decreases seasonally in MJJ, it is not enough to reduce the SST cold bias. Other factors, such as alongshore wind stress driving off-shore Ekman transport and wind stress curl

controlling the strong coastal upwelling along the Peru coast also contribute to the SST variations here.

9.3.5.2 Atlantic Ocean

In the Atlantic a warm bias along the eastern coastal region is prominent, enhancing slightly in OND (Figs. 9.15, 9.16). A net positive radiation bias in OND just north of the equator warms the surface (Fig. 9.17), where the MLD deepens sharply from east to west from ~10m to 60m (Fig. 9.19). However with a much flatter MLD in MJJ (Fig. 9.20), the warm SST bias reduces with much reduced bias in net radiation (Figs. 9.16, 9.18).

The cold SST bias in the western Atlantic is associated with a negative bias in net radiation over a deeper MLD in OND, but persists in MJJ in spite of an increase in net radiation (bias turns positive) over a shallower MLD. This suggests a deficiency in the model's seasonal generation of warm core rings along the Gulf Stream which would typically last for a week or more and act to warm the surface.

In the eastern Atlantic a warm SST bias in MJJ is also in contrast to a net negative radiation bias. Extra-tropical westerlies (which have a weak bias) strengthen off the US northeast in the western Atlantic in OND, while easterlies expand in the eastern Atlantic in MJJ, which could induce seasonal changes in the surface ocean (Figs. 9.22, 9.23).

Moving south of the equator, a net radiative cool bias does not seem to drive the warm SST bias in OND, but a cold SST bias in a seasonally deeper MLD appears to be consistent with negative radiation bias in MJJ (Figs. 9.16, 9.18, 9.20).

9.3.5.3 Indian Ocean

The warm SST bias in the equatorial Indian Ocean (IO) increases from MJJ to OND (Figs. 9.16, 9.15), with a slight increase in net positive radiation bias in the eastern equatorial basin (Figs. 9.18, 9.17). The warm bias in the east is an extension of the Indo-Pacific warm pool bias, where the atmospheric convection and coupled air-sea interaction play a dominant role in driving the SST pattern. The Indian summer monsoon driven by land-sea temperature contrast controls the seasonal wind pattern, likely influencing the SST variations in the tropical IO as well. Thus in addition to radiation biases, wind biases likely play an additional role in driving the net ocean heat uptake and hence the SST bias.

During the monsoon season, westerlies dominate the northern IO in MJJ, driving southwesterly winds into the Indian subcontinent (Fig. 9.23). This likely cools the sea surface by deepening the MLD (30–40m) and mixing the upper ocean well (Fig. 9.20). However, a weak bias in the westerly winds in P5 mostly reduces the sea surface cooling (Fig. 9.23), driving a warm SST bias in MJJ (Fig. 9.16). Unlike in OSCAR, where strong eastward surface currents are seen in MJJ, a westward surface current in P5 keeps the relatively colder western equatorial waters from spreading eastward, thereby enhancing the warm SST bias.

During OND, the westerlies weaken and are pushed equatorward (Fig. 9.22), driving a basin-wide MLD shoaling (20–30m) warming the upper ocean (Fig. 9.19). However, along the equator, the meridional width of these equatorial westerlies in P5 are too narrow and weak compared to ERA5 (Fig. 9.22). Thus although the wind-driven equatorial D20 deepens in the east, it has a shallow bias. Hence an increased ocean heat uptake driving the warm SST bias in OND is not clearly evident, suggesting there are other basin-wide processes involved that remain to be explained. However, the overly strong eastward surface currents in OND likely push warmer waters eastward, contributing to the warm SST bias.

10. Prototype 6 (P6)

10.1 Configuration and Settings

10.1.1 Fractional grid

In the preceding prototypes, P1–P5, each atmospheric grid cell consisted of either 100% land or 100% water, depending on which of those two was spatially dominant; water cells could be partially covered with ice. Starting with P6, a fractional grid capability has been introduced which allows partial land, open water, and sea ice areas to coexist in a single atmospheric cell, as shown in Fig. 10.1. Thus, property fluxes through the bottom of an atmospheric grid column are now evaluated as weighted averages of fluxes over land, open water (including lakes), and sea ice. Compositing weights reflect the extent to which an atmospheric grid cell is covered by the various surface types. Weighting by fractional surface area is also used for state variables like roughness length or sea surface temperature.

Partial land is allowed to exist on the atmospheric grid if the land fraction is greater than 0.01%. Grid cells in the ocean model continue to be either 100% water or 100% land.

If the ocean and atmosphere models reside on different grids (which is the case here), a fractional grid permits better matching of the coastline on the two grids and, importantly, makes it possible to rigorously conserve flux quantities during air-sea coupling. No such conservation can be achieved by the nearest-neighbor mapping scheme used in the previous (non-fractional) model prototypes.

In order to achieve conservation (*i.e.*, global/basinwide preservation of surface integrals) of fluxes passed through the mediator, fluxes must always be "fraction-weighted" on the *source* grid before being passed to the target grid. To illustrate the difference between unweighted and fraction-weighted fluxes, consider rainfall rate (in mm/hr) versus water mass falling into a specific area (in kg/hr) as an example. It is the latter quantity whose global integral the atmospheric and oceanic submodel must agree on.

In the case of air-sea fluxes over open water, fraction-weighting is a two-step procedure because these fluxes are typically computed in the PBL module residing on the atmospheric grid, whereas the ice-free fraction of a grid cell is determined by the sea ice model residing on the ocean grid. To resolve this, the ice-free fraction must first be passed through the mediator to the atmosphere before the fraction-weighted open-water flux can be computed and passed to the ocean model.

This situation does not arise with fluxes over sea ice passed to the atmosphere because the ice model computes those fluxes (sensible/latent/momentum flux, longwave radiation, reflected sunlight) and ice fraction on the same grid. Also, no conservation issues arise with fluxes over land and lake surfaces as long as the land/lake models reside on the atmospheric grid.

By using subscripts *lnd, opw, ice* to denote land, open water, and sea ice, respectively, and indicating by *a2o(x)* and *o2a(x)* the passage of an argument "*x*" through the mediator in atmosphere-to-ocean and ocean-to-atmosphere direction, respectively, we can express the above considerations in the form:

$$w_{atm} * F_{atm} = w_{lnd} * F_{lnd} + \left[o2a\left(w_{opw}\right) * F_{opw}\right] + o2a\left(w_{ice} * F_{ice}^{top}\right)$$

$$w_{ocn} * F_{ocn} = a2o[o2a\left(w_{opw}\right) * F_{opw}\,] + w_{ice} * F_{ice}^{bot}$$

Note that the weights w_{lnd}, w_{opw}, w_{ice} in these expressions are neither normalized nor dimensionless but represent fractional grid areas. Quantities w_{atm} and w_{ocn} stand for cell size on the atmospheric and oceanic grid, respectively. Superscripts *top and bot* are introduced to distinguish ice surface from ice bottom fluxes. The portion of the incoming radiation that does not become part of the open-water flux is fraction-weighted appropriately—i.e., multiplied by $o2a(w_{ice})$—before being directed to the ice model. This detail is omitted from the above expressions to reduce their visual complexity.

Most fractional variables (surface upward potential latent heat flux, kinematic surface upward latent heat flux, surface specific humidity, snow depth and its liquid equivalent, kinematic surface upward sensible heat flux and soil heat flux, surface and skin temperature, surface exchange coefficients, bulk Richardson number, stress, friction velocity, and surface stability parameters) in P6 are composited by a simple arithmetic average, *i.e.*,

$$w_{atm} * X_{comp} = w_{lnd} * X_{lnd} + w_{opw} * X_{opw} + w_{ice} * X_{ice}.$$

The exception are the kinematic and thermodynamic roughness lengths ("*zorl*" and "*ztmax*") which are composited by a geometric mean,

$$Z_{comp} = exp\left[\{w_{lnd} * ln(Z_{lnd}) + w_{opw} * ln\left(Z_{opw}\right) + w_{ice} * ln(Z_{ice})\}/w_{atm}\right]$$

A geometric mean is appropriate in this case as roughness length spans several orders of magnitude between land and open water.

A separate lake mask defines the extent to which an atmospheric cell is covered by lakes. In P6, the global lake mask is derived from GLDBv2 (Choulga *et al.*, 2014). For historical reasons, the lake mask is temporarily taken as an integer. If less than 50% of a grid cell is covered by one or more lakes, its lake fraction is set to 0, *i.e.*, the cell is treated as 100% land. Otherwise the cell becomes an all-lake cell. Note that the 50% threshold is arbitrary and can be replaced by another value. In the interest of simplicity, no lakes are presently allowed in atmospheric cells partially covered by ocean.

10.1.2 Vertical Resolution and Physics Updates

Starting with P6, the atmospheric model top is extended from 54 km to 80 km, and the number of vertical layers is increased from 64 to 127. The atmospheric physics package is updated from GFSv15.2 (used in P1–P5) to GFSv16. The GFSv16 Physics includes the following additions and updates:

- PBL/turbulence: replaced Hybrid EDMF (Han et *al.*, 2016) with sa-TKE-EDMF (scale-aware turbulent kinetic energy-based EDMF; Han and Bretherton, 2019), and revised background diffusivity as a stability-dependent function;
- Gravity wave drag parameterization: added a parameterization for subgrid scale nonstationary gravity wave drag;
- Radiation: updated calculation of solar radiation absorption by water clouds, and updated cloud overlap assumptions;
- Microphysics: updated GFDL microphysics scheme for computing ice cloud effective radius;
- Noah LSM: revised heat flux calculation over snow covered surface, and introduced vegetation impact on surface energy budget over urban areas.

10.2 Initial Conditions

With the introduction of a fractional atmospheric grid, additional surface variables were introduced to allow for both temperature and roughness length to have distinct, separate values over land and water. The separation of value over land and ocean—starting with the initial conditions and continuing throughout the forecast—is limited to these two variables as they have the greatest difference between land and water, as shown in Fig. 10.2. Due to the changes in the land mask and the vertical resolution in the atmospheric model, the atmospheric initial conditions used in P6 were re-generated from CDAS1.

10.3 Evaluation

10.3.1 Benchmark AC Scores

The AC scores of P6 for Week 1 and Week 2 for most considered fields are similar to those of P5 (Fig. 10.3). Global tropical SST scores are slightly lower starting with Week 2. The most

pronounced difference in scores between the two prototypes is for the CONUS T2m at Week 3, and 3&4. For these lead times P5 had a marked degradation of score relative to its preceding prototype, which was hypothesized to be related to the lack of lake ice in P5 and consequently a warm lake temperature bias. The lake ice in P6 has been corrected to allow for non-zero values, which may explain the improvement in forecast skill. It should be noted, however, that in P6 (and in the following prototype, P7) lake ice does not change over the length of an individual forecast.

10.3.2 Annual Mean Biases

The large-scale pattern of annual mean SST bias is similar between P6 and P5 (Fig. 10.4), but some prominent differences become more apparent with increasing lead time. There are warmer SSTs in the eastern equatorial Pacific and eastern Equatorial Atlantic (representing exacerbation of the warm bias), and cooler SSTs in the western equatorial Pacific/Indian Ocean (representing an amelioration of the warm bias). The annual mean surface temperatures of regions affected by sea ice in both hemispheres are cooler in P6 than in P5.

The annual T2m has a reduced warm bias at high latitudes, especially in the northern hemisphere (Fig. 10.5). Warm bias is also reduced over the western tropical Pacific and Indian Oceans, consistent with the changes in SST. Also consistent with the SST changes is the increased warm bias in the eastern Pacific and eastern Atlantic basins. The change in T2max (Fig. 10.6) is smaller in magnitude than the change in T2min (Fig. 10.7), *i.e.*, the reduced T2m warm bias at high latitudes predominantly stems from lowering of T2min.

Occasional spikes of very high T2m (not shown) appear on some coastal grid points. This behavior in P6 is attributed to the fractional grid compositing of some variables that, as was discussed earlier, are calculated as simple arithmetic averages in P6. After updating the compositing methodology, these spikes are no longer present.

The large-scale pattern of annual mean precipitation bias is broadly similar between P6 and P5 (Fig. 10.8) but there are pronounced differences in the western equatorial Pacific and Indian Ocean (wetter), and in the eastern Pacific just north of the equator, the Amazon region, and equatorial Africa (drier).

The positive annual mean OLR bias (Fig. 10.9) is reduced in P6 compared to P5. The reduction of bias is prominent in the Indian Ocean, Maritime Continent, and eastern tropical Pacific Ocean.

The mean differences between P5 and P6 Weeks 3&4 forecasts initialized in winter (DJF) and summer (JJA) for various fields are shown in Figs. 10.10 and 10.11 to illustrate the season-dependent impact of model changes.

10.3.3 MJO

The RMM1 anomaly correlation remains very similar between P4, P5 and P6 (Fig. 10.12, top). All three prototypes' RMM1 AC remains above 0.6 until roughly forecast day 14. However, in P6, the

AC skill score remains close to 0.6 until forecast day 17 before it decreases further, indicating slower propagation of MJO. P6 loses about one day in RMM2 skill as AC drops to 0.6 on day 17.5 compared to forecast day 19 in P4 and P5. In terms of the combined RMM, both P5 and P6 show degradation by one to one and a half days compared to P4. In terms of component variables (Fig. 10.12, bottom), P6 shows decrease in AC compared to P4 and P5 for both U850 and U200 at longer lead times, and for OLR at all leads.

Among the P4, P5 and P6, there is a slightly larger RMSE in P6 than P4 and P5, mostly contributed by OLR (Figs. 10.13, 10.14). Overall, MJO tends to be forecasted stronger and slower (Fig.10.15). Initially strong (i.e. amplitude of the MJO from analysis at the initial time > 1) MJOs tend to be forecasted weaker and move slower than initially weak (amplitude of the MJO from analysis at the initial time < 1) MJOs in early lead time. The amplitude error in P6 tends to slightly better than P5 for initially strong MJO (up to ~day 24), and worse than P5 for initially weak MJO. The phase error in P6 tends to be better than in P5 for initially strong MJO, but worse than in P5 for initially weak MJO. OLR is the component with the most noticeable RMS error increase compared to P5.

10.3.4 Sea Ice

The forecast Arctic ice area has changed very little between P5 and P6 (Fig. 10.16) and remains close to the observational estimate derived from NASA-Team ice concentrations and below the NOAA/NSIDC-CDR -based estimate.

10.3.5 Ocean Weeks 3&4

The P6 SST skill in Weeks 3&4 forecasts very closely resembles that of P5 (Fig. 9.13, top). $\boldsymbol{\tau}_x$ skill drops in the tropical Atlantic, tropical Indian Ocean and southeastern Pacific. However, a slight reduction in $\boldsymbol{\tau}_x$ bias is noticeable in the tropical–extratropical regions of the north Pacific, and central-western equatorial Pacific (Fig. 9.13, bottom). A slight decrease in the too strong Niño 3.4 SST anomaly for the 2015/16 El Niño and the 2017/18 La Niña event is also noticeable (Fig. 9.14).

The SST bias pattern is very similar to that in P5. Regions of warm SST bias are accompanied by an increase in net radiative positive bias in regions of shallow MLD. The narrow band of eastern coastal warm bias (along SE Pacific and SE Atlantic) are associated with a positive net radiation bias in a shallower MLD. However, negative radiation bias seems to drive the cold SST bias further away from the shore in SE Pacific, with a gradual deepening of the D20. Weak easterlies dominate the tropics, and weak westerlies dominate the extratropics. However, the $\boldsymbol{\tau}_x$ in the IO show minimal bias compared to P5 and P7.

The seasonal pattern of SST biases remains the same as in P5, with most of the changes in P6 focused on the equator, and along the eastern coastal regions of the southern hemisphere, in tandem with changes in both the net radiative heating and MLD. Warm SST bias is amplified in

the summer hemispheres and mostly associated with a net increase in positive radiation bias from P5 to P6.

10.3.5.1 Pacific Ocean

In the Pacific SST bias is seen to increase in specific regions from P5 to P6. For example, the eastern coastal warm bias increases, particularly in OND (Fig. 9.15), and is mostly from an increase in net radiative heating biases (Fig. 9.17). The warm bias along the US west coast also increases and is amplified on merging with the meridionally expanded warm bias along the ITCZ in MJJ (Fig. 9.16). However, although this warm bias along the US west coast could be attributed to an increase in net radiation positive bias from P5 to P6, but not for the warm bias along ITCZ (Fig. 9.18), indicating again the likely role of latent and sensible heat fluxes.

We now focus on regions where the SST bias decreases from P5 to P6. The western Pacific warm pool SST bias decreases and is most likely due to improved net radiative fluxes. The eastern equatorial cold tongue bias in the Pacific also decreases due to reduced negative radiation bias particularly in OND (Figs. 9.15, 9.16). However the negative radiation bias in the eastern equatorial cold tongue in P6 extends too far west in OND. D20 shows a slight shoaling along the eastern coastline and a slight deepening around the Galapagos Islands from P5 to P6, particularly in this season (Fig. 9.21). A wide stretch of D20 shoaling from the eastern coastline westward in MJJ from P5 to P6 indicates reduced upper ocean heating, inconsistent with a warmer SST. Hence changes in SST bias from P5 to P6 are driven mostly by radiative changes and not due to equatorial D20 changes.

The westward equatorial surface current shows spatial asymmetry from P5 to P6, strengthening in the far east and weakening in the west in OND (Fig. 9.24). The shear generated between westward flowing North Equatorial Current (NEC) and eastward flowing North Equatorial Countercurrent (NECC) strengthens from P5 to P6 in OND, as in observations (OSCAR). The most interesting improvement in surface currents from P5 to P6 is perhaps the weakening of eastward Equatorial Countercurrent (ECC) in MJJ, much like that in OSCAR.

In the equatorial subsurface, the eastward flowing Equatorial Undercurrent (EUC) weakens slightly (season of strengthened EUC) along with a strengthening of South Equatorial Current (SEqC) from P5 to P6 in MJJ (Fig. 9.25). This likely enhances the vertical shear driven mixing in the east from P5 to P6 where EUC shoals along D20, that is shallower in P6. However, in OND with a seasonally deeper but weaker EUC core, westward currents dominate above D20 in the equatorial subsurface although weaker than P5. However in the far east SEqC is still strong in P6, driving waters off-shore. This shoals D20 from P5 to P6 and cools the equatorial subsurface in central-eastern equatorial Pacific and east of the Galapagos Islands, particularly in MJJ.

10.3.5.2 Atlantic Ocean

A warmer SST in the equatorial Atlantic is consistent with a deeper D20 from P5 to P6 (Figs. 9.15, 9.16), suggesting a role of ocean heat uptake in SST warming. Seasonally, an increase in SST warm bias in OND (Fig. 9.15) is associated with increase in net positive radiation bias (Fig. 9.17),

in a shallower MLD (Fig. 9.19) and a deeper of D20 (Fig. 9.21) from P5 to P6, indicating role of both net radiation and upper ocean heat uptake. But an increase in SST warm bias in MJJ (Fig. 9.16) is accompanied with a deeper D20 (Fig. 9.21) only, and minimal changes in net radiation and MLD (Figs. 9.19, 9.20), indicating upper ocean heating mostly contributes to the SST warming in P6.

The zonal wind stress, τ_x, in general does not change much from P5 to P6 in the Atlantic basin, except for a slight strengthening in the north tropical Atlantic in MJJ (Fig. 9.23). The surface currents do not change much except for slightly weaker NEC and NECC across the equatorial Atlantic from P5 to P6 (Fig. 9.24). The equatorial subsurface (upper 100m) warms P5 to P6 by at least 0.8°C, in particular in the central-eastern basin (Fig. 9.25). The warmer surface connects to the warmer subsurface and expands in MJJ from P5 to P6. The westward current in the upper 80m weakens and the deeper EUC strengthens from P5 to P6, irrespective of seasons (Fig. 9.26).

10.3.5.3 Indian Ocean

The tropical Indian Ocean (IO) shows much improvement in SST bias from P5 irrespective of the season. This is accomplished by both a decrease in positive net radiation bias accompanied with a deeper ML, and a deeper D20 from P5 to P6 driven by improved τ_x variations.

The equatorial westerlies are latitudinally too broad in OND, resulting in a westerly τ_x bias just south of the equator (Fig. 9.22). τ_x in the eastern part of the tropical IO shows least bias, compared to the weaker westerlies in the west in the northern IO, and too strong easterlies in the west south of the equator. At the equator the westerlies are very well simulated showing minimal bias compared to ERA5. In OND, an increase in equatorial westerlies from P5 to P6 likely drive a deeper eastern equatorial D20 in the IO (Figs. 9.22, 9.21). However the most obvious improvement in the equatorial westerlies are seen in MJJ, when P6 captures the zonal pattern of the westerlies much more efficiently than P5, and deepens the D20 by more than 10m in the east, thereby warming the surface (Figs. 9.23, 9.21).

The weaker westerlies in the northern IO are also strengthened and are much closer to ERA5 in P6 in MJJ indicating an improved monsoonal winds in the Indian subcontinent. Thus overall, an improved τ_x in P6 plays an important role in reducing the warm SST bias of the IO in P6.

Surface currents are also improved in P6, i.e. not turning westward in MJJ as in P5 (Fig. 9.24) and eastward currents strengthening from P5 to P6 in OND. This is closer to that seen in OSCAR, where seasonally eastward current strengthens in OND, but only weakens in MJJ.

The equatorial subsurface warms from P5 to P6 by ~1°C, most likely from strengthened equatorial westerlies deepening the equatorial D20 (Fig. 9.25). Subsurface currents above D20 are predominantly eastward in OND, and westward in MJJ. The eastward subsurface current strengthens from P5 to P6 in OND, but the westward current in MJJ turns eastward in P6 (Fig. 9.26).

10.3.6 Impact of SST bias on CONUS temperature and precipitation

Stan et al. (2022), using regression analysis, found a significant relationship between the biases of tropical Pacific SST and CONUS precipitation and temperature bias in P5. The impact of the SST biases was shown to be growing with lead week. The connection between SST biases and CONUS forecast skill was found to be less direct, and more pronounced in the forecasts for temperature than for precipitation.

The evaluation of impact of tropical Pacific SST biases on the forecast skill of CONUS precipitation in P5 and P6 shows only small differences between the two prototypes. Despite these small differences, the impact of SST biases on the precipitation biases can be quite large. An example is shown in Fig. 10.17, which compares the precipitation bias in Week 3 during boreal summer (JJAS) along with the influence of the tropical SST bias. Tropical SST bias is represented as the time series of the area average SST bias over the region 120°E–180° and 10°S–10°N where the largest SST bias is located (See Fig. 9.16). The region with relatively small precipitation bias in the northwestern U.S. seen in P5 (Fig. 10.17a) expands and increases in P6 (Fig. 10.17b). The influence of tropical SST bias on the CONUS precipitation bias measured by the regression coefficient, shows an increase of the SST bias influence in P6 (Fig. 10.17d) relative to the influence seen in P5 (Fig. 10.17c).

11. Prototype 7 (P7)

11.1 Configuration and settings

11.1.1 Surface datasets

For P7, the orographic datasets are updated to allow fractional lakes to exist on the grid if the lake fraction is greater than 20%. A new VIIRS-based surface type climatology is used to define the vegetation categories for the land grids based on the NESDIS Surface Type product (https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/ATBD_VIIRS-SurfaceType_v2.0.pdf). A “climatological” surface type dataset was created by NESDIS and UMD collaborators (Xiwu Zhan and Chengquan Huang) using the NESDIS VIIRS Surface Type product from 2012 to 2019 (Fig. 11.1, left). Additional tundra categories are added using Köppen-Geiger climate classifications (Fig. 11.1, right). New surface fixed fields are then created to be consistent with the new orographic and surface type datasets.

The old coarse resolution 5° x 5° Optical Properties of Aerosols and Clouds (OPAC) data are replaced by 0.5° x 0.625° MERRA-2 aerosol climatology (Randles *et al.* 2017). The MERRA-2 aerosol has more reasonable sea salt, sulfate, and dust distribution over North America in addition to global. For implementation analysis, see Cheng and Yang (2022).

11.1.2 Model Physics Updates

Land Model Changed to Noah-MP: The land surface model (LSM) is changed from the Noah LSM to the Noah-MP LSM in P7. Noah-MP is a community-based model that was developed specifically to address known limitations of the Noah LSM and to use multiple options for key land-atmosphere interactions. It improves upon Noah in multiple ways. For example, it uses (1) a tiled approach to separate vegetation and bare soil, (2) Ball-Berry stomatal resistance related to photosynthesis, (3) a dynamic vegetation scheme, (4) a multi-component, separate vegetation canopy, (5) a two-stream radiative transfer approach along with shading effects necessary to achieve proper surface energy and water transfer processes including under-canopy snow processes, (6) a multi-layer snow pack with liquid water storage and melt/refreeze capability and a snow-interception model describing loading/unloading, melt/refreeze capability, and sublimation of canopy-intercepted snow, (7) multiple options for surface water infiltration and runoff and groundwater transfer and storage including water table depth to an unconfined aquifer. These additional features have the potential to significantly improve the realism of the land surface, hydrology, and atmosphere interactions.

Gravity Wave Physics Updated: The Unified Gravity Wave Physics (UGWP) implemented in P7 consists of a suite of drag parameterizations that represent the drag forces imparted on atmospheric flow due to subgrid scale (unresolved) orographic and non-stationary sources. The orographic drag parameterizations consist of 1) a gravity wave drag parameterization, which represents the vertical propagation and breaking of topographic gravity waves throughout the free atmosphere from just above the planetary boundary layer (PBL) up to the mesosphere; 2) a low-level mountain blocking scheme; 3) a small-scale gravity wave drag parameterization that represents gravity wave drag produced by waves induced within highly stable, typically nocturnal PBLs by small-horizontal-scale topography; and 4) a turbulent orographic form drag scheme that represents the effects on PBL turbulence by small-horizontal-scale topography. These orographic drag schemes were part of the “GSL” drag suite that is currently implemented in the RAP/HRRR regional NWP models. These have been joined with the non-stationary gravity wave drag scheme of Yudin *et al.* (2020), which represents convective and other non-stationary sources of gravity wave drag, to form the “unified” gravity wave physics suite used in P7.

Convection, PBL and surface layer updates: Significant model physics updates to cumulus convection, vertical turbulent mixing and surface layer physics are included in P7. The updates aim mainly to enhance the underestimated surface-based convective available potential energy, to reduce the tropospheric cold biases and the planetary boundary layer (PBL) dry biases over the Tropics, to reduce the excessive vertical turbulence mixing in strong shear environment as well as the excessive PBL growth, to better predict the surface inversion as well as the capping inversion near the PBL top, to reduce the nighttime cold and daytime warm T2m biases over forested regions, and to better represent sub-grid scale turbulence variability in the surface layer. The updated physics not only showed a significant improvement in Z500 anomaly correlations in uncoupled simulations, especially for a forecast lead time of 48–96-h and in the tropical wind vector root mean square errors, but it also improved a Madden-Julian Oscillation intensity and

propagation especially for longer lead time. For details, see Han *et al.* (2021). A summary of the changes are listed below:

- Cumulus convection
    - Stricter convection trigger
    - Reduced entrainment rate below cloud base
    - Enhanced downdraft detrainments starting from 60mb above the ground surface
    - Reduced rain evaporation
    - Modification of cloud depth separating shallow convection from deep convection
- PBL and surface layer
    - Inclusion of wind shear effect reducing characteristic mixing length
    - Reduction of background diffusivity in the inversion layers as a function of surface roughness and green vegetation fraction
    - PBL updraft overshooting limited by bulk Richardson number-based-PBL depth
    - Inclusion of new canopy heat storage parameterization
    - Modification of thermal roughness calculation over land
    - Increase of momentum roughness length over sea
    - Inclusion of sea spray effect parameterization
    - Inclusion of new maximum z/L parameterization
    - The constraint of stability parameter z/L is applied over ocean and sea ice.
    - Introduction of a seasonally varying momentum roughness length
    - Correction of the diagnostic stability parameters

Cellular automata convective organization introduced: A novel method for representation of sub-grid, and cross-grid, convective organization, based on cellular automata (CA), is introduced into P7. A CA is a mathematical model used to describe self-organizing behavior in physical systems exhibiting birth/death processes, and it is used here to address the effect of convective organization in a bulk-plume cumulus convection parameterizations (saSAS), where this type of organization has to be represented in terms of how the resolved flow would “feel” convection if more coherent structures were present on the subgrid. Since the model physical processes are parameterized in columns, and the horizontal interaction between physical processes takes place only through advection and diffusion, the addition of the CA offers a unique approach to organize convection across adjacent grid-boxes by enhancing the probability of triggering of convection in nearby areas, representing subgrid fluctuations in temperature and humidity, and triggering in premoistened regions if convection is triggered in a cluster. The stochastic nature of the CA may enhance organization in different directions within the grid-box, and across grid-boxes, depending on the initial seed. If the model is run as an ensemble, the convection scheme's stochastic triggering function can help to improve uncertainty estimates associated with subgrid fluctuations of temperature and humidity and randomness in organization. In P7, model grid boxes in which the CA's largest connecting plume exceeds a given threshold will be considered as candidates for convective activation, in addition to saSAS’s current triggering criteria. More details including the scientific outline of the scheme, the CA rulesets explored, and the impact on convectively coupled equatorial waves can be found in the following references: Bengtsson *et al.* (2013), Bengtsson *et al.* (2019), and Bengtsson *et al.* (2021).

Mushy Thermodynamics for Sea Ice: The mushy thermodynamics scheme is implemented in the CICE6 model. This changes the sea ice salinity from an assumed fixed salinity profile to one in which salinity evolves. Sea ice is modeled as a mushy layer which is composed of brine inclusions and pure ice. Both sea ice salinity and enthalpy become prognostic variables.

Near-Surface Sea Temperature Model Added: The NSSTM (Near-Surface Sea Temperature Model) is turned on in P7. The NSSTM simulates ocean skin layer diurnal warming and cooling profiles, and calculates a temperature (NSST) profile just below the ocean surface. The NSST profile is the sum of the foundation temperature and the balance of the diurnal warming and cooling. The foundation temperature is calculated by NSSTM as the ocean model's top layer temperature corrected by the vertically integrated balance of diurnal warming and cooling profiles. The surface value of the NSST profile is used as the surface temperature for atmospheric flux calculations in P7. In previous prototypes, this role was played by the temperature of the top ocean layer, considered as the SST.

Wave Model Updated: The wave, ice and ocean models were all updated to their latest versions.

Fractional Grid Compositing: To resolve the issue of sporadic spikes of very high temperatures encountered on some coastal grid points, the methodology for fractional grid compositing for some types of variables is revised in P7.

The compositing of surface skin temperature variables “tsfc” and “tsurf”, which was done by simple arithmetic averaging in P6, is revised in P7 to become a flux-weighted averaging, namely

$$T_{comp} = \frac{w_{lnd} * Ch_{lnd} * \ T_{lnd} + w_{opw} * Ch_{opw} * T_{opw} + w_{ice} * Ch_{ice} * T_{ice}}{w_{lnd} * Ch_{lnd} + w_{opw} * Ch_{opw} + w_{ice} * Ch_{ice}}$$

In the above expression, subscripts *lnd, opw, ice* denote land, open water, and sea ice, respectively, w represents the fractional grid area of each component, and Ch is the surface heat exchange coefficient (separate over each surface type).

The fractional grid compositing for variables such as surface exchange coefficients, bulk Richardson number, stress, friction velocity, and surface stability parameters is also no longer calculated as a simple arithmetic average, as was done in P6. The composite values of these fields in P7 are computed by taking the composited values of surface roughness and surface temperature and making an additional call to a subroutine called “stability” which ensures that in cases where different surface types span different stability regimes (*e.g.,* unstable over land but stable over open water), a single and physically consistent stability profile is computed.

11.2 Initial Conditions

Atmospheric initial conditions are provided from GEFSv12 Reanalysis (Hamill *et al.*, 2022), which was run on the same horizontal resolution (C384). Wave initial conditions are created by running

eight streams of spin-up runs forced by GEFSv12 Reanalysis winds with a 6 hour time frequency. Each initial condition has at least a one-month spin-up to ensure that the swell has fully developed. Land initial conditions are provided from a land-only simulation using a combination of Global Soil Wetness Project (http://hydro.iis.u-tokyo.ac.jp/GSWP3) and GDAS atmospheric forcing. The Noah-MP land model was executed on the C384 FV3 grid beginning in 2000. Soil temperature, soil liquid water content and soil total water content were then transferred to surface initial files. Land model snow states are provided from GEFSv12 Reanalysis.

11.3 Evaluation

11.3.1 Benchmark AC Scores

The benchmark anomaly correlation scores for Week 1 in P7 are similar to P6 for most considered fields, and slightly higher for southern hemisphere Z500 and Niño 3.4 precipitation (Fig. 11.2). Note that scores for SST (tropical and ENSO) for P7 are calculated using the NSST temperature at the surface, instead of the actual SST, and are thus somewhat inaccurate; however, they remain similar to preceding prototypes. The AC scores of CONUS T2m and precipitation decay faster with lead time—the Week 3 scores in P7 are as low as the week 4 scores in P6. The AC scores of the southern hemisphere Z500 also decays faster with lead time; the southern hemisphere Z500 scores are, by contrast, higher for Week 1 and Week 2, but lower for Week 3 and Weeks 3&4.

On balance, the Weeks 3&4 AC scores in P7, with the exception of Niño 3.4 precipitation, are amongst the lowest of all prototypes.

11.3.2 Annual Mean Biases

The surface temperature (Fig. 11.3) seen by the atmospheric model (defined with NSST in P7, SST in P6) is cooler over the ocean from 30°S to 30°N, and over regions with sea ice. The difference in surface temperature over tropical oceans is on the order of −0.5°C; this difference does not increase with lead time and is more pronounced in the winter hemisphere. Offline experiments demonstrate that this relative cooling is a direct result of the NSST correction. The cooling over sea ice is larger in magnitude, up to −5°C, and increases with lead time. This cooling is likely at least in part a consequence of the activation of Mushy thermodynamics which leads to overall larger sea ice concentration and thickness in P7 compared to P6 (Figs. 11.9 and 11.10, *icec* and *icetk*), consistent with the findings in long-term simulations with Mushy TD in CESM (Bailey *et al.*, 2020).

The annually averaged T2m (Fig. 11.4a) over ocean are also lower in P7 compared to P6, leading to a slight negative bias with respect to the CFSR reanalysis. Note however that relative to the ERA5 reanalysis, whose T2m over ocean are higher than the CFSR (not shown), P7's lower T2m represents a reduction of bias.

Over land, the annually averaged T2m (Fig. 11.4a) show a large increase of warm bias for most regions. The erroneous warming is particularly pronounced in northern Eurasia, northern North America, Greenland, Antarctica, and the humid tropics of Africa and South America. Desert regions on all continents, on the other hand, tend to show a small reduction in overall warm bias. Over land the T2m biases are similar in nature if ERA5, instead of CFSR, is used as reference (not shown).

Over the high latitudes and humid tropics, the relative warming is pronounced in T2min (Fig. 11.6), and to a lesser degree in T2max (Fig. 11.5). Over arid regions, on the other hand, T2min (Fig. 11.6) are pronouncedly cooler, while the T2max (Fig. 11.5) may be either warmer (Sahara, Australia) or cooler (Western US, Atacama).

The annual mean precipitation (Fig. 11.7) in P7 is reduced over southeast Asia eastward of 90°E, the Maritime Continent, Eastern US, most of South America, Africa south of the equator, and the western basins of the Pacific and Atlantic Oceans. Precipitation is progressively enhanced with lead time over much of the Indian Ocean, eastern equatorial Pacific, and sub-Saharan Africa. The regions of reduced/increased OLR (Fig. 11.8) generally correspond to the regions of increased/reduced precipitation.

Seasonal analyses demonstrate that the warming of P7 over northern Eurasia and northern North America is present in winter, spring, and fall, but not summer (Fig. 11.4b). Focusing on the differences between P7 and P6 in the winter season (Fig. 11.9) demonstrates that over these regions the relative warming is accompanied by reduced snow depth (*snod*); increased ground (*gflux*) and sensible (*shtfl*) heat flux; decreased total cloud cover (*cloudtot*), primarily from a decrease in low clouds (cloudlow); and increased surface longwave radiation, both upward (*ulwrf*) and downward (*dlwrf*). Further evaluation revealed that in P7 (a) soil temperatures at initialization are warmer than P6, (b) snow depths are initialized with lower values and (c) the rate of compaction is much faster than in P6 and compared to reanalyses. It is hypothesized that the warming in winter high latitudes is partially attributable to the warmer soil temperature and thinner snow cover, leading to increased upward ground heat flux. An additional factor contributing to this warming is the invocation of the updated gravity wave drag parameterization, uGWDv1, which in atmosphere-only experiments produces wintertime warming over a similar region.

The warming of P7 over Greenland and Antarctica is present in their respective hemisphere's spring and summer but not in winter or fall (Fig. 11.4b), suggesting the relevance of shortwave radiation. Examination of the differences between P7 and P6 over Antarctica in DJF (Fig. 11.9) and Greenland in JJA (Fig. 11.10) and demonstrates an increase in middle level cloudiness (*cloudmid*) and moderate reduction in the surface downward shortwave radiation flux (*dswrf*) but a very large reduction of the surface upward shortwave radiation flux (*uswrf*), suggesting reduced albedo. The snow depths (*snod*) are smaller, possibly explaining the reduced albedo. Thus the hypothesized explanation for the summertime warming of Greenland is increased radiative heating due to reduced reflection and increased absorption of solar radiation resulting from reductions in surface albedo tied to reduced snow cover.

The warming of P7 in the humid tropics of Africa and South America is present in all seasons (Fig. 11.4b) and is accompanied (Figs. 11.9, 11.10) by reduced total cloud cover (*cloudtot*) primarily from reduced middle and high cloud cover (*cloudmid, cloudhi*); reduced precipitation (*prate*); increased sensible heat flux (*shtfl*); and reduced latent heat flux (lhtfl). It is hypothesized that the reduction of latent heat flux results from both drier soil initial conditions and changes in the parameterization introduced in Noah-MP and is the root cause of the warmer temperatures in these regions. Even though the same stomatal resistance formulation is used in both the Noah and Noah-MP models, Noah-MP produces significantly lower latent heat flux for the same soil moisture. The reason for this is being investigated.

Over deserts regions, the difference between P7 and P6 (Figs. 11.9, 11.10) shows strongly reduced surface upward shortwave radiation regardless of the sign of change in downward shortwave radiation, implying a decreased albedo and thus increased absorption of solar radiation leading to an increase in daytime temperatures. Due to the relatively low amount of cloud cover in these regions, the daytime warming is likely compensated by increased nighttime cooling resulting in little change in the daily average temperature.

11.3.3 MJO

The MJO RMM AC shows improvement in P7 compared to both P5 and P6. The combined RMM (RMM1+RMM2) AC in P7 is ≥ 0.6 through forecast day 18.5 compared to forecast day 16/17 in the preceding two prototypes (Fig. 11.11). The majority of improvement comes from a longer duration of forecast skill in RMM1 (AC ≥ 0.6 through forecast day 16 in P7 compared to forecast day 14 in P5 and P6). The RMM2 AC is slightly improved in P7 compared to that in P6 and is similar to P5.

Tests with an uncoupled version of the model indicate that some of the improvements of MJO forecast skill in P7 is likely a consequence of the physics changes, particularly those related to the parameterization of convection, including the stochastic CA, and PBL. The more accurate representation of diurnal sea-surface skin temperature variability resulting from introducing NSST in P7 is likely another contributing factor, given the recognized impact of such variability on MJO (see *e.g.* Slingo *et al*, 2003; Seo *et al.*, 2014; Shinoda *et al.* 2021). More tests to isolate the impact of individual contributions would be necessary in the future to fully diagnose how they benefit MJO prediction individually, and in tandem.

In terms of component variables, the P7 AC is slightly higher than P6 for OLR and U850. The RMSE in RMMs and components is smaller in P7 than in P6, but the RMSE of OLR is still larger in P7 compared to P5 (Fig. 11.12). As a result of the updated physics, there is a decrease of RMSE in U850, U200 and OLR in P7 compared to P6, although the RMSE of OLR in P7 is still larger than P5 (Fig. 11.13). Compared to P6, the MJO amplitude in P7 is weaker.

Once the forecasts are separated into those with initial strong and initial weak MJOs (Fig. 11.14), it is revealed that for forecasts with initially strong MJO, P7 has a negative amplitude bias beyond

Week 2. For forecasts with initially weak MJO, P7's amplitude bias is positive, and slightly smaller than in P6. There is some improvement (speeding up) in the propagation speed of initially strong MJOs but some degradation (slowing down) in the propagation speed of initially weak MJOs in P7.

11.3.4 Sea Ice

As a result of the overall increase in sea ice concentration in P7, particularly in the cold season (Figs. 11.9, 11.10, *icec*), the forecast Arctic ice area in P7 is larger than in P6 during most months, most pronounced for forecasts initialized on dates from October through March (Fig. 11.15). For these months, the increase is such that the P7 Arctic ice area forecast is closer to the NOAA/NSIDC CDR estimate than to NASA-Team. For forecasts initialized from May to August, it is closer to NASA-Team.

A limited set of runs made with intermediary prototypes (P7a and P7b) can be used to gauge the impact of Mushy thermodynamics on the forecast. These runs were made for 8 initial conditions (1st of January, April, July, and October of 2012 and 2013) with two versions of the model (P7a and P7b). The latter differed from the former by the inclusion of several features: (a) the update of gravity wave drag parameterization, (b) the update of surface compositing and albedo/emissivity calculations, (c) NSST, and (d) Mushy thermodynamics. It is assumed that of these changes, the largest changes to sea ice are attributable to (d), as any impact of (a)--(c) on sea ice is indirect.

The comparison between these runs (P7b minus P7a) for the first week of forecasts demonstrates the resulting increase in Arctic ice concentration (Fig. 11.16) and ice thickness (Fig. 11.17) in the January, April, and October, but not perceptibly in the July runs, and an increase in ocean salinity in the upper 10 meters in all four months (Fig. 11.18). The increase in ocean salinity in January, April, and October can be explained by the relatively larger ice formation, resulting in removal of a larger volume of water from the ocean. In the case of July, which is during the ice melt season, there is little change of ice concentration and thickness between the two runs, but the water gained from ice melt in the case of Mushy thermodynamics being invoked is saltier, still resulting in increased salinity of the upper ocean. In later forecast weeks (not shown) ice concentration and thickness remain larger for the simulations with Mushy thermodynamics, while the impact on upper ocean salinity is mixed.

11.3.5 Ocean Weeks 3&4

The SST skill in P7 Weeks 3&4 forecasts closely resembles that of P5 and P6 (Fig. 9.13). The skill in $\boldsymbol{\tau}_x$ forecasts remains unaltered from P6. Overall there seems to be no improvement in Niño 3.4 index and DMI in P7 (Fig. 9.14). The mean SST bias pattern remains very similar to that in P6. In general, the changes from P5 to P6 are reverted back in P6 to P7, for example, the warm bias along eastern boundary region increased from P5 to P6 but decreased from P6 to P7; the warm bias in the eastern equatorial IO decreased from P5 to P6 but increased from P6 to P7. However, this change is not robust in seasonal averages nor in respective regions of the global

ocean. For example, away from the equator, the warm bias along the US west coast, and in extratropical regions apparently increases from P6 to P7.

A slight decrease in warm SST bias from P6 to P7 at the eastern flank of the ITCZ is due to a decrease in net radiation bias from P6 to P7. τ_x bias increases in P7, with easterly winds widening meridionally off the Central American coastline. An enhanced net radiation bias westward towards the warm pool does not change the SST bias from P6 to P7. An increase in warm SST bias spreading south-eastward from the warm pool region is noticeable in P7, which is consistent with net radiation changes.

Along the western boundaries and tropical Indian Ocean a net radiative cooling from P5 to P6 reverses to a net radiative heating from P6 to P7. But regions of net radiative heating from P5 to P6 do not uniformly reverse to a net radiative cooling from P6 to P7. For example, a net radiative heating along the eastern boundary in the Pacific from P5 to P6, that further increases from P6 to P7, but does little to change the warm SST there. However, a decrease in net radiative heating along the eastern boundary in the Atlantic from P6 to P7 does decrease the warm SST bias in P7. A likely improvement in the stratus cloud deck cover in the south-eastern Pacific could reduce the cold SST bias here from P6 to P7 by warming the SST over a shallower MLD.

In general MLD shoals from P6 to P7 in regions where it deepened from P5 to P6. However, regions where MLD shoaled from P5 to P6, remained mostly unchanged from P6 to P7. D20 shows most of its changes very close to the equator in the east, mostly shoaling in the IO and Pacific, but deepening in the Atlantic from P6 to P7. Changes in D20 are not consistent with changes in mean SST, as it shoals in regions of SST cooling from P6 to P7.

The zonal wind stress, τ_x, shows most of its improvement in the high-latitude, where too weak τ_x in P6 (south of 40°S) are greatly improved in P7. The tropics do not show much changes in τ_x from P6 to P7, except for slight strengthening in the tropical north Atlantic and southern IO, and weakening of westerlies in the north IO.

The seasonal pattern of SST biases remains the same as in P6, with most of the changes focused in the equator, and along the eastern coastal regions of the southern hemisphere. As in P5, and P6, warm SST bias in P7 is also amplified in the summer hemispheres. A decreased (increased) cold (warm) bias in net radiation in the southern (northern) hemisphere increases the warming into the ocean in OND (MJJ) from P6 to P7.

10.3.5.1 Pacific Ocean

The cold SST bias in the equatorial cold tongue region reduces from P6 to P7 in OND, but increases slightly in MJJ (Figs. 9.15, 9.16, bottom row). The shoaling of D20 from P6 to P7 (Fig. 9.21) is consistent with the increase in cold bias, particularly in MJJ that an increased radiative heating fails to explain.

Further eastward the coastal equatorial warm bias increases in OND, likely due to reduced cooling from net radiation in P7. Weakened near-equatorial westerly winds in MJJ close to the coast, and a stronger westward South Equatorial Current push water off-shore likely increasing the coastal upwelling (Figs. 9.23, 9.24). This results in a shallower D20 driven cooling of the region in P7.The coastal warm SST bias along the Peru coast remains almost unchanged from P6 to P7 although the net radiative heating bias increases (Figs 9.15, 9.16, 9.17, 9.18).

The western Pacific warm pool warm SST bias increases from P6 to P7 most likely due to increase in net radiation heating bias in MJJ.

Along the US west coast, there is a slight decrease in cold SST bias in OND, but the warm bias in MJJ increases further from P6 to P7 and apparently merges southward with the warm bias along the ITCZ (Figs. 9.15, 9.16). Net radiative biases drive the warm SST bias along the US west coast, but is not consistent with the changes in warm bias along ITCZ, indicating the likely role of latent and sensible heat fluxes in the later.

The South Equatorial Current strengthens from P6 to P7 across the equatorial Pacific in OND (Fig. 9.24). The shear between North Equatorial Current (NEC, westward) and North Equatorial Countercurrent (NECC, eastward) remains almost unchanged from P6 to P7 in OND, but increases slightly in the far east in MJJ. The eastward Equatorial Countercurrent weakens further from P6 to P7 in MJJ, getting much closer to that seen in OSCAR. In MJJ, the eastward flowing EUC weakens slightly along with a strengthened South Equatorial Current from P6 to P7 in the eastern equatorial region (Fig. 9.22). This could enhance the vertical current shear in the east from P6 to P7 where EUC shoals along the D20, that shoals from P6 to P7, cooling the subsurface at that depth. However, in OND the EUC remains almost unchanged, but the westward South Equatorial Current that dominates above D20 strengthens from P6 to P7. The far eastern coastal subsurface above 20m warms from P6 to P7.

10.3.5.2 Atlantic Ocean

SST along the eastern boundary of the Atlantic Ocean is cooler in P7 than P6 consistent with decrease in net radiative heating in a deeper MLD in P7. However the changes in D20 opposes the surface cooling, i.e. it deepens in regions of SST cooling in P7. However, cooler SST in P7 in the SE Atlantic basin is consistent with a slight shoaling in D20 in MJJ. $\boldsymbol{\tau}_x$ in general, is stronger in P7 than P6 enhancing the mean bias, particularly along the equator.

The equatorial surface currents (NEC) show large changes in P7, weakening significantly from P6, particularly in OND, and to a lesser extent in MJJ. However eastward NECC apparently strengthens just north of NEC in P7.

10.3.5.3 Indian Ocean

Much of the improvement in SST bias in P6 is retained in P7 in the tropical Indian Ocean, except for a slight increase in warm bias in the eastern equatorial regions, which likely is due to an

increase in net radiative heating bias in P7. Both MLD and D20 shoals in P7 (particularly in MJJ) indicating upper ocean processes play an additional role in SST changes.

The equatorial westerlies weaken in P7, particularly in MJJ (season of weak westerlies), and comparatively less in OND (season of strong westerlies). This drives a shallower D20 in P7 compared to P6. The mean τ_x bias increases in P7 overall in the IO, with strengthened easterlies in the south and weakened westerlies in the north, particularly in MJJ (monsoon season).

Eastward surface currents in OND weakens slightly in P7, but reverses its direction in MJJ. This reveals biases in surface currents seasonally as eastward current only weakens in MJJ in OSCAR.

D20 shoaling also cools the equatorial subsurface in P7. The weakening of equatorial westerlies in MJJ in P7 likely drives the shoaling of equatorial D20. Although subsurface warming in P6 is reversed - as in subsurface cooling P7 in MJJ, however such reversal is not obvious in OND. The subsurface currents above D20 also reverses particularly in MJJ, and only weakens in OND.

11.3.6 Land

The Surface Radiation Budget Network (SURFRAD, Augustine *et al.*, 2000) has seven stations across the continental United States (CONUS) and provides accurate, continuous, long-term measurements pertaining to the surface radiation budget. Here three stations, Desert Rock, Nevada (with a bare soil surface type); Sioux Falls, South Dakota (croplands); and Goodwin Creek, Mississippi (cropland/natural vegetation mosaic), are selected to represent the western, northern and southeast CONUS, respectively. Comparisons are shown between UFS (P6 and P7) and SURFRAD measurements for UFS forecasts initialized on 00 UTC, 1 Jan 2012–2018 and 1 Jul 2011–2017. Note that the UFS output is only available at 6-hourly intervals (00:00, 06:00, 12:00 and 18:00 UTC), while the SURFRAD observations are available at hourly intervals.

The 35-day-average diurnal cycle of land surface skin temperature (Tskin) at Desert Rock from the January runs (Fig. 11.19) shows that P6 has overall small (generally cool) bias in both nighttime and daytime Tskin, while P7 has a more pronounced cold bias during both daytime and nighttime. In July at Desert Rock (Fig. 11.20), P6 still has small biases during daytime and nighttime. P7 has small biases during daytime but large cold biases appear during nighttime reaching up to −10°C.

At Sioux Falls in January (Fig. 11.21a), P6 shows relatively small and generally cool Tskin biases at all times of the day. In most years, P7 overestimates the January Tskin at all times of the day, with the exception of 2014 and 2015 when it displays cold bias. As discussed previously, snow in P7 compacts faster than in P6, leading to overall lower depths and more frequent instances of snow-free ground. The temperature biases are strongly related to snow presence—P7 is systematically warmer than P6 when snow cover is present in P6 but absent in P7. The lack of such warming 2014 and 2015 is consistent with these two years having the fewest (none, in the case of 2014) snow-free P7 forecast times (Fig 11.21b).

In July at Sioux Falls (Fig. 11.22), P6 has small (generally warm) nighttime biases and generally cool daytime biases, except for 2012 and 2016 which have warmer daytime temperatures than the observations. P7 has small biases during nighttime but relatively cool biases of −3 to −8°C during daytime. The July Tskin at Sioux Falls in P7 exhibits markedly smaller interannual variability and reduced diurnal range than both P6 and observations.

At Goodwin Creek in January (Fig. 11.23) the nighttime Tskin in both P6 and P7 generally agrees well with observations, but the daytime Tskin is underestimated in most years, translating into an underestimation of the diurnal range. The interannual variability of the January Tskin is also generally underestimated by both P6 and P7. In July at Goodwin Creek (Fig. 11.24) the daytime and nighttime Tskin is overestimated in both P6 and P7, and the overestimation is more pronounced in P7 from late evening and into early morning.

As mentioned previously, both P6 and P7 present large warm T2m biases over Antarctica and Greenland where land is covered by glacial surface, and the warm bias of P7 is larger than that of P6 (Fig. 11.4a). The comparative warming for these two regions is most pronounced in summer, *i.e.*, DJF for Antarctica and JJA for Greenland (Fig. 11.4b). In-situ measurements at the South Pole Observatory, obtained from the public ftp file archive of the ESRL Global Monitoring Laboratory (https://gml.noaa.gov/dv/data/?site=spo) are selected for verification of surface fields over Antarctica. Fig. 11.25 shows the averaged diurnal cycles of Tskin at South Pole for 4 weeks initiated at 00 UTC, 1 Jan from 2012 to 2017. The warm biases appear in all years and can reach up to +10°C in P7, substantially larger than P6. The time series of Tskin further indicates that P7 starts warming just after the first few days (not shown). Fig. 11.26 gives the time series of albedo (ratio of upward SW flux to downward SW flux) for 2012 and 2013. P6 using Noah has a constant albedo (0.82) over snow. The albedo in P7 using Noah-MP is derived according to the snow parameters and varies from 0.7 to 0.8, much smaller than the observation (from 0.8 to 0.9). The lower albedo in P7 results in a lower upward short-wave and retains more heat energy at the surface, which would contribute to the warm bias.

In the winter seasons at the South Pole there is no solar radiation. Both P6 and P7 show warm biases (Fig. 11.27), but the warm biases in P7 are smaller than those in P6. Analysis of the downward long-wave radiation at the surface also indicates the positive biases in both prototypes (Fig. 11.28). The positive surface downward long-wave radiation bias may be closely related to unrealistically large amounts of clouds in the lower troposphere (not shown).

A basin-scale water budget analysis was conducted over ten large river basins that have diverse climate regimes. The river basins sorted by observed precipitation magnitude are shown in Fig. 11.29. P7 produces lower precipitation in all but the Yukon River Basin (Fig. 11.29, left). Using the CPC gridded precipitation analysis as a reference, this represents an improvement in eight of the ten basins (the Yukon and Murray basins are the exceptions). The partitioning of precipitation into evapotranspiration (ET; Fig. 11.29, right) shows that P7 has lower ET in all basins, and substantially lower in tropical basins (Amazon and Congo). This contributes to the high temperature biases referenced above and seen in Figs. 11.4a and 11.4b. However, the reduced

ET is an improvement compared to P6 in all temperate climate basins. For the lower-precipitation Arctic basins (Lena and Yukon), this reduction is a slight degradation.

11.3.7 Aerosols (P7.2)

Prognostic aerosols were introduced in the UFS in a version otherwise similar to P7 (P7.2) and tested in a set of eight model runs (initial dates on the 1st of January, April, July, and October of 2012 and 2013). Aside from the addition of prognostic aerosols, P7.2 differed from P7 in the following: (a) a bug fix was made in sfcsub.f ensuring that a sea ice loop only operates on non-land points, (b) the orographic wave drag parameter *cdmbgwd* in the gfs_physics_nml namelist was changed from “1.1,0.72,1.0,1.0” to "1.0,2.2,1.0,1.0", and (c) CA parameterization was turned off. The original P7 initial conditions were modified to update: (a) the atmosphere forcing source of the land spin up (NASA GLDAS instead of GDAS); (b) snow initial conditions (derived from ERA5 instead of GEFSv12 Reanalysis); (c) ice thickness initialization over fractional lakes; and (d) wave model initialization. These changes do result in somewhat different from P7 model forecasts.

P7.2 Prognostic aerosols component

A novel, atmosphere-driven prognostic aerosol component (UFS-Aerosols) was developed for the UFS through a NOAA-NASA collaborative effort as a deliverable for the UFS-R2O Project (Kinter *et al.*, 2020: Sect. 1.2, Key Goal 1). UFS-Aerosols embeds NASA’s 2nd-generation GOddard Chemistry Aerosol Radiation & Transport (GOCART) model (https://github.com/GEOS-ESM/GOCART) in a NUOPC-compliant model component that ties in the UFS coupling infrastructure and provides interoperability across U.S. modeling centers. The component predicts tropospheric mixing ratios of major aerosol species such as:

- Dust and sea salt (both as 5-bin particle size distributions);
- Hydrophilic and hydrophobic black (BC) and organic (OC) carbon;
- Sulfur dioxide, sulfate, dimethylsufide (DMS), and methanesulfonic acid (MSA) to model the atmospheric sulfur cycle (see e.g. Chin *et al*. 2000a,b);
- Nitrates (3 size bins) along with ammonia and ammonium to better represent radiative forcing.

Particulate matter (PM10 and PM2.5) is further computed as a model diagnostics

UFS-Aerosols features an updated FENGSHA (Dong *et al.*, 2016) dust scheme providing a more accurate description of dust fluxes, and is able to ingest time-varying anthropogenic, biogenic, volcanic, and biomass burning emissions from high-resolution inventories. In the UFS prototype configuration, the aerosol component is coupled directly to the atmospheric model and exchanges aerosol mixing ratios with it at high frequency on the same grid. While bidirectional data exchanges occur, this coupling configuration is considered one-way from a scientific standpoint since prognostic aerosols have no impact on meteorology.

The integration of UFS-Aerosols into UFS incorporated the MAPL (https://github.com/GEOS-ESM/MAPL) infrastructure layer, which provides the framework to run GOCART and perform disk I/O (*e.g.* read emissions, write diagnostics). MAPL is the foundation of NASA GEOS-5 and gives the UFS community unprecedented access to NASA's expertise, datasets, and additional GEOS-5 modeling components and tools. Collaborative development of MAPL for the UFS led to multiple NASA/NOAA MAPL releases, starting with version 2.4 on Nov. 20, 2020 (see *e.g.* v2.7.3: https://doi.org/10.5281/zenodo.5027614).

P7.2 Setup

Interpolated MERRA-2 aerosol mixing ratio values were added to the atmosphere's initial condition files to initialize global aerosol concentrations. The uptake of dust and sea salt is dynamically predicted during model integration, while anthropogenic, biogenic, wildfire, and volcanic emissions are continuously prescribed. The following surface emissions inventories were used throughout the model runs:

- Anthropogenic emissions:
  - Community Emissions Data System (CEDS), release 2019 (Hoesly *et al.*, 2019)
  - Hemispheric Transport of Air Pollution (HTAP) v2.2 (Janssens-Maenhout *et al*., 2015) for aviation emissions during the three phases of flight
- Biogenic emissions:
  - Model of Emissions of Gases and Aerosols from Nature (MEGAN) offline datasets (Guenther *et al.*, 2012)
- Wildfire emissions:
  - NASA's Quick Fire Emissions Dataset (QFED) (Darmenov and da Silva, 2015)
- Volcanic emissions:
  - Carn's Ozone Monitoring Instrument (OMI)-based SO2 emission dataset (Carn, 2019)

P7.2 Results

Upon initial examination of the results, the predicted total AOD values from the July 2012 experiment were found to be unexpectedly low. This was traced back to incorrect initial conditions for that date. Consequently, aerosol predictions from the July 2012 experiment are discarded and will not be discussed further.

The Week 1, Week 2, and Week 3&4 total AOD predictions of the remaining seven experiments are presented below. In interpreting the results, it should be noted that:

- The FENGSHA dust scheme introduces additional detail (and possibly accuracy) in the description of dust emission processes, compared to the Ginoux scheme (Ginoux *et al*., 2001) used in GEOS-5, which could partially explain some of the discrepancies for dust (and total) AOD between P7.2 and MERRA-2.
- The NASA QFED product was used to provide emissions from biomass burning in P7.2. This dataset tends to underestimate the OC sources when compared to the Blended

Global Biomass Burning Emissions Product, version 3 (GBBEPx V3, https://www.ospo.noaa.gov/Products/land/gbbepx/) emissions (Zhang *et al.*, 2012) used in the operational medium-range global aerosol forecast model at NOAA (GEFS-Aerosol).

Means in Week 1

In the January Week 1 forecasts (Fig. 11.30), the largest overpredictions of total AOD (*i.e.*, P7.2 AOD exceeding MERRA-2 values) occur predominantly over the Western Coast of Africa, where dust transport from the Sahara Desert (Fig. 11.31) and biomass burning in Africa's central region (Fig. 11.32) are major contributors to the total aerosol concentrations. The P7.2 overprediction of AOD compared to MERRA-2 is mostly localized near emission sources and amounts to about 0.2. Additional AOD overprediction (between 0.1 and 0.2) is observed over East Asia, and is contributed primarily from OC (Fig. 11.32), and sulfate (Fig. 11.33). Underpredictions of total AOD in high NH latitudes in January and April Week 1 forecasts are also contributed by sulfates.

In the April Week 1 forecasts (Fig. 11.30), the largest positive differences are associated with overprediction of OC from forest fires in Southeast Asia (Fig. 11.32). Dust contributions from the Sahara Desert (Fig. 11.31) account for the AOD overpredictions near the emission source, while the dust plume transported across the Atlantic Ocean is largely underpredicted. Underestimation of dust contributions in the Middle East (Fig. 11.31) accounts for the underprediction of the total AOD in that region for both years.

Large AOD overprediction (above 0.2) appears along the Western Coast of Africa, North America, and Central Asia for the July 2013 experiment (Fig. 11.30). Dust AOD (Fig. 11.31) is the largest contributor to the high total AOD bias over the Gobi and Taklamakan Deserts in Central Asia as well as in Northern Africa. Over North America, AOD overestimation is contributed by OC and AC (Fig. 11.32) AOD, thus it appears to be linked to emissions from wildfires, which are prevalent in this season.

In the October simulations, the Week 1 differences between P7.2 and MERRA-2 values are smaller in magnitude than in the other three seasons (Fig. 11.30). Dust overprediction over Africa is lower than in other seasons (Fig. 11.31), however some total AOD overprediction (~0.1–0.2) from biomass burning regions in South America (Fig. 11.32) and anthropogenic pollution over China (Fig. 11.33) still occurs.

Compared to the other species, the Week 1 difference between P7.2 and MERRA-2 AOD is smallest for sea salt (Fig. 11.34). Sea salt AOD differences (both over and underprediction) for Week 1 are generally in the range of 0.1–0.15 for all seven runs. The largest magnitude sea salt AOD differences are negative, *i.e.*, strong underprediction is more prevalent than strong overprediction, and occur in regions generally associated with strong surface winds.

Means in Week 2

Across all experiments, over the regions where the Week 1 predicted total AOD exceeds the MERRA-2 values, the predicted Week 2 values increase further (Fig. 11.35).

In the January 2012 and 2013 simulations, a total AOD overprediction in the range of ~0.4–0.5 is seen along the coast of Africa, Central Africa and Europe. A similar increase is observed over the Eastern Asia forest fire region, with underprediction over India.

In the April runs, a total AOD overprediction by 0.4 and above is present over the Sahara Desert and Southeast Asia. A few of the biomass burning plumes (contributing to high OC AOD) from Asia reach high latitudes (north of 60°N) and accumulate over time, likely due to insufficient atmospheric removal.

AOD overprediction from North American biomass burning plumes dominates the Week 2 results for July 2013, followed by dust contributions over Northern Africa and Central Asia, and sulfate contributions over northern China. Increased AOD bias over the Southern Ocean is also evident as a result of elevated sea salt concentrations.

AOD overprediction in Week 2 over the South American biomass burning zone is largest in the October simulations for both years. Dust AOD over the Saharan desert and sulfate AOD over China show growth in overprediction when compared to Week 1.

Means in Weeks 3&4

A consistent feature of the Weeks 3&4 forecast across all experiments (Fig. 11.36) is the overprediction of the total AOD across the Atlantic Ocean, predominantly contributed by overly high dust AOD (Fig. 11.37). This overeprediction is generally larger than in Week 1 and Week 2, and is least pronounced in the January 2013 experiment.

In the April simulations (Fig. 11.36), total AOD overestimations larger than 0.6 are seen in the dust plume over the Sahara Desert and the Atlantic Ocean (Fig. 11.37). Additionally, AOD overprediction from dust and biomass burning (OC) is noted across Northern and Southern African regions, respectively (Figs. 11.37 and 11.38). Dust originating from the Taklamakan Desert continues to contribute to total AOD overprediction (~0.5 and higher), while dust originating from the Middle East continues to be largely underpredicted (Fig. 11.37). The increase in AOD overestimation over East Asia's biomass burning regions (Fig. 11.38) amounts to 0.4 or larger AOD differences at high Northern latitudes.

In July 2013, the largest overprediction of total AOD (Fig. 11.36) is contributed by Saharan dust over Northern Africa and by dust from the Taklamakan desert in East Asia (Fig. 11.377). The total AOD overprediction in these regions is greater than 0.5. The model underpredicts total AOD over the Middle East, largely due to underprediction of dust AOD in the region (Fig. 11.37).

Dust overprediction over Africa (Fig. 11.37) is the largest contributor to the total AOD differences (Fig. 11.36) in the October simulations, followed by underprediction of anthropogenic pollution (sulfates) over Eastern Asia (Fig. 11.39).

Compared to other species, differences in sea salt AOD (Fig. 11.40) remain smal for all months and retain similar magnitude between Week 1 and Weeks 3&4. As in earlier weeks,

underpredictions are more pronounced than overpredictions, and are found near areas that typically experience strong surface winds.

Figure 11.41 shows the Weeks 3&4 total AOD differences when compared with MODIS. The pattern is similar to that computed using MERRA-2 with one notable difference: as MODIS data is available mostly between 60°S and 60°N, high latitude regions are missing. Both data sets show overprediction of the total AOD over the Sahara Desert and the dust outflow region over the Atlantic Ocean for all seasons, along with major biomass burning regions (Southern Africa, South America, North America, and Asia).

RMSE

Figures 11.42, 11.43, and 11.44 show the RMSE of total AOD against MERRA-2 for the 2012 and 2013 experiments for Week 1, Week 2, and Weeks 3&4. For Week 1 (Fig. 11.42), the RMSE is in the range of 0.2–0.3 over most regions of the globe, except over the Eastern Asia wildfire region (in April) and over the Sahara desert (in July) where it exceeds 0.5. For Week 2, RMSE exceeding 0.8 is found over the Taklamakan Desert, the North American biomass burning region and its outflow region (both in July), the Saharan dust region, and the Eastern Asia fire region (in April). Figure 11.44 shows that Weeks 3&4 RMSE grows over the East Asia wildfire outflow region and the Southern Africa forest fire source region.

Figure 11.45 shows global RMSE of total and all species AOD compared to MERRA-2 for day 1 to day 28 of all seven runs. The lowest global RMSE (and, as discussed above, smallest amplitude of local differences) for total AOD is seen in the October runs, while the highest global RMSE is found in April and July simulations. In the January, April, and July simulations, dust is the largest contributor to the total AOD RMSE, followed by either OC or sulfates.

Dust AOD maintains relatively low global RMSE (below 0.1) up to day 10 in both January simulations, gradually increasing to 0.15 and above at longer forecast leads. Some increase in sulfate AOD RMSE is seen starting with the second week of integration in the January and April simulations. In April, dust (from Sahara), OC (from East Asian fires), and sulfates are the main contributors to total AOD global RMSE. The total AOD global RMSE in the July 2013 run is largely contributed by discrepancies associated with dust from the Saharan and Taklamakan desert and OC from wildfires in Southern Africa and the Amazon. The RMSE of all species remains under 0.1 for both October runs.

12. Summary

12.1 Prototypes

The main characteristics of the seven coupled UFS prototypes are summarized in Table 2.1 (model configuration), Table 2.2 (initial conditions), and Table 2.3 (lake ice specification). Additional details are provided in the corresponding section for each prototype (P1–P7, Sections 5–11).

12.2 Benchmark AC Scores

The prototypes' Benchmark AC skill scores for Week 1, Week 2, and combined Weeks 3&4 forecasts are summarized and compared to CFSv2 in Table 12.1 (RAW scores, *i.e.*, with forecast anomalies calculated by subtracting the observed climatology) and Table 12.2 (SEC scores, *i.e.*, with forecast anomalies calculated by subtracting each prototype's own climatology).

All benchmark RAW AC scores (Table 12.1) for Week 1 are an improvement over CFSv2 in all prototypes, with the exception of: (a) Niño 3.4 SST in P1, the only prototype not using CPC-3DVar ocean initial conditions, and (b) T2m CONUS land in P7 (whose score matches the CFSv2 score). Results are similar in nature for the SEC AC scores (Table 12.2), showing improvement over CFSv2 in all prototypes except for (a) SST in P1 and (b) T2m CONUS land in P6 (matching CFSv2) and P7 (below CFSv2). Note that the SEC AC scores are higher than the RAW AC scores for all leads and variables, reflecting the beneficial impact of bias correction. The usefulness of bias correction is particularly clear for SST, and increasingly so at longer lead times.

For Week 2, the Niño 3.4 SST AC scores (whether RAW or SEC) in all prototypes are an improvement over CFSv2. This is also true of the Global Tropics SST with the exception of the RAW AC score in P6, which is below the CFSv2; and of the Niño 3.4 precipitation, with the exception of P1 (RAW score matching CFSv2, SEC score below CFSv2). The Week 2 scores (both RAW and SEC) for 500mb heights in both hemispheres are above CFSv2 in all prototypes without exception. Week two CONUS land T2m scores, whether RAW or SEC, are above CFSv2 in three prototypes (P1, P2, P6) and below CFSv2 in four prototypes (P3.1, P4, P5, P7). For CONUS land precipitation, RAW and SEC AC scores are above CFSv2 in five prototypes (P1, P2, P3, P5, P6) and either below or matching CFSv2 in two (P4, P7).

The Weeks 3&4 RAW AC scores for Niño 3.4 SST are above those of CFSv2 in P1–P3.1, and below CFSv2 in P4–P7; the SEC AC scores are above CFSv2 in all seven prototypes. For Weeks 3&4 Global Tropics SST, both the RAW and SEC AC scores are an improvement over CFSv2 in all prototypes, with the only exception being the RAW score in P6. The AC scores (RAW and SEC) over the northern hemisphere 500mb heights are better than CFSv2 in P1–P6, but worse in P7. Over the southern hemisphere, both RAW and SEC scores for 500mb heights in all seven prototypes are an improvement over CFSv2. The Weeks 3&4 CONUS T2m AC scores (RAW and SEC) are better than CFSv2 in P1, P2, P3.1, P4, and P6, and worse in P5 and P7. The Weeks 3&4 scores for CONUS precipitation are exceedingly low across the board, indicating very little useful skill at that lead time. Be that as it may, the P2–P7 scores are above those of CFSv2, most clearly in P6; the scores for P1 and P7 are lower than the already very low scores of CFSv2. Using output from P2, Krishamurty *et al.* 2022 demonstrated that despite the very low CONUS precipitation skill beyond week two, there is an intraseasonal oscillation mode of predictability stemming from large-scale circulation 180°–150°W/20°S–41°N which is captured by the prototypes and can provide some forecast skill in summer at longer time scales.

12.3 Annual Mean Biases

The spatial pattern of annual mean SST bias is very similar in all prototypes (Fig. 12.1) and consists of predominantly warm bias in the tropics (Indian Ocean, western equatorial Pacific, extending southeastward into the southern Pacific and eastward along the equator across the central and into the eastern equatorial Pacific, equatorial Atlantic, and along the western coasts of tropical South America and Africa) and cold bias in the cold tongue region. The warm bias is increasingly prominent in more recent prototypes. The magnitude of both the warm and cold biases increases with lead time from Week 1 into Weeks 3&4. The warm annual mean bias is predominantly contributed by the corresponding hemisphere's summer seasons (Figs. 9.15 and 9.16).

In CFSv2, the annual mean Weeks 3&4 T2m bias (Fig. 12.2, top left) is relatively mild, cool over the high latitudes of Asia and along the eastern coasts of North America, South America and Australia, and warm elsewhere, most notably over desert regions. In P1 (Fig. 12.2, top right), the warm biases are warmer and more spread out, covering most land from 30°S to 30°N. The high latitudes of the northern hemisphere land in P1 are marked by cold bias. Starting with P2 (Fig. 12.2, second row, left), the annual mean Weeks 3&4 T2m bias is warm over virtually all land areas. T2m over the high latitude lakes in P1–P3.1 has a pronounced cold bias, as a result of lake temperatures being set to the nearest ocean point; starting with P4, that is no longer the case. Elsewhere, the T2m biases differ slightly between prototypes in P2–P5, but the general pattern, magnitude, and sign remain consistent. The land T2m biases are somewhat reduced in P6, but then noticeably amplified in P7 (Fig. 12.2, bottom right). The largest warm bias increases in P7 are encountered over (a) the humid tropics of South America and Africa (all seasons), (b) the high latitudes of the northern hemisphere (annual warming stemming from winter, spring and fall), , and (c) ice-covered land (annual warming stemming from the corresponding hemisphere's spring and summer).

The annual mean Weeks 3&4 land T2max bias in CFSv2 (Fig. 12.3, top left) is overwhelmingly cool, with some exceptions in the tropics. T2max in P1 (Fig. 12.3, top right) is, on average, warmer than in CFSv2, although still carrying a cool bias over most land areas outside of the tropics; the P1 bias is warm over most tropical land. With P2 (Fig. 12.3, second row, left), T2max became warmer yet (especially at higher latitudes), increasing the warm biases and reducing the cool biases found in P1. Between P2 and P6, there were some fluctuations in the bias but the overall pattern and magnitudes remained similar. In P7 (Fig. 12.3, bottom right), both the extent and magnitude of warm biases increase.

The annual mean Weeks 3&4 land T2min bias in CFSv2 (Fig. 12.4, top left) is warm over CONUS, most of northern Africa, southern South America, the Indian subcontinent, and central and northeastern Asia, and cool over Greenland, South America north of 30°S, central Africa, southeastern and northwestern Asia, and the Maritime Continent land. The pattern of bias in P1 (Fig. 12.4, top right) is similar to that of CFSv2. In P2 (Fig. 12.4, second row, left), T2min is, on average, warmer than CFSv2. The T2min bias remains similar in pattern and magnitude in P2–P5, shifting slightly cooler at high latitudes in P6 (Fig. 12.4, bottom left). Going from P6 to P7 (Fig.

12.4, bottom right), T2min warms up at high latitudes and most of the humid tropics (most of South America except for its west coast, Africa between 20°S and 5°N, and Maritime Continent land), and cools down over most of the dry tropics (western CONUS, Sahara, Australia, Arabian peninsula and Asia between 20°N and 40°N, and the western coast of South America). P7 is the only prototype with warm T2min bias over South America north of 30°S.

The annual mean Weeks 3&4 precipitation bias in CFSv2 (Fig. 12.5, top left) is wet over most tropical oceans, with a gap over the equatorial Pacific and eastern South Pacific cold tongue region. The CFSv2 bias is also wet over some land areas—along the western coast of South America, most of CONUS, the Tibetan Plateau, and along the Sahel. CFSv2 exhibits a pronounced dry bias over southeast Asia, the Maritime Continent and South America north of 30°S. In P1 (Fig. 12.5, top right), the magnitude of dry biases over southeast Asia, the Maritime Continent and South America is reduced. There is a more pronounced dry bias into the equatorial Indian Ocean and western equatorial Pacific, and a spread of wet bias into the central and eastern equatorial Pacific, into the eastern South Pacific cold tongue region, and into the eastern tropical South Atlantic. The wet bias over CONUS is reduced, and the wet bias along the Sahel is no longer present. The general pattern of precipitation bias is similar in P2–P5. In P6 (Fig. 12.5, bottom left), wet bias becomes more prominent over the Maritime Continent and central equatorial Pacific. In P7 (Fig. 12.5, bottom right) the central equatorial Pacific wet bias is further increased, and, similarly to CFSv2, there is wet bias over the Sahel, and strong dry bias over South America north of 30°S.

In CFSv2, the annual mean Weeks 3&4 OLR bias (Fig. 12.6, top left) is markedly positive (suggesting underestimation of convection) over the eastern Indian Ocean/Maritime Continent/western equatorial Pacific, the cold tongue region, South America north of 30°S, and equatorial Africa. The bias is strongly negative over the western equatorial Atlantic, Sahara, the Arabian peninsula, parts of south Asia, and the southern Indian Ocean and western Australia. In P1 (Fig. 12.6, top right), positive OLR bias covers the entire globe, but most of the maxima are located in similar regions as in CFSv2. The large positive bias spreads further eastward into the Pacific and has a secondary maximum (absent in CFSv2) around 120°W/10°N. Unlike CFSv2, the cold tongue region does not have a local maximum in OLR bias. Also unlike CFSv2, P1 has pronounced maxima in OLR poleward of 60°S and 60°N. In P2 (Fig. 12.6, second row, left), the overall positive biases are substantially reduced, but the structure of local maxima remains similar to P1. In P3.1–P5 the high latitude OLR biases are eliminated but the tropical pattern of biases remains in place. The Weeks 3&4 biases are smallest in P6 (Fig. 12.6, bottom left). In P7 (Fig. 12.6, bottom right), the positive biases are amplified in the eastern Indian Ocean/Maritime Continent/western equatorial Pacific; equatorial Africa; and, particularly strongly, over South America north of 30°S. The positive bias over the western equatorial Pacific, centered around 120°W and 10°N, is, on the other hand, reduced.

12.4 MJO Representation

Table 12.3 shows the tabulated RMM1, RMM2, and combined RMM skill, in terms of approximate forecast days until the corresponding AC decreases to 0.6, for the operational coupled model

(CFSv2) and all seven coupled UFS prototypes. The skill of all prototypes is above that of CFSv2 for both the individual (RMM1 and RMM2) and the combined RMM. The gain in forecast skill lead time is on the order of 3–5 days for RMM1 and RMM2, and 4–7 days for the combined RMM.

The duration of skillful forecast in RMM1 increases from P2 to P3.1 by about 2 days, decreases from P3.1 to P4 by about 1.5 days and remains at that level until P7, when it regains the skill lost in the previous three prototypes. The forecasts of RMM2 (representing primarily the Western Pacific/Indian Ocean convection) remain skillful for a few days beyond RMM1 (representing primarily the Maritime Continent/Western Hemisphere/Africa convection); the differences in skillful forecast length between RMM1 and RMM2 is smallest in the most recent prototype, P7. The duration of skill for the combined RMM forecast was shortest in P5 (16 days) and longest in P2 (19 days), closely followed by P7 (18.5 days).

When considering a set of forecasts from the S2S database (Vitart *et al.*, 2017) over the period 1999–2010, Vitart (2017) found that the lead time at which the models' control run bivariate RMM AC reached 0.6 had a large range across models. The lead was approximately 6 days in one (HMCR), 10–14 days in six (ECCC, CMA, BOM, JMA, CNRM, ISAC), 15 days in NCEP-CFSv2, 17 days in the UKMO model, and 21 days in the ECMWF model. More recently, examining a collection of SubX (Pegion *et al.,* 2019) and a few of the S2S ensemble boreal winter reforecasts, Kim *et al*. (2019) found the lead times when RMM AC reaches 0.6 to be approximately 17–18 days in three of the model ensembles (RSMAS-CCSM4, Navy-ESPC, and ESRL-FIM), 20–22 days in five (KMA-GloSea5, NCEP-GEFSv12, NASA-GEOS5, and NCAR-CESM1), and approximately 27 days in ECMWF-CY43R3.

The coupled UFS prototypes' AC skill (16–19 days to RMM AC of 0.6) compares very favorably to the deterministic S2S models' skill discussed in Vitart (2017), losing out only to the ECMWF forecast. Taking into account that the MJO forecast skill is generally higher in boreal winter, and that the skill of ensemble forecasts exceeds that of deterministic forecasts, the coupled UFS skill prototypes' AC skill (deterministic, all year) comparison with the more contemporary SubX models in the Kim *et al.* 2019 (ensemble, winter) study is also encouraging.

The MJO amplitude and phase errors are examined in P3.1–P7. All exhibit a negative phase bias (*i.e.*, an underestimation of the MJO propagation speed), growing with lead time. The MJO amplitude bias in P3.1–P6 is positive (*i.e.*, the MJO strength is overestimated) and growing with lead time. In P7 the MJO amplitude bias is close to zero at most lead times, briefly going negative during Week 4. Forecasts initialized with strong MJO tend to have smaller bias than those initialized with weak MJO, while the phase bias is independent of the initial MJO strength. In the case of P7, the forecasts with initially strong MJO underestimate its amplitude, while those with initially weak MJO overestimate it.

12.5 Sea Ice

Arctic ice area from the prototypes is compared to that of CFSv2 and to that derived from two satellite-based estimates of sea ice concentration, NOAA/NSIDC-CDR (version 3) and NASA-

Team. The Arctic ice area from NOAA/NSIDC-CDR is systematically larger than that from NASA-Team in all months and years examined here. The CFSv2 forecasts are closer to the larger estimate, exceeding it during seasons of low ice. In forecasts initialized in the ice melt season, the CFSv2 area decreases too slowly, while for those initialized in the ice growth season it increases too rapidly.

In P1 and P2, the forecast Arctic ice area is reduced relative to CFSv2, and, for initialization dates between January and June—closer to NASA-Team than to NOAA/NSIDC-CDR. The switching of ice initialization from CFSv2/SIS1 to CPC sea ice analysis (CPC-CSIS) starting with P3.1 further reduces the forecast Arctic ice area in June–September, bringing it even closer to the NASA-Team estimate. From P3.1 to P6 changes in this metric are minor.

In P7, the activation of the mushy thermodynamics option in the ice model brings about an overall increase in ice concentration, particularly during the cold season. As a result, the Arctic ice area in P7 increases, most prominently—between September and April. At these times, it is close to NOAA/NSIDC-CDR. For May to August, on the other hand, the P7 Arctic ice area is closer to the values from NASA-Team.

12.6 Ocean Weeks 3&4

Ocean Weeks 3&4 forecasts were evaluated and intercompared for the three latest prototypes, P5–P7.

SST skill for Weeks 3&4 forecasts is the highest in ENSO active region of tropical Pacific whereas $\boldsymbol{\tau}_x$ skill is highest just west of it (Fig. 9.13). Although the SST skill is retained from P5 to P7, $\boldsymbol{\tau}_x$ skill drops in specific regions of the global ocean. Overall, warm SST bias mostly dominates the tropics and is generally associated with a positive net radiation bias in regions with shallow MLD, and is amplified in the summer hemispheres. In many regions of the ocean the changes from P5 to P6 in SST, MLD, and net radiation revert back in P7, but there are significant instances where this is not consistently true (Figs. 9.15, 9.16). For example, the warm bias along the US west coast increases from P5 to that in P7.

The eastern equatorial Pacific cold tongue SST bias decreases from P5 to P7, primarily due to reduced net radiation bias in P7 (Figs. 9.16, 9.18). Seasonally, a slight increase in the cold bias in MJJ in P7 is accompanied by a stronger westward off-shore South Equatorial Current, which shoals the equatorial D20. Further eastward along the coast, the warm SST bias increases from P5 to P7, likely originating from the ocean initial conditions with additional contribution from radiative biases. The seasonal strengthening of shear generated between the equatorial Pacific currents is much improved in P7 and closer to that in OSCAR. The shoaling of the Equatorial Undercurrent and D20, along with strengthening of South Equatorial Current in P7, increases the vertical shear-driven mixing, driving a cooler upper ocean in P7 (Figs. 9.26, 9.24, 9.25).

In ITCZ the net increase/decrease in SST bias between the prototypes is not driven by changes in net radiation biases only. Instead, latent and sensible heat flux biases likely play an additional

role in the SST variations. Similarly, in the southeastern Pacific, wind stress curl driven near shore coastal upwelling likely play an additional role in changes in the SST bias between prototypes.

In the equatorial Atlantic, an increase in warm SST bias from P5 to P6 is primarily associated with a deeper D20, indicating an increased upper ocean heat uptake driving the warmer SST bias. From P6 to P7, on the other hand, the equatorial Atlantic warm SST bias is reduced despite a deeper D20, indicating that the SST change in this case is dictated by changes in radiative heating. The positioning of the Gulf Stream and the overlying winds—and not radiative biases—likely influence the North Atlantic SST bias, which remains almost unchanged among the prototypes. For example, a possible deficiency in simulating warm-core rings (which may be due to the ocean model resolution) along the Gulf Stream may explain the persistence of cold SST bias along the western North Atlantic.

The zonal wind stress, $\boldsymbol{\tau}_x$, is generally weak in all prototypes compared to ERA5. The largest improvement in $\boldsymbol{\tau}_x$ is noticeable in the Indian Ocean in P6, where the zonal pattern of the westerlies is more accurately represented than in either P5 or P7 (Figs. 9.22, 9.23). This plays a role in improving the SST bias in the tropical Indian Ocean in P6. Along with a decrease in positive net radiation bias, a deeper D20 driven by $\boldsymbol{\tau}_x$ contributes to improved SST mean state forecast in P6. Specifically, westerlies in the northern Indian Ocean are much closer to ERA5 in P6 in MJJ, indicating improved monsoonal winds over the Indian subcontinent. Surface currents are also significantly improved in P6 and much closer to OSCAR (not shown). Much of the improvement in the SST bias is retained in P7, except for a slight increase in warm bias in the eastern equatorial regions, which likely comes from an increase in net radiative heating bias in P7. The mean $\boldsymbol{\tau}_x$ bias however increases in P7, and that affects the equatorial D20 and the subsurface temperature. The warming of the subsurface (a deeper D20) in P6 is reversed in P7, but only seasonally in MJJ. Such reversal of subsurface temperatures is not obvious in OND (Fig. 9.21).

12.7 Land

Land surface processes play a key role in defining the partitioning of radiation, heat, moisture and momentum fluxes at the land-atmosphere interface.

A change in the parameterization of thermal roughness between P3.1 and P4 (reverted in P5) provided an opportunity to illustrate its impact on 2-meter temperatures via changes in the simulated skin temperature due to the modification of land-atmosphere heat exchanges.

The Noah land surface model was used in P1–P6, and replaced by Noah-MP in P7. The additional features of Noah-MP have the potential to significantly improve the realism of the land surface, hydrology, and atmosphere interactions. Analysis of the P7 forecast suggests that tuning is required for fully capitalizing on this potential.

Snow in P7 is found to be initialized too thin and compacting overly rapidly, resulting in thinner snow cover and consequently increased upward ground heat flux, and lower albedoes. Both

processes lead to warmer surface skin temperatures and warming of the 2-m atmospheric temperature. The impact of ground heat flux increase is evident at high latitudes during cold seasons, while the impact of albedo changes dominates in glacial regions during warm seasons.

Even though the same stomatal resistance formulation is used in both the Noah and Noah-MP models, Noah-MP produces significantly lower latent heat flux for the same soil moisture. As a result, the latent heat flux over humid tropical land is overly low in P7, due to both drier soil initial conditions and changes in the parameterization introduced in Noah-MP and is the root cause of the increased temperatures in these regions, though shallow model soils (2 meters) may also contribute.

Over deserts regions, the difference between P7 and the preceding prototype shows strongly reduced surface upward shortwave radiation regardless of the sign of change in downward shortwave radiation, implying a decreased albedo and thus increased absorption of solar radiation leading to overly high daytime temperatures. The difference is due to how soil albedo is prescribed in the Noah and Noah-MP configurations.

A basin-scale water budget analysis over ten large river basins with diverse climate regimes demonstrates improvement in eight of the ten basins. The partitioning of precipitation into evapotranspiration shows that P7 has lower ET in all basins, and substantially lower in tropical basins. The reduced ET is a slight degradation for the lower-precipitation Arctic basins, but an improvement in all temperate climate basins.

12.8 Aerosols

A one-way, atmosphere-coupled aerosol prediction model (UFS-Aerosols) was introduced into the coupled UFS in prototype P7.2. The aerosol predictions were evaluated in terms of total and species-specific aerosol optical depths (AOD) from a limited set of experiments.

The P7.2 AOD forecasts initialized in four different seasons of 2012 and 2013 show promising results when validated against MERRA-2, which uses aerosol data assimilation for AOD analysis. The global RMSE of total AOD ranges between 0.15 and 0.30, and is smallest in the October simulations, followed by January, July, and April.

The largest discrepancies between P7.2 and MERRA-2 AODs may be attributed to differences in dust concentrations, followed by either OC or sulfates depending on the season and location. While OC and sulfate concentrations are ultimately dependent on prescribed emissions, dust surface emission fluxes are dynamically modeled by the FENGSHA dust scheme as a function of friction velocity and high-resolution land surface properties provided as external inputs. Independent UFS-Aerosols evaluation studies (Siyuan Wang, NOAA/OAR/CSL, internal communication), performed with atmosphere-only coupling at C96 horizontal resolution, showed agreement between the predicted and observed dust distributions during all four NASA's Atmospheric Tomography Mission (ATom) campaigns. This may suggest that the distinct dust

schemes—FENGSHA in UFS-Aerosols and Ginoux in GEOS-5—may be partially responsible for the observed dust differences between the P7.2 and MERRA-2 datasets.

13. Conclusions and Further Work

This NCEP technical note describes the development and testing of a UFS-based coupled modeling system configurable into MWR and S2S applications targeting GFSv17, GEFSv13, and eventually SFSv1. The system development is coordinated with UFS-R2O objectives and includes significant efforts of the wider UFS community. The initial transition to operations of this coupled modeling system is tentatively scheduled for FY24 for MRW (GFSv17) and subseasonal (GEFSv13) predictions, and FY26 for seasonal (SFSv1) predictions.

The prototype coupled global modeling system for MRW and S2S applications uses CMEPS architecture for inter-component coupling and CCPP infrastructure for dynamics-physics coupling in the atmospheric model. All the coupled model components are combined to form a single UFS. The CMEPS coupling infrastructure makes it relatively easy to add, replace or remove component models. At present, the coupled UFS consists of the FV3-based atmospheric model, MOM6 ocean model, CICE6 sea-ice model, WAVEWATCH III wave model and GOCART aerosol model. The land surface model, Noah MP, is currently not a separate component module but directly connected with the atmospheric model. With the exception of GOCART, all components are two-way coupled. Work is underway for two-way coupling of GOCART as well, and for the addition of a lake model.

The coupled UFS described herein has undergone evaluations through a series of prototypes from P1 to P7 to assess the system's technical and scientific readiness towards finalizing the configuration of MRW and S2S applications for operational transition at NWS/NCEP. As documented in this note, **the present coupled UFS demonstrates deterministic skill exceeding that of the currently-operational coupled system, CFSv2**. Nevertheless, there clearly is significant room for further improvement, as illustrated by the biases and other errors in the system. Further advances and adjustments of the physics parameterizations, and introduction of a modular advanced workflow are part of ongoing developments to be implemented in the near future in the next system prototype, P8.

Priority development focus areas for the coupled UFS encompass both scientific and technical goals, and include the following:

- Implementing advanced data assimilation algorithms to improve the initialization of the coupled system.

- Revising and tuning physics parameterizations to reduce model systematic biases, such as those associated with land, convection, PBL, gravity-wave drag, microphysics, and land-atmosphere interactions.
- Improving computational efficiency and numerical instability of physics parameterizations.
- Improving the model representation of large-scale variability modes known to influence subseasonal predictability, such as the MJO and the stratospheric quasi-biennial oscillation (QBO).
- System reengineering to separate the land surface model as a modular component of the coupled modeling system, with Noah-MP and LM4 land models chosen for further development.
- Developing an ensemble prediction system and improving the representation of model uncertainty, particularly near model component interfaces.
- Adding reanalysis and reforecast capability to the coupled UFS.

Work in all these directions is currently underway.

All coupled model system developments—past, current and future—require strong infrastructure support that includes code management, coupled model development, workflow development, connection with data assimilation systems, pre-processing tools, website management, documentation and community support. Public code repositories, definitive practices and rules for code management, access and engagement from all active developers, open forums for discussions, and transparency of decision making are all necessary to facilitate continuing community involvement. These will need to continue in the future to support overall modeling infrastructure with core functionality of coupling interfaces for model components, parallel grid remapping, inter-component communication, modular design of atmospheric physics suites, and other utilities needed for model development on high-performance platforms.

Acknowledgments

The development of the global UFS-based coupled modeling system has drawn on the expertise of all multi-institutional cross-cutting development teams (Atmospheric Chemistry, Marine Components, Atmospheric Physics and Dynamics, Data Assimilation, Application Support, Modeling Infrastructure and Verification and Post processing Infrastructure) identified under UFS (https://ufscommunity.org/science/working-groups). Co-leads and members of these development teams represent the resident expertise in all requisite areas of successful development of the UFS-based future operational applications extending from weather to sub-seasonal to seasonal scales.

The progress in developing the global coupled UFS has been achieved with the substantial contributions from the scientists and developers listed below whose effort is gratefully acknowledged:

Ali Abdolali, Alistair Adcroft, Alper Altuntas, Barry Baker, Jian-Wen Bao, Alexei Belochitski, Rusty Benson, Ligia Bernardet, Alexandra Bozec, Patrick Campbell, Benjamin Cash, Eric Chassignet, Anning Cheng, Chris Cox, Gokhan Danabasoglu, Clara Draper, Rocky Dunlap, Gregory Frost, Bing Fu, George Gayno, Maria Gehne, Georg Grell, Stephen Griffies, Robert Grumbine, Hong Guang, Robert Hallberg, Matthew Harrison, Jian He, Katherine Hedstrom, Dom Heinzeller, Songyou Hong, Ethan Hughes, Christiane Jablonowski, Minsuk Ji, Sajal Kar, James Kinter, Walter Kolczynski, Shobha Kondragunta, Venkataramanaiah Krishnamurthy , Zachary Lawrence, Bin Li, Weiwei Li, Keith Lindsay, Qingfu Liu, Yanyun Liu, Gustavo Marques, Lawrence Marx, Matthew Masarik, Stuart McKeen, Jeff McQueen, Joseph Olson, Philip Pegion Brandon Reichl, Akella Santha, Rick Saylor, Eunkyo Seo, Andrew Shao, Eric Sinsky, Tanya Smirnova, Amy Solomon, Ivanka Stajner, Ruiyu Sun, Dustin Swales, Vijay Tallapragada, Gerhard Theurich, Michael Toy, Huug van den Dool, Mariana Vertenstein, Alan Wallcraft, Jun Wang, Siyuan Wang, Wanqiu Wang, Marshall Ward, Helin Wei, Yihua Wu, Rongqian Yang, Bo Yang, Valery Yudin, Li (Kate) Zhang, Xiaoyang Zhang, Jieshun Zhu.

This work is supported by NOAA/EMC, the NWS Office of Science and Technology Integration (OSTI) and the Weather Program Office (WPO) of NOAA Oceanic and Atmospheric Research (OAR) via multiple programs (UFS-R2O, JTTI, CTB). The authors would like to express their appreciation to Brian Gross, Neil Barton and Yuejian Zhu from EMC for their thoughtful reviews and comments, and Mary Hart for her copyediting and technical assistance. Finally, we would like to thank NCEP/EMC, UFS-R2O and ESRL/PSL management for their support.

References

Adcroft, A., W. Anderson, V. Balaji, C. Blanton, M. Bushuk, C. O. Dufour, and coauthors, 2019: The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *J. Adv. Model. Earth Syst.*, **11**, https://doi.org/10.1029/2019MS001726

Ardhuin, F., E. Rogers, A. V. Babanin, J.-F. Filipot, R. Magne, A. Roland, A. van der Westhuysen, P. Queffeulou, J.-M. Lefevre, L. Aouf, and F. Collard, 2010: Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. *J. Phys. Oceanogr.*, **40**(9), 1917–1941, https://doi.org/10.1175/2010JPO4324.1

Augustine, J. A., J. J. DeLuisi, and C. N. Long, 2000: SURFRAD—A national surface radiation budget network for atmospheric research, *Bull. Amer. Meteorol. Soc.*, **81**, 2341–2357, https://doi.org/10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2

Bailey, D. A., M. M. Holland, A. K. DuVivier, E. C. Hunke, and A. K. Turner, 2020: Impact of a new sea ice thermodynamic formulation in the CESM2 sea ice component. *J. Adv. Model. Earth Syst.,* **12**(11), https://doi.org/10.1029/2020MS002154

Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME, *J. Climate*, **27**(15), 5891–5906, https://doi.org/10.1175/JCLI-D-13-00597.1

Behringer, D.W., and Y. Xue, 2004: Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. Eighth Symposium on Integrated Observing and Assimilation

Systems for Atmosphere, Oceans, and Land Surface, *8th Symposium on Integrated Observing and Assimilation System for Atmosphere, Ocean, and Land Surface/84th Annual AMS Meeting*, Seattle, WA, 2.3, https://origin.cpc.ncep.noaa.gov/products/people/yxue/pub/13.pdf [*Accessed May 20, 2022]*

Bengtsson, L., M. Steinheimer, P. Bechtold, and J. F. Geleyn, 2013: A stochastic parameterization for deep convection using cellular automata. *Q. J. Roy. Meteorol. Soc.*, **139**(675), 1533–1543, https://doi.org/10.1002/qj.2108

Bengtsson, L., J. W. Bao, P. Pegion, C. Penland, S. Michelson, and J. Whitaker, 2019: A model framework for stochastic representation of uncertainties associated with physical processes in NOAA's next generation global prediction system (NGGPS). *Mon. Weather Rev.*, **147**(3), 893–911, https://doi.org/10.1175/MWR-D-18-0238.1

Bengtsson, L., J. Dias, S. Tulich, M. Gehne, and J. W. Bao, 2021: A stochastic parameterization of organized tropical convection using cellular automata for global forecasts in NOAA's Unified Forecast System. *J. Adv. Model. Earth Syst.*, **13**(1), https://doi.org/10.1029/2020MS002260

Bonjean, F., and G. S. E. Lagerloef, 2002: Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean. *J. Phys. Oceanogr.*, **32**(10), 2938–2954, https://doi.org/10.1175/1520-0485(2002)032%3C2938:DMAAOT%3E2.0.CO;2

Brassington G. B., M. J. Martin, H. L. Tolman, S. Akella, M. Balmeseda, C. R. S. Chambers, E. Chassignet, J. A. Cummings, Y. Drillet, P. A. E. M. Jansen, P. Laloyaux, D. Lea, A. Mehra, I. Mirouze, H. Ritchie, G. Samson, P.A. Sandery, G.C. Smith, and M. Suarez, 2015: Progress and challenges in short-to medium-range coupled prediction. *J. Oper. Oceanogr.*, **8**(sup2): s239-s258, https://doi.org/10.1080/1755876X.2015.1049875

Carn, S., 2019: Multi-Satellite Volcanic Sulfur Dioxide L4 Long-Term Global Database V3, Greenbelt, MD, USA, Goddard Earth Science Data and Information Services Center (GES DISC), [*Accessed 28 Feb 2019*], 10.5067/MEASURES/SO2/DATA404

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally, 1996, updated yearly: Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. *NASA National Snow and Ice Data Center Distributed Active Archive Center*, Boulder, CO, https://doi.org/10.5067/8GQ8LZQVL0VL

Cheng, A., and F. Yang, 2022: Direct radiative effects of aerosols on numerical weather forecast—a comparison of two aerosol datasets in the NCEP GFS. *NCEP Office Note* **507**, 30pp, https://doi.org/10.25923/rb1n-za92

Chin, M., R. B. Rood, S.-J. Lin, J.-F. Müller, A. M. Thompson, 2000a: Atmospheric sulfur cycle in the global model GOCART: Model description and global properties. *J. Geophys. Res.: Atmos.*, **105**(D20), 24671–24687, https://doi.org/10.1029/2000JD900384

Chin, M., D. L. Savoie, B. J. Huebert, A. R. Bandy, D. C. Thornton, T. S. Bates, P. K. Quinn, E. S. Saltzman, W. J. De Bruyn, 2000b: Atmospheric sulfur cycle in the global model GOCART: Comparison with field observations and regional budgets, *J. Geophys. Res.: Atmos.*, **105**(D20), 24689–24712, https://doi.org/10.1029/2000JD900385

Choulga, M., E. Kourzeneva, E. Zakharova, and A. Doganovsky, 2014: Estimation of the mean depth of boreal lakes for use in numerical weather prediction and climate modeling, *Tellus A: Dyn. Meteorol. Oceanogr.*, **66**(1), https://doi.org/10.3402/tellusa.v66.21295

Collow, T., Y. Liu, W. Wang, and A. Kumar, 2019: Overview of the CPC Sea Ice Initialization System (CSIS) and its use in experimental sea ice forecasting at the NOAA Climate Prediction Center. *Science and Technology Infusion Climate Bulletin*, NOAA NWS, 44th NOAA Annual Climate Diagnostics and Prediction Workshop, Durham, NC, 22–24 October 2019, https://www.nws.noaa.gov/ost/climate/STIP/44CDPW/44cdpw-TCollow.pdf [*Accessed May 20, 2022*]

Darmenov, A. and A. da Silva 2015: The Quick Fire Emissions Dataset (QFED): Documentation of versions 2.1, 2.2 and 2.4, NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TM-2015-104606, V**38**, http://gmao.gsfc.nasa.gov/pubs/docs/Darmenov796.pdf [*Accessed Jun 8, 2022*]

Doelling, D. R., N. G. Loeb, D. F. Keyes, M. L. Nordeen, D. Morstad, C. Nguyen, B. A. Wielicki, D. F. Young, and M. Sun, 2013: Geostationary Enhanced Temporal Interpolation for CERES Flux Products, *J. Atmos. Oceanic Technol.,* **30**(6), 1072–1090, https://doi.org/10.1175/JTECH-D-12-00136.1

Dong, X., J. S. Fu, K. Huang, D. Tong, and G. Zhuang, 2016: Model development of dust emission and heterogeneous chemistry within the Community Multiscale Air Quality modeling system and its application over East Asia. *Atmos. Chem. Phys.*, **16**, 8157–8180, doi:10.5194/acp-16-8157-2016.

Entekhabi, D., S. Yueh, P. E. O'Neill, K. H. Kellogg, and Coauthors, 2014: SMAP Handbook. Jet Propulsion Laboratory Publ, JPL, NASA. 400–1567, 182pp. https://smap.jpl.nasa.gov/system/internal_resources/details/original/178_SMAP_Handbook_FINAL_1_JULY_2014_Web.pdf [*Accessed May 20 2022*].

Gentile, E.S., S. L. Gray, J. F. Barlow, H. W. Lewis, and J. M. Edwards, 2021: The impact of atmosphere–ocean–wave coupling on the near-surface wind speed in forecasts of extratropical cyclones. *Boundary Layer Meteorol*.,**180**, 105–129, https://doi.org/10.1007/s10546-021-00614-4

Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. Holben, O. Dubovik, and S.-J. Lin, 2001: Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geophys. Res.*, **106**, 20 255–20 273, https://doi.org/10.1029/2000JD000053

Global Modeling and Assimilation Office (GMAO), 2015: MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly,Time-averaged,Single-Level,Assimilation,Aerosol Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), [*Accessed: Jun 8, 2022*], https://doi.org/10.5067/KLICLTZ8EM9D

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. Geophys. Res.: Oceans*, **118**, 6704–6716, https://doi.org/10.1002/2013JC009067

Guenther, A. B., X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, and X. Wang, 2012: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, *Geosci. Model Dev.*, **5**, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012

Guilyardi, E., A. Wittenberg, A. Fedorov, M. Collins, C. Wang, A. Capotondi, G. J. van Oldenborgh, and T. Stockdale, 2009: Understanding El Niño in Ocean–Atmosphere General Circulation Models: Progress and Challenges, *Bull. Amer. Meteorol. Soc*., **90**(3), 325–340, https://doi.org/10.1175/2008BAMS2387.1

Hamill, T. M., J. S. Whitaker, A. Shlyaeva, G. Bates, S. Fredrick, P. Pegion, E. Sinsky, Y. Zhu, V. Tallapragada, H. Guan, X. Zhou, and J. Woollen, 2022: The Reanalysis for the Global Ensemble Forecast System, Version 12. *Mon. Weather Rev.*, **150**(1), 59–79, https://doi.org/10.1175/MWR-D-21-0023.

Han, J., and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Weather Forecasting*, **26**(4), 520–533, https://doi.org/10.1175/WAF-D-10-05038.1

Han, J., M. L. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. Fletcher, and C. S. Bretherton, 2016: Implementation in the NCEP GFS of a hybrid eddy-diffusivity mass-flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Weather Forecasting*, **31**(1), 341–352, https://doi.org/10.1175/WAF-D-15-0053.1

Han, J., and C. S. Bretherton, 2019: TKE-Based Moist Eddy-Diffusivity Mass-Flux (EDMF) Parameterization for Vertical Turbulent Mixing. *Weather Forecasting*, 34, 869–886, https://doi.org/10.1175/WAF-D-18-0146.1

Han J., W. Li, F. Yang, E. Strobach, W. Zheng, and R. Sun, 2021: Updates in the NCEP GFS cumulus convection, vertical turbulent mixing, and surface layer physics. *NCEP Office Note 505*, 18 pp., https://doi.org/10.25923/cybh-w893

Harris, L., X. Chen, W. Putman, L. Zhou, and J. H. Chen, 2021: A scientific description of the GFDL Finite-Volume Cubed-Sphere Dynamical Core. *NOAA technical memorandum OAR GFDL*, **2021**(001), https://doi.org/10.25923/6nhs-5897

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J.Nicolas, C. Peubey, R. Radu, R., D. Schepers, A. Simmons, C. Soci, and Coauthors, 2020. The ERA5 global reanalysis. *Q. J. R. Meteorolog. Soc.*, **146**(730), 1999–2049, https://doi.org/10.1002/qj.3803

Hoesly, R, P. O'Rourke, C. Braun, L. Feng, S. J. Smith, T. Pitkanen, J. J. Seibert, L. Vu, P. Muwan, R. Bolt, B. Goldstein, N. Kholod, 2019: Community Emissions Data System (Version Dec-23-2019). *Zenodo*. http://doi.org/10.5281/zenodo.3592073

Holte, J., L. D. Talley, J. Gilson, and D. Roemmich, 2017: An Argo mixed layer climatology and database, *Geophys. Res. Lett.*, **44**, 5618–5626, https://doi.org/10.1002/2017GL073426

Huffman, G.J., E.F. Stocker, D.T. Bolvin, E.J. Nelkin, J. Tan, 2019: GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06, Edited by A. Savtchenko, Greenbelt,

MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/GPM/IMERGDF/DAY/06

Janssens-Maenhout, G., M. Crippa, D. Guizzardi, F. Dentener, M. Muntean, G. Pouliot, T. Keating, Q. Zhang, J. Kurokawa, R. Wankmüller, H. Denier van der Gon, J. J. P. Kuenen, Z. Klimont, G. Frost, S. Darras, B. Koffi, and M. Li, 2015: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, *Atmos. Chem. Phys.*, **15**, 11411–11432, https://doi.org/10.5194/acp-15-11411-2015

Johansson, A., C. Thiaw, and S. Saha, 2007: CFS retrospective forecast daily climatology in the EMC/NCEP CFS public server, 42pp. http://cfs.ncep.noaa.gov/cfs.daily.climatology.doc [*Accessed Jun 8, 2022*]

Kato, S., F. G. Rose, D. A. Rutan, T. E. Thorsen, N. G. Loeb, D. R. Doelling, X. Huang, W. L. Smith, W. Su, and S.-H. Ham, 2018: Surface irradiances of Edition 4.0 Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data product, *J. Clim.*, **31**, 4501–4527, https://doi.org/10.1175/JCLI-D-17-0523.1

Kim, H., M. A. Janiga, and K. Pegion, 2019: MJO propagation processes and mean biases in the SubX and S2S reforecasts. *J. Geophys. Res. Atmosph.*, **124**, 9314– 9331, https://doi.org/10.1029/2019JD031139

Kinter, J., V. Tallapragada, J. Whitaker (leads), 2020: Unified Forecast System Research-to-Operations (UFS-R2O) project proposal. https://www.weather.gov/media/sti/UFS-R2O-Project-Proposal-Public.pdf [*Accessed Jun 8, 2022*]

Krishnamurthy, V., J. Meixner, L. Stefanova, J. Wang, D. Worthen, S. Moorthi, B. Li, T. Sluka, and C. Stan, 2021: Sources of subseasonal predictability over CONUS during boreal summer. *J. Clim.*, **39**, 3273–3294, https://doi.org/10.1175/JCLI-D-20-0586.s1

Krishnamurthy, V., and C. Stan, 2022: Prediction of extreme events in precipitation and temperature over CONUS during the boreal summer in the UFS coupled model. *Clim. Dyn*. (2022), https://doi.org/10.1007/s00382-021-06120-0

Lee, H.-T., and NOAA CDR Program, 2011: NOAA Climate Data Record (CDR) of daily Outgoing Longwave Radiation (OLR), Version 1.2. *NOAA National Climatic Data Center*, https://doi.org/10.7289/V5SJ1HH2

Levy, R. C., S. Mattoo, L. A. Munchak, L. A. Remer, A. M, Sayer, F. Patadia, and N.C. Hsu, 2013: The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.*, **6**, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013

Liu, Y., W. Wang, A. Kumar, and T. Collow, 2019: Assessment of CPC Sea Ice Initialization System (CSIS) and CPC weekly experimental sea ice forecasts. *Science and Technology Infusion Climate Bulletin*, NOAA NWS, 44th NOAA Annual Climate Diagnostics and Prediction Workshop, Durham, NC, 22–24 October 2019, https://www.nws.noaa.gov/ost/climate/STIP/44CDPW/44cdpw-YLiu.pdf [*Accessed May 20, 2022*]

Lu, F., M. J. Harrison, A. Rosati, T. L. Delworth, X. Yang, Wi. F. Cooke, L. Jia, C. McHugh, N. C. Johnson, M. Bushuk, Y. Zhang, and A. Adcroft, 2020: GFDL's SPEAR seasonal

prediction system: Initialization and ocean tendency adjustment (OTA) for coupled model predictions. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002149, https://doi.org/10.1029/2020MS002149

Martens, B., D. G. Miralles, H. Lievens, R. van der Schalie, R. A. M. de Jeu, D. Fernández-Prieto, H. E. Beck, W. A. Dorigo, and N. E. C. Verhoest, 2017: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev.*, **10**, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017

Mecklenburg, S., M. Drusch, Y. H. Kerr, J. Font, M. Martin-Neira, S. Delwart, G. Buenadicha, N. Reul, E. Daganzo-Eusebio, R. Oliva, and R. Crapolicchio, 2012: ESA's soil moisture and ocean salinity mission: Mission performance and operations. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1354–1366, https://doi.org/10.1109/TGRS.2012.2187666

Meier, W. N., F. Fetterer, M. Savoie, S. Mallory, R. Duerr, and J. Stroeve, 2017: NOAA/NSIDC Climate Data Record of passive microwave sea ice concentration, Version 3. *NSIDC: National Snow and Ice Data Center,* Boulder, Colorado USA. https://doi.org/10.7265/N59P2ZTG

Mitchel, K. and Collaborators, 2005: The Community Noah Land-Surface Model (LSM) user's guide, public release version 2.7.1, https://ral.ucar.edu/sites/default/files/public/product-tool/unified-noah-lsm/Noah_LSM_USERGUIDE_2.7.1.pdf [*Accessed May 20, 2022*]

Mogensen, S. K., L. Magnusson, and J.-R. Bidlot, 2017: Tropical cyclone sensitivity to ocean coupling in the ECMWF. *J. Geophys. Res.: Oceans*, **122**(5), https://doi.org/10.1002/2017JC012753

NOAA, 2019: Upgrade NCEP Global Forecast Systems (GFS) to v15.1: Effective June 12, 2019. *Service Change Notice **19-40***. National Weather Service Headquarters, Silver Spring MD, https://www.weather.gov/media/notification/scn19-40gfs_v15_1.pdf *[Accessed May 20, 2022].*

NOAA, 2020: Announcement of upgrade to the Global Ensemble Forecast System (GEFS), coupled with the Global Wave Ensemble System (GWES) and the NEMS GFS aerosol component (NGAC): Effective September 23, 2020 and request for comments. *Service Change Notice **20-75***. National Weather Service Headquarters, Silver Spring MD, https://www.weather.gov/media/notification/SCN_20-75_GEFSv12_Changes.pdf *[Accessed May 20, 2022].*

NOAA, 2021: Upgrade NCEP Global Forecast Systems (GFS) to v16: Effective March 17, 2021. *Service Change Notice **21-20***, Updated. National Weather Service Headquarters, Silver Spring MD, https://www.weather.gov/media/notification/scn_21-20_gfsv16.0_aaa_update.pdf *[Accessed May 20, 2022].*

Pegion, K., B. P. Kirtman, E. Becker, D. C. Collins, E. LaJoie, R. Burgman, R. Bell, T. DelSole, D. Min, Y. Zhu, W. Li, E. Sinsky, H. Guan, J. Gottschalck, E. J. Metzger, N. P. Barton, D. Achuthavarier, J. Marshak, R. D. Koster, H. Lin, N. Gagnon, M. Bell, M. K.

Tippett, A. W. Robertson, S. Sun, S. G. Benjamin, B. W. Green, R. Bleck, and H. Kim, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment, *Bull. Amer. Meteorol. Soc.*, **100**(10), 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1

Randles, C. A., A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, J. Hair, Y. Shinozuka, and C. J. Flynn, 2017: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *J. Clim.* **30**(17), 6823–6850, https://doi.org/10.1175/JCLI-D-16-0609.1

Saha, S., S. Moorthi, H. L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, and H. Liu, 2010: The NCEP Climate Forecast System reanalysis. *Bull. Amer. Meteorol. Soc.*, **91**(8), 1015–1058, https://doi.org/10.1175/2010BAMS3001.1

Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker, 2014: The NCEP Climate Forecast System Version 2, *J. Clim.*, **27**(6), 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1

Seo, H., A. C. Subranian, A. J. Miller, N. R. Cavanaugh 2014: Coupled impacts of the diurnal cycle of sea surface temperature on the Madden-Julian oscillation. *J. Clim.,* **27**(22), 8422–8443, https://doi.org/10.1175/JCLI-D-14-00141.1

Shi, W., 2019: Global daily surface air temperature analyses, https://www.ftp.cpc.ncep.noaa.gov/precip/PEOPLE/wd52ws/global_temp/CPC-GLOBAL-T.pdf *[Accessed May 20, 2022].*

Shinoda, T., S. Pei, W. Wang, J. X. Fu, R.-C. Lien, H. Seo, A. Soloviev, 2021: Climate Process Team: Improvement of ocean component of NOAA Climate Forecast System relevant to Madden-Julian Oscillation Simulations. *J. Adv. Model. Earth Syst.*, **13**(12), e2021MS002658, https://doi.org/10.1029/2021MS002658

Slingo, J. M., P. Inness, R. Neale, S. Woolnough, G.-Y. Yang, 2003: Scale interactions on diurnal to seasonal timescales and their relevance to model systematic errors. *Ann. Geophys.*, **46**(1), 139–155, https://doi.org/10.4401/ag-3383

Smith, G. C., J. -M. Bélanger, F. Roy, P. Pellerin, H. Ritchie, K. Onu, M. Roch, A. Zadra, D. S. Colan, B. Winter, J. -S. Fontecilla, and D. Deacu, 2018: Impact of coupling with an ice–ocean model on global medium-range NWP forecast skill. *Mon. Weather Rev.*, **146**(4), 1157-1180, https://doi.org/10.1002/qj.2555

Stan, C., V. Krishnamurthy, H. Bai, B. Li, A. Mehra, J. Meixner, S. Moorthi, L. Stefanova, J. Wang, D. Worthen, and F. Yang, 2022: The impact of tropical Pacific SST biases on the S2S forecast skill over North America in the UFS coupled model. *J. Clim. (in review).*

Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch, 2007: OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system. *Proceedings of Oceans '07 IEEE Conference, 'Marine Challenges: Coastline to Deep Sea',* 18–21 June 2007, Aberdeen, UK, https://doi.org/10.1109/OCEANSE.2007.4302251

Taburet, G., F. Mertz, and J.-F. Legeais, 2021: Product user guide and specification (sea level). https://datastore.copernicus-climate.eu/documents/satellite-sea-

level/vDT2021/D3.SL.1-v2.0_PUGS_of_v2DT2021_SeaLevel_products_v1.1_APPROVED_Ver1.pdf [*Accessed May 20, 2022*]

Tallapragada, V., 2018: Development and implementation strategies for NOAA's Unified Forecast System for operational NWP at NCEP. In *EGU General Assembly Conference Abstracts*, v**20**, p. 10325, https://meetingorganizer.copernicus.org/EGU2018/EGU2018-10325-1.pdf [*Accessed May 31, 2022*]

Theurich, G., C. DeLuca, T. Campbell, F. Liu, K. Saint, M. Vertenstein, J. Chen, R. Oehmke, J. Doyle, T. Whitcomb, A. Wallcraft, M. Iredell, T. Black, A. M. Da Silva, T. Clune, R. Ferraro, P. Li, M. Kelley, I. Aleinov, V. Balaji, N. Zadeh, R. Jacob, B. Kirtman, F. Giraldo, D. McCarren, S Sandgathe, S. Peckham, and R. Dunlap IV, 2016: The Earth System Prediction Suite: Toward a coordinated U.S. modeling capability. *Bull. Amer. Meteorol. Soc.*, **97**(7), 1229–1247, https://doi.org/10.1175/BAMS-D-14-00164.1

Trier, S. B., M. A. LeMone, F. Chen, and K. W. Manning, 2011: Effects of surface heat and moisture exchange on ARW-WRF warm-season precipitation forecasts over the central United States. *Weather Forecasting*, **26**, 3–25, https://doi.org/10.1175/2010WAF2222426.1

Van den Dool, H. and S. Saha, 2008: Verification of daily CFS forecasts 1981-2005. *NOAA CTB-COLA,* Joint Seminar, May 28, 2008, https://www.nws.noaa.gov/ost/climate/STIP/CTB-COLA/Huug_052808.pdf [*Accessed May 20, 2022]*

Vecchi, G.A. and Wittenberg, A.T., 2010: El Niño and our future climate: where do we stand? *WIREs Clim. Change*, **1**, 260–270, https://doi.org/10.1002/wcc.33

Vellinga, M., D. Copsey, T. Graham, S. Milton, and T. Johns, 2020: Evaluating benefits of two-way ocean–atmosphere coupling for global NWP forecasts", *Weather Forecasting*, **35**(5), 2127-2144, https://doi.org/10.1175/WAF-D-20-0035.1

Vitart, F., 2017: Madden–Julian Oscillation prediction and teleconnections in the S2S database. *Q.J.R. Meteorol. Soc*, **143**, 2210–2220, https://doi.org/10.1002/qj.3079

Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H.-S. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Mallas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu, D.-J. Won, H. Xiao, R. Zaripov, and L. Zhang, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteorol. Soc.,* **98**(1),163–173, https://doi.org/10.1175/BAMS-D-16-0017.1

Wyser, K., 1998: The effective radius in ice clouds. *J. Clim.*, **11**, 1793–1802, https://doi.org/10.1175/1520-0442(1998)011<1793:TERIIC>2.0.CO;2

Xie, P., M. Chen, W. Shi, 2010: CPC unified gauge-based analysis of global daily precipitation. 24th Conf. on Hydrology, Atlanta, GA, *Amer. Meteorol. Soc.*, 2.3A, https://ams.confex.com/ams/90annual/techprogram/paper_163676.htm

Yudin, V. A., F. Yang., S. Karol, T. Fuller-Rowell, A. Kubaryk, H. Juang, S. Kar, J. Alpert, and Z. Li, 2020: The Unified Gravity Wave Physics in the Vertically Extended Atmosphere Models of NGGPS and UFS. *1st UFS Users' Workshop*,

https://dtcenter.org/sites/default/files/events/2020/4-valery-yudin.pdf [*Accessed May 20, 2022*]

Zhang, X., S. Kondragunta, J. Ram, C. Schmidt, and H.C. Huang, 2012: Near real time global biomass burning emissions product from geostationary satellite constellation. *J. Geophys. Re*s, **117**, D14201, https://doi.org/10.1029/2012JD017459

Zheng, W., H. Wei, Z. Wang, X. Zeng, J. Meng, M. Ek, K. Mitchell, and J. Derber, 2012: Improvement of daytime land surface skin temperature over arid regions in the NCEP GFS model and its impact on satellite data assimilation. *J. Geophys. Res*., **117**, D06117, https://doi.org/10.1029/2011JD015901

Zhou, X., Y. Zhu, D. Hou, B. Fu, W. Li, H. Guan, E. Sinsky, W. Kolczynski, X. Xue, Y. Luo, J. Peng, B. Yang, V. Tallapragada, and P. Pegion, 2022: The Development of the NCEP Global Ensemble Forecast System Version 12. *Weather Forecasting* (2022), https://doi.org/10.1175/WAF-D-21-0112.1

Zuo, H., M. A. Balmaseda, K. Mogensen, and S. Tietsche, 2018: OCEAN5: The ECMWF Ocean Reanalysis System and its real-time analysis component. *Technical Report 823*, ECMWF, Reading, UK, https://doi.org/10.21957/la2v0442

FIGURES
P1 Figures

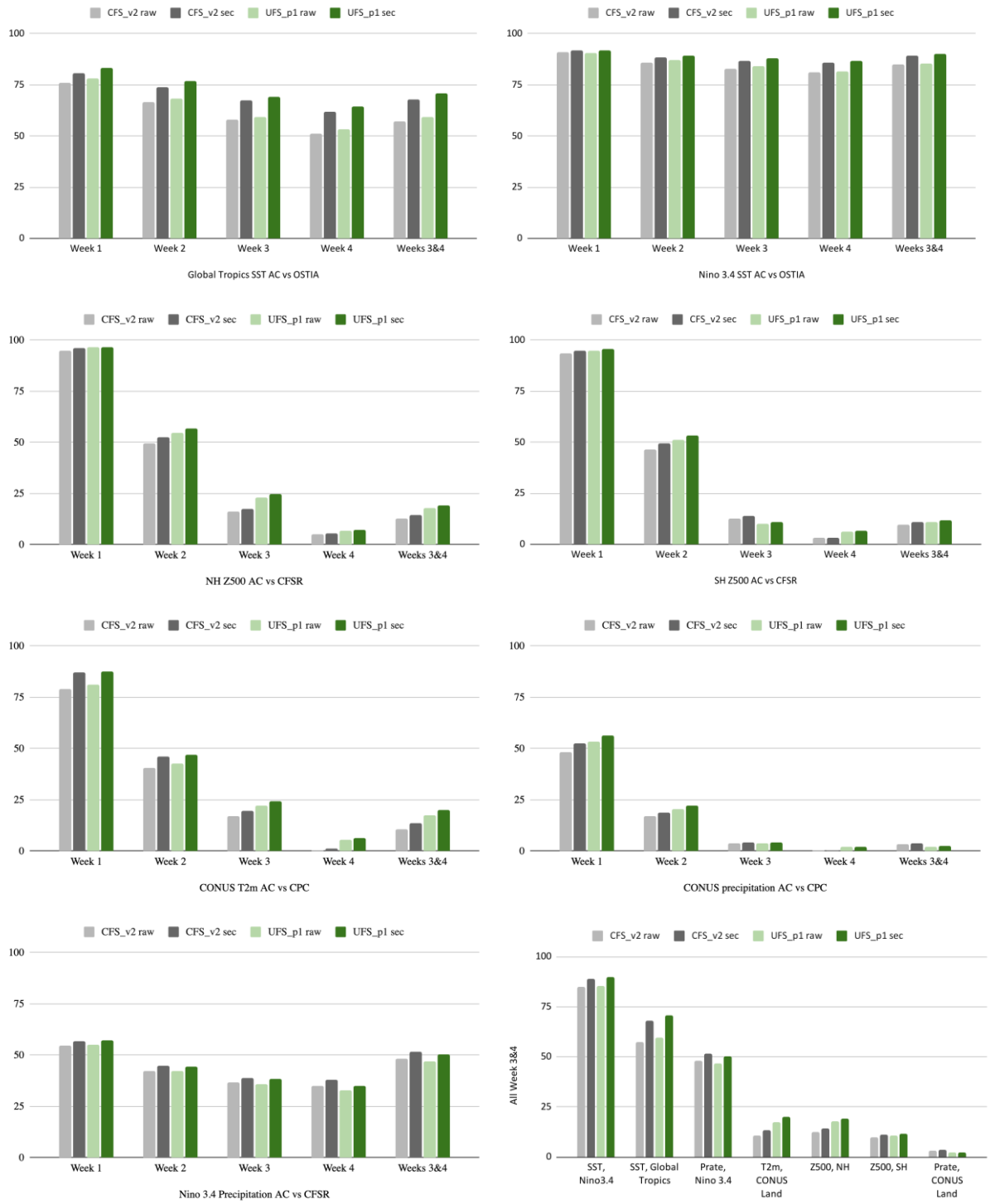

Figure 5.1: Anomaly correlation scores by forecast week of CFS_v2 (RAW-pale gray, SEC-dark gray) and of P1 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of anomaly correlation scores in Weeks 3&4.

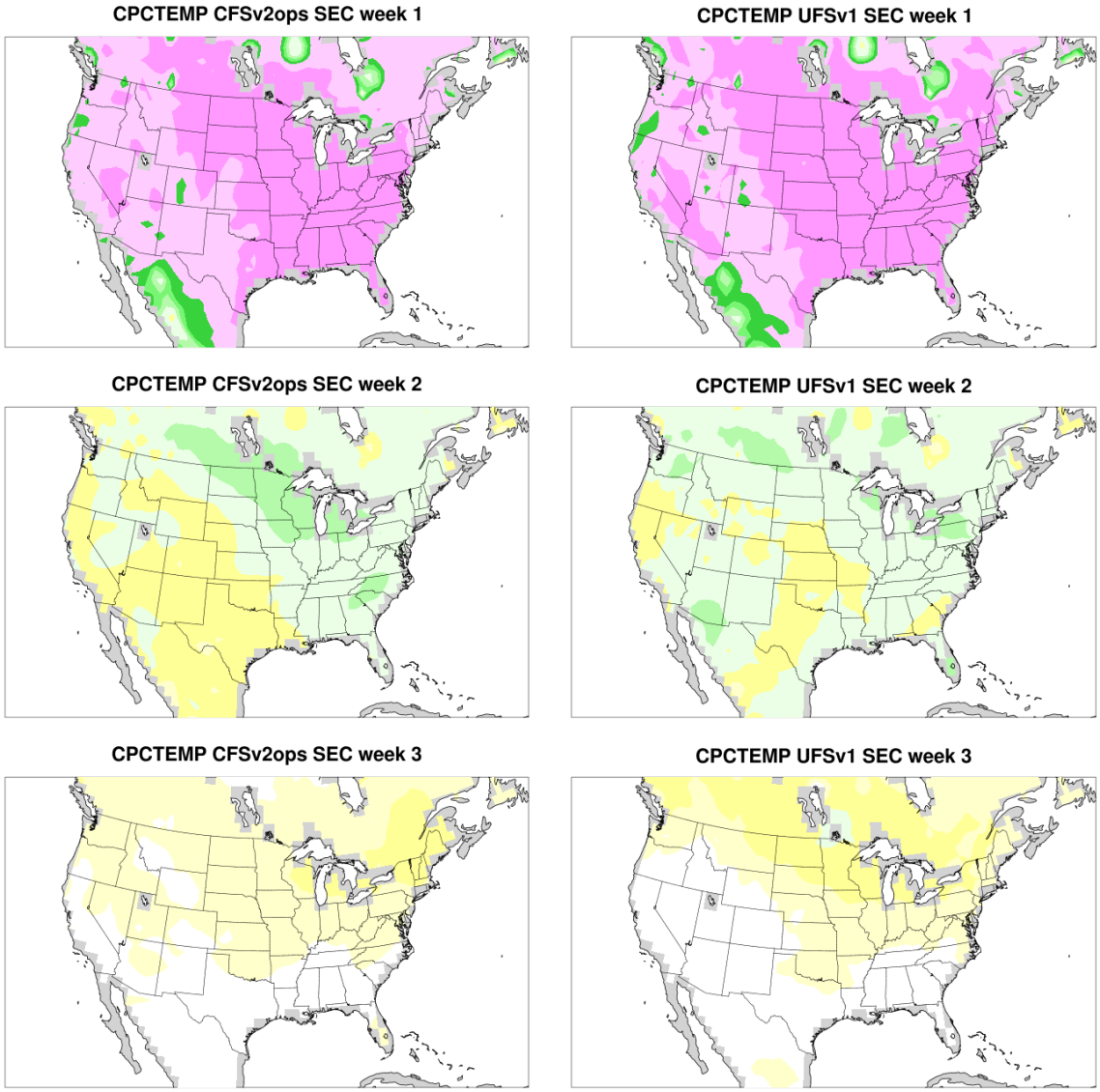

Figure 5.2: Week 1 (top row), Week 2 (middle row) and Week 3 (bottom row): Annually aggregated SEC AC maps for CONUS 2-m temperature in CFSv2 (left) and P1 (right).

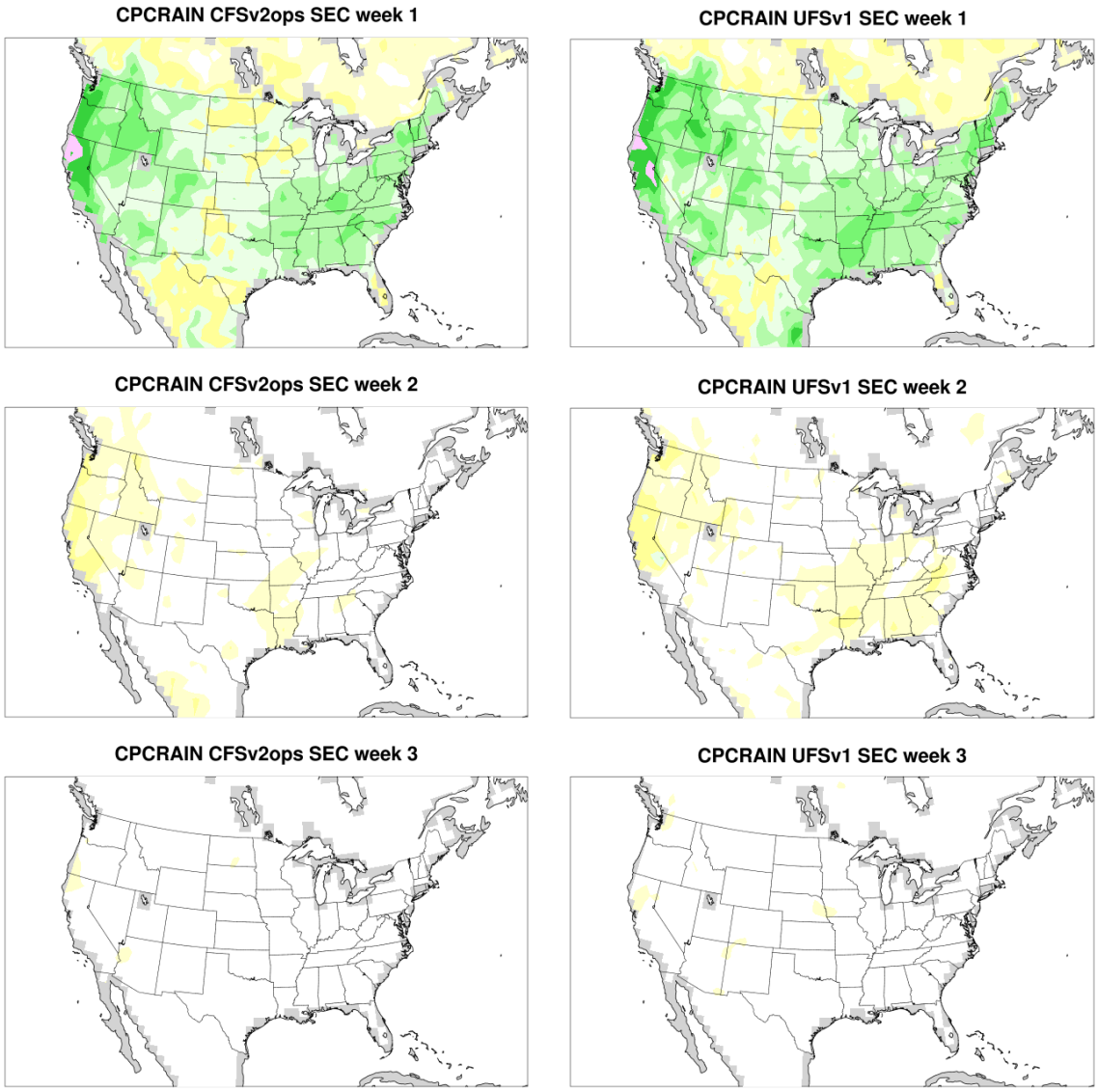

Figure 5.3: Week 1 (top row), week 2 (middle row) and week 3 (bottom row): Annually aggregated AC maps for CONUS 2-m precipitation in CFSv2 (left column) and P1 (right column).

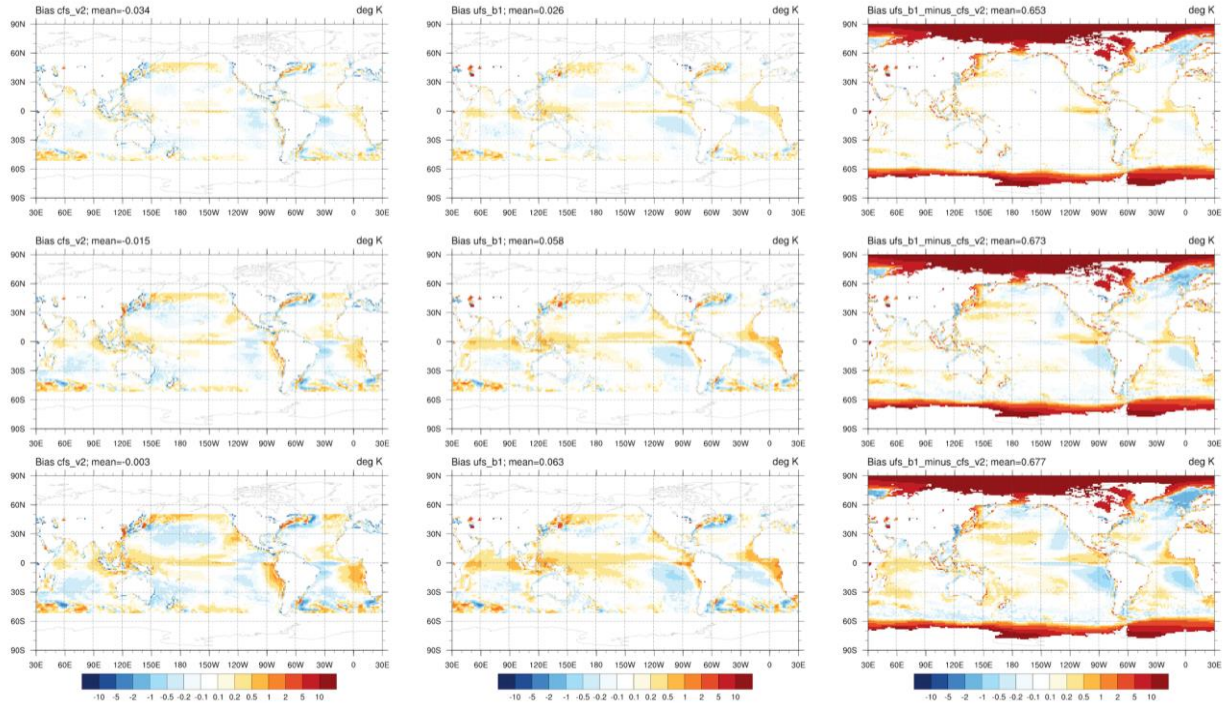

Figure 5.4: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean SST bias wrt OSTIA from 50°S to 50°N in CFSv2 (left column) and P1 (middle column); Difference (P1 minus CFSv2) in surface temperatures over ocean grid points, including sea ice (right column).

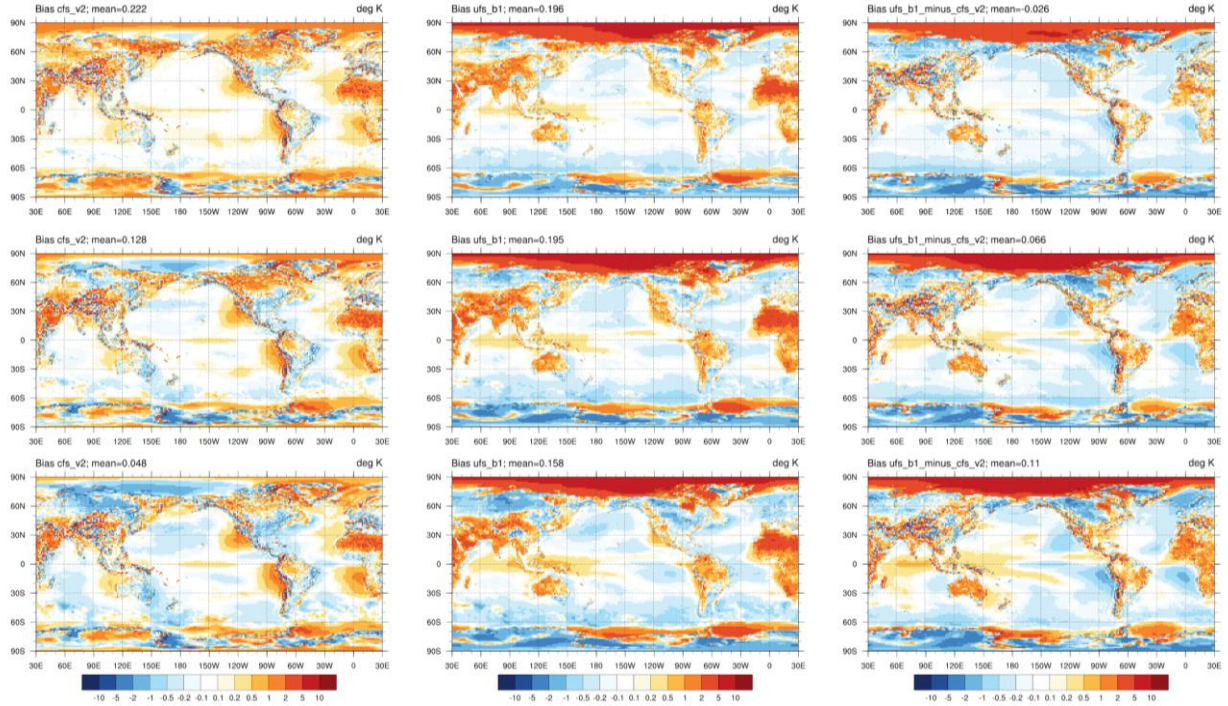

Figure 5.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in CFSv2 (left column) and P1 (middle column); Difference (P1 minus CFSv2) in 2-m air temperature (right column).

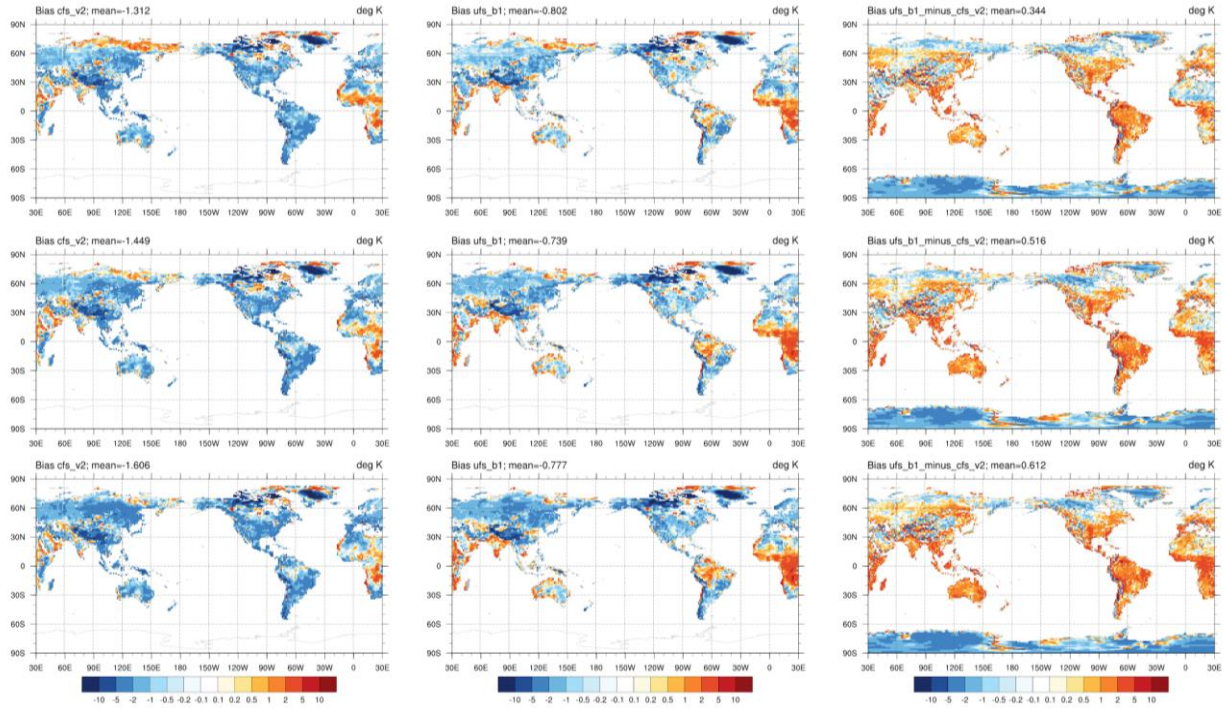

Figure 5.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in CFSv2 (left column) and P1 (middle column); Difference (P1 minus CFSv2) in 2-m daily maximum air temperature (right column).

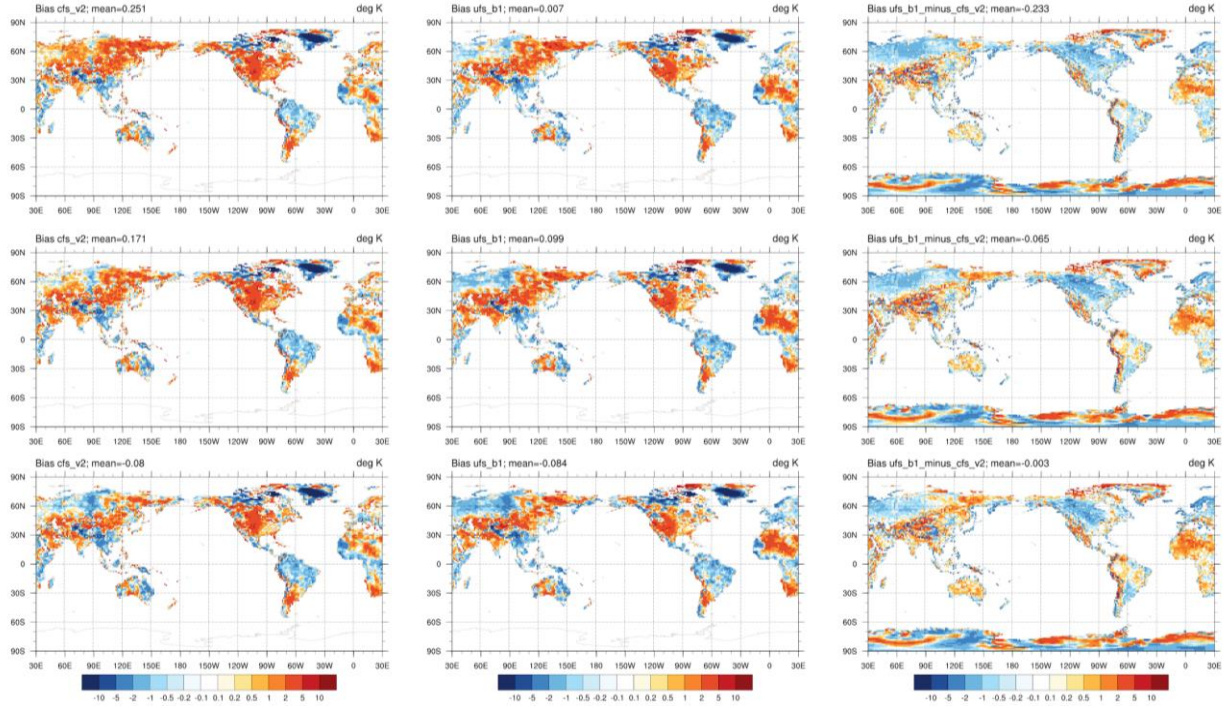

Figure 5.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in CFSv2 (left column) and P1 (middle column); Difference (P1 minus CFSv2) in 2-m daily minimum air temperature (right column).

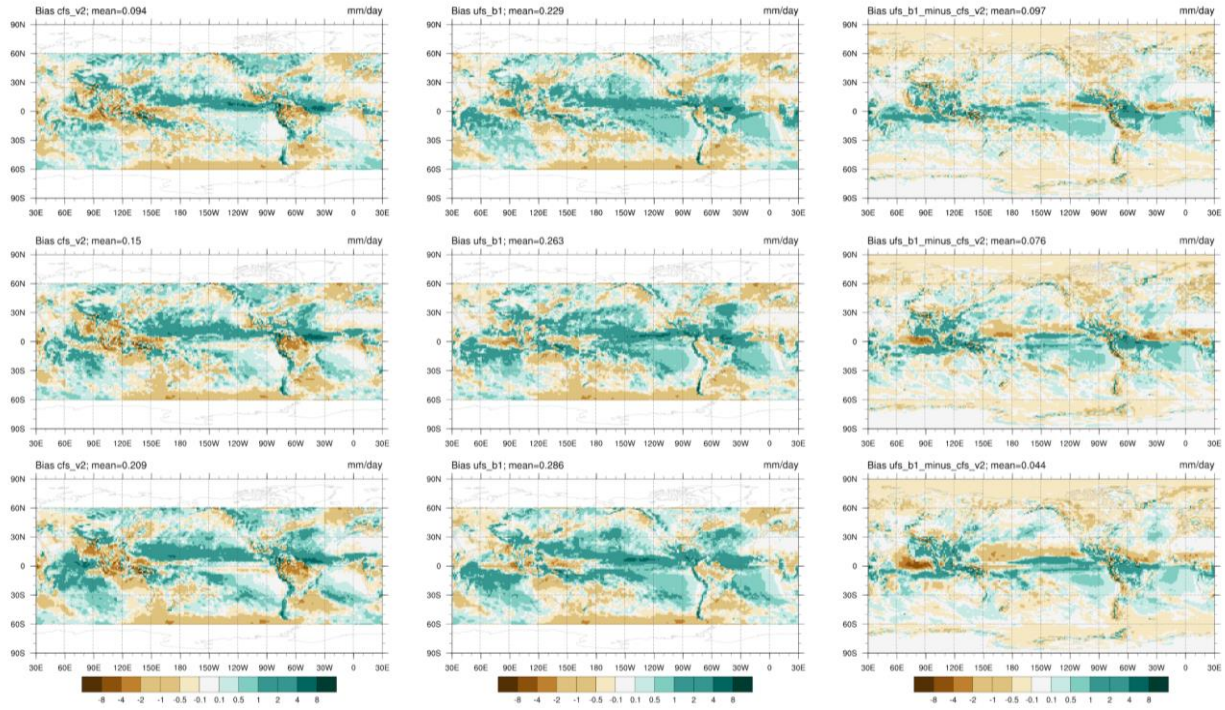

Figure 5.8: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in CFSv2 (left column) and P1 (middle column); Difference (P1 minus CFSv2) in precipitation rate (right column).

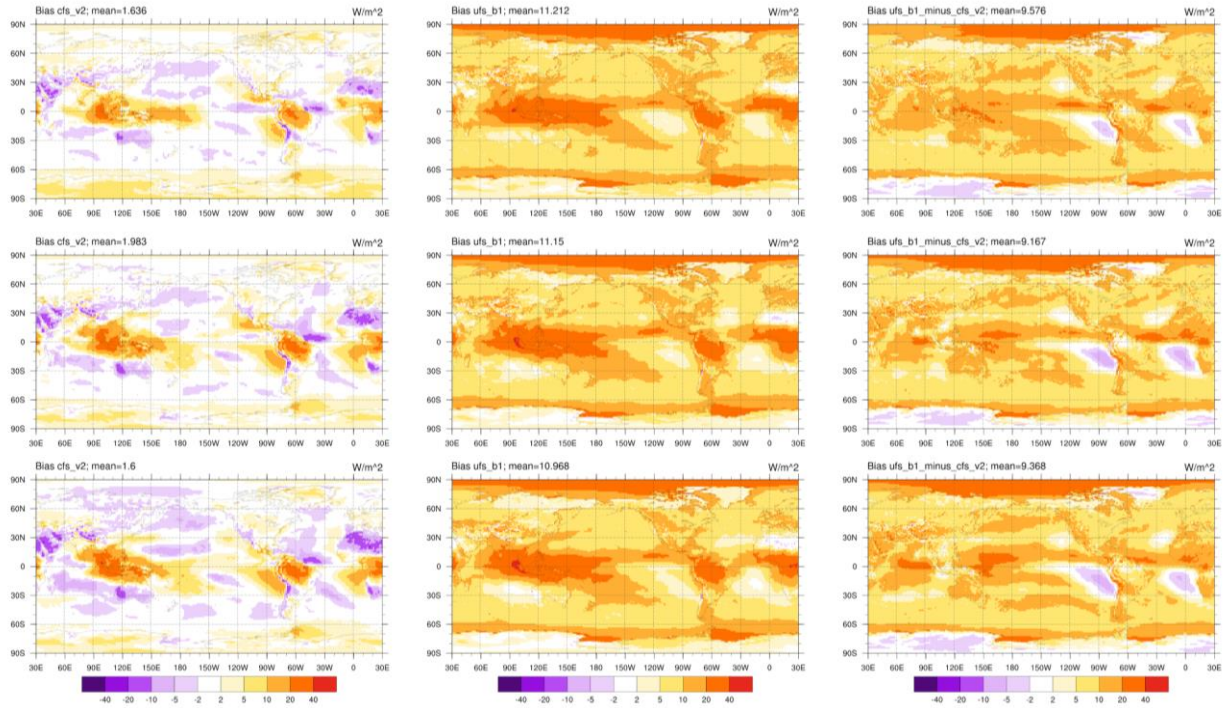

Figure 5.9: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in CFSv2 (left column) and P1 (middle column); Difference (P1 minus CFSv2) in OLR (right column).

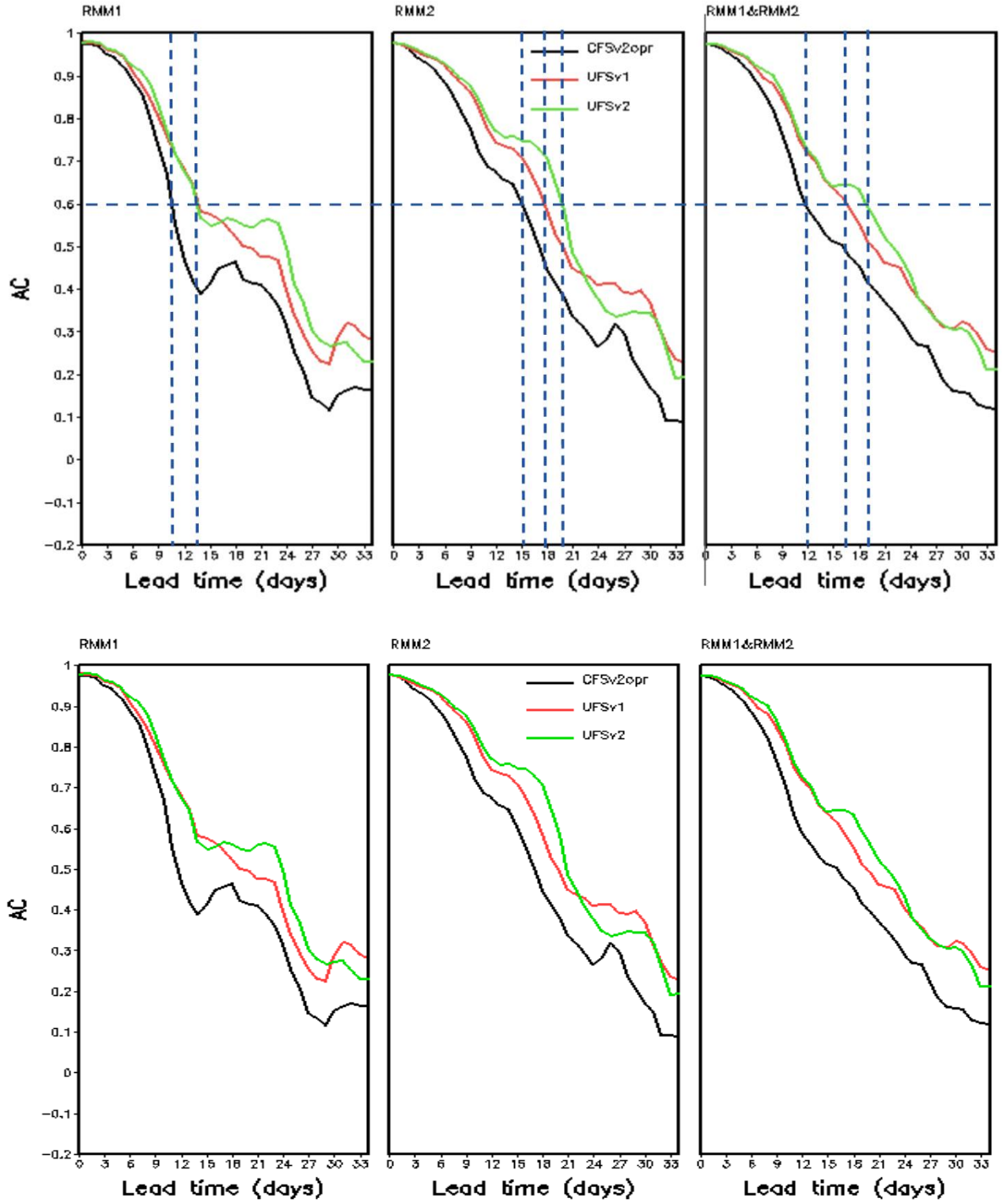

Figure 5.10: Top: AC of MJO RMM1 (left), RMM2 (middle) and combined RMM1&RMM2 (right). Bottom: AC of equatorial filtered zonal wind at 850mb (left) and 200mb (middle), and equatorial filtered OLR (right) in CFSv2 (black line), P1 (red line) and P2 (green line).

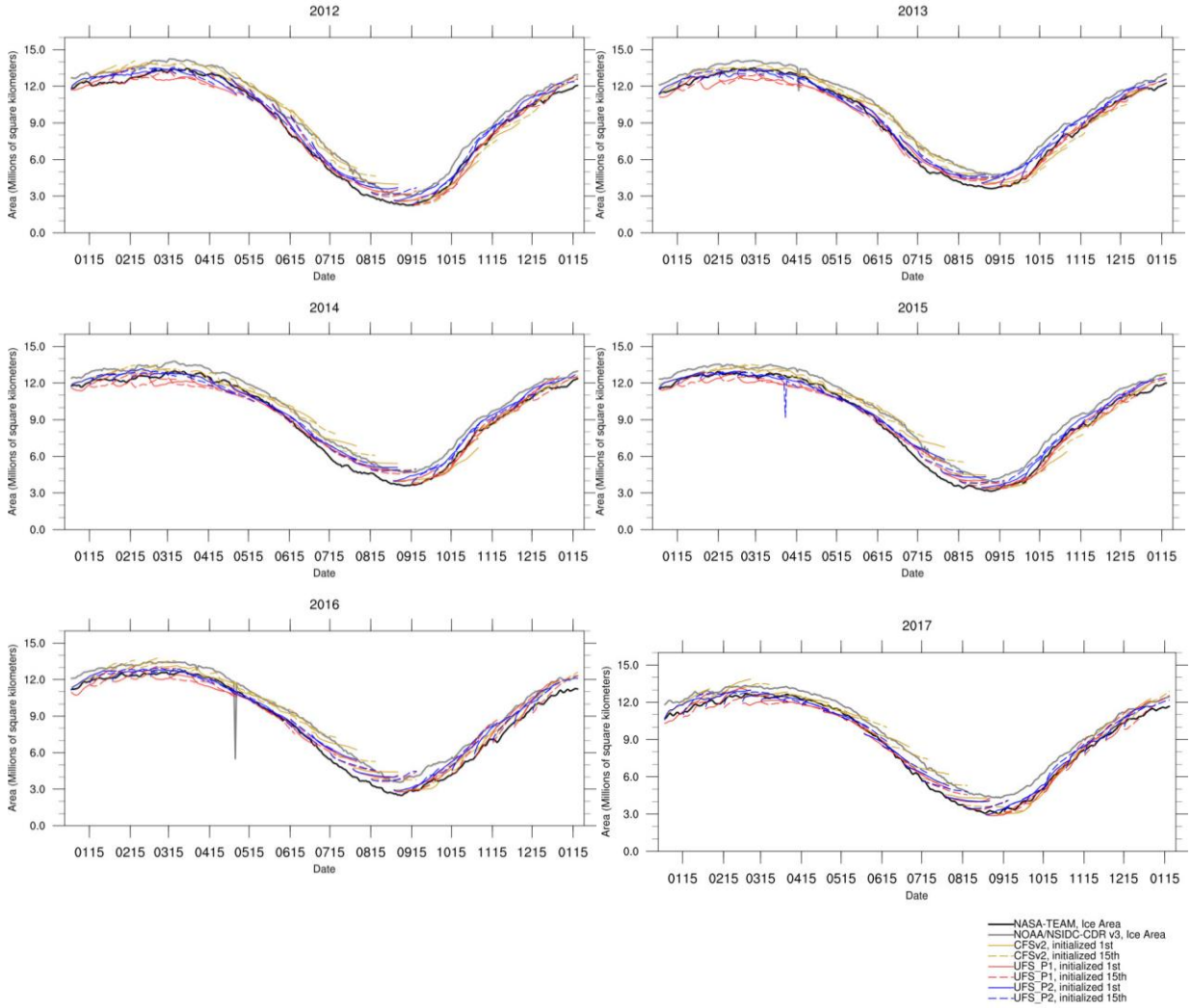

Figure 5.11: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow) and P1 (red). Thick lines correspond to the observational estimate derived from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date.

P2 Figures

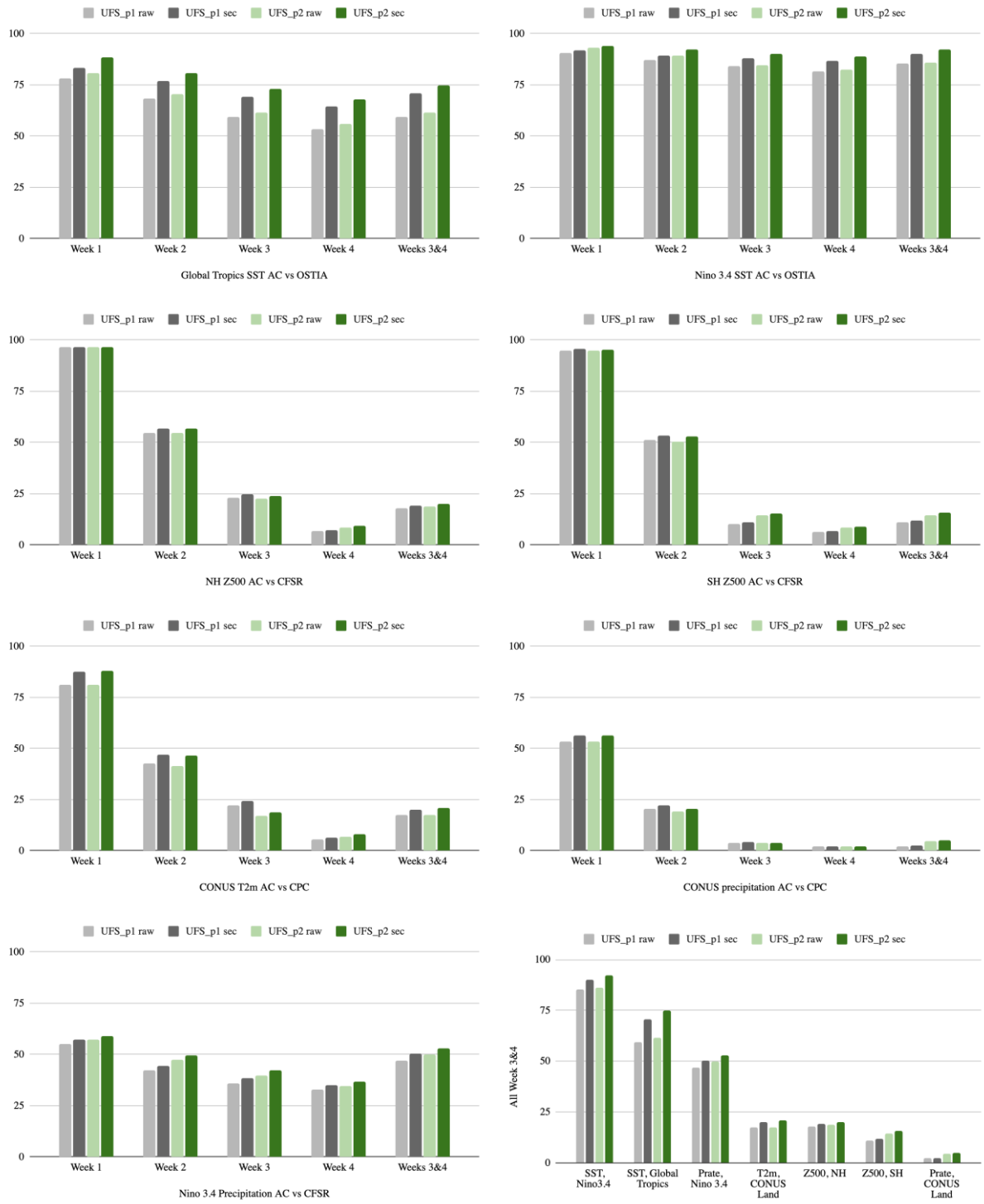

Figure 6.1: Anomaly correlation scores by forecast week of P1 (RAW-pale gray, SEC-dark gray) and P2 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of the anomaly correlation scores for Weeks 3&4.

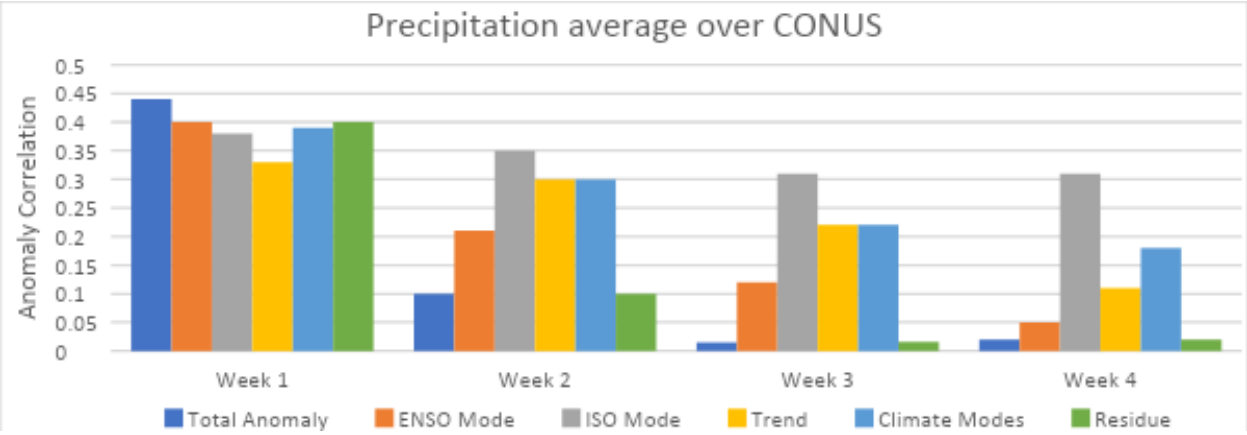

Figure 6.2: Spatial correlation of weekly-average precipitation over CONUS between CPC and P2 reforecasts during June-September 2011-2017. Daily reforecasts anomalies (dark blue) are decomposed into two modes of variability [ENSO (orange) and ISO (gray)] and a long-term trend (yellow). The residue (green) is defined as the difference between the total daily anomaly and the sum of the modes and trend. The forecast skill of this sum is shown as the climate modes (light blue).

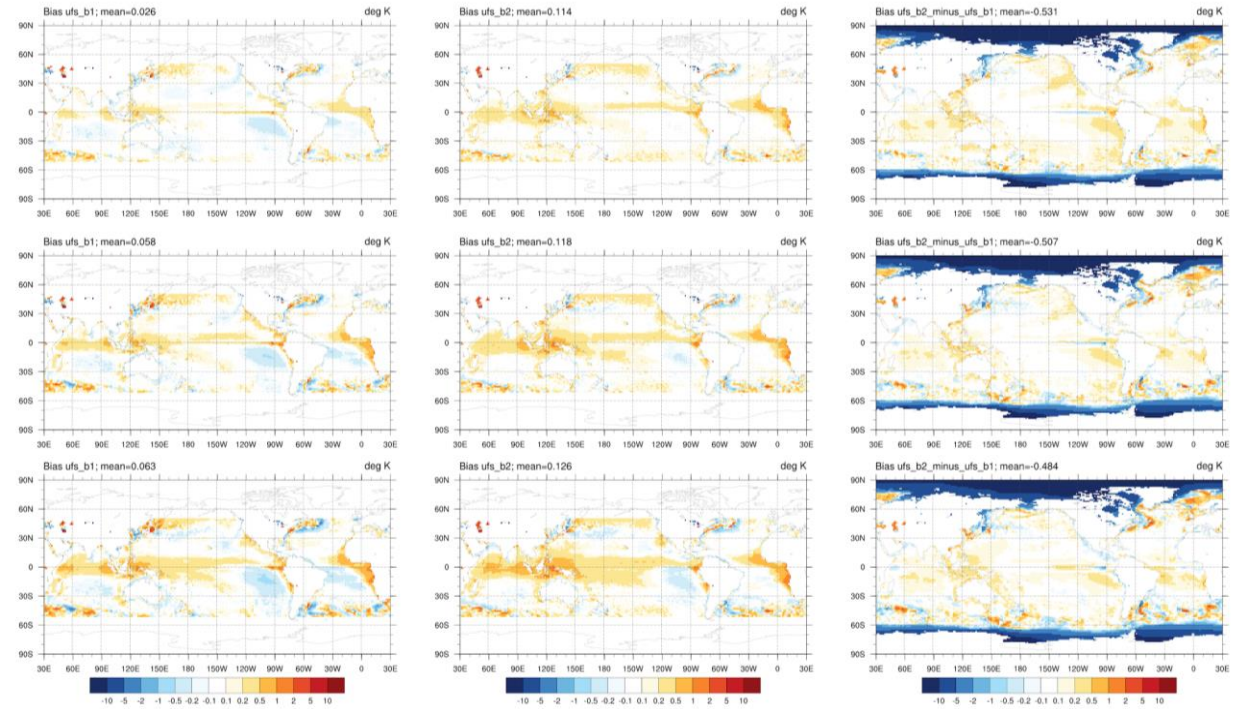

Figure 6.3: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean SST bias wrt OSTIA from 50°S to 50°N in P1 (right column) and P2 (middle column); Difference (P2 minus P1) in surface temperatures over ocean grid points, including sea ice (left column).

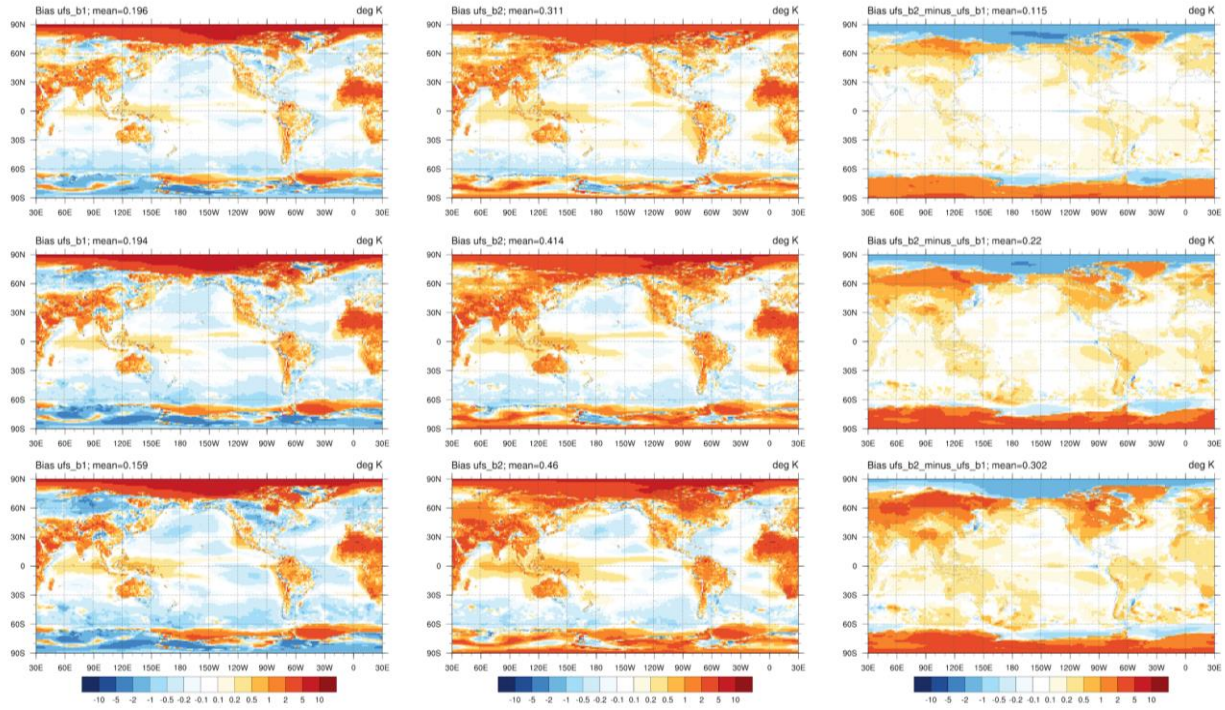

Figure 6.4: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in P1 (left column) and P2 (middle column); Difference (P2 minus P1) in 2-m air temperature (right column).

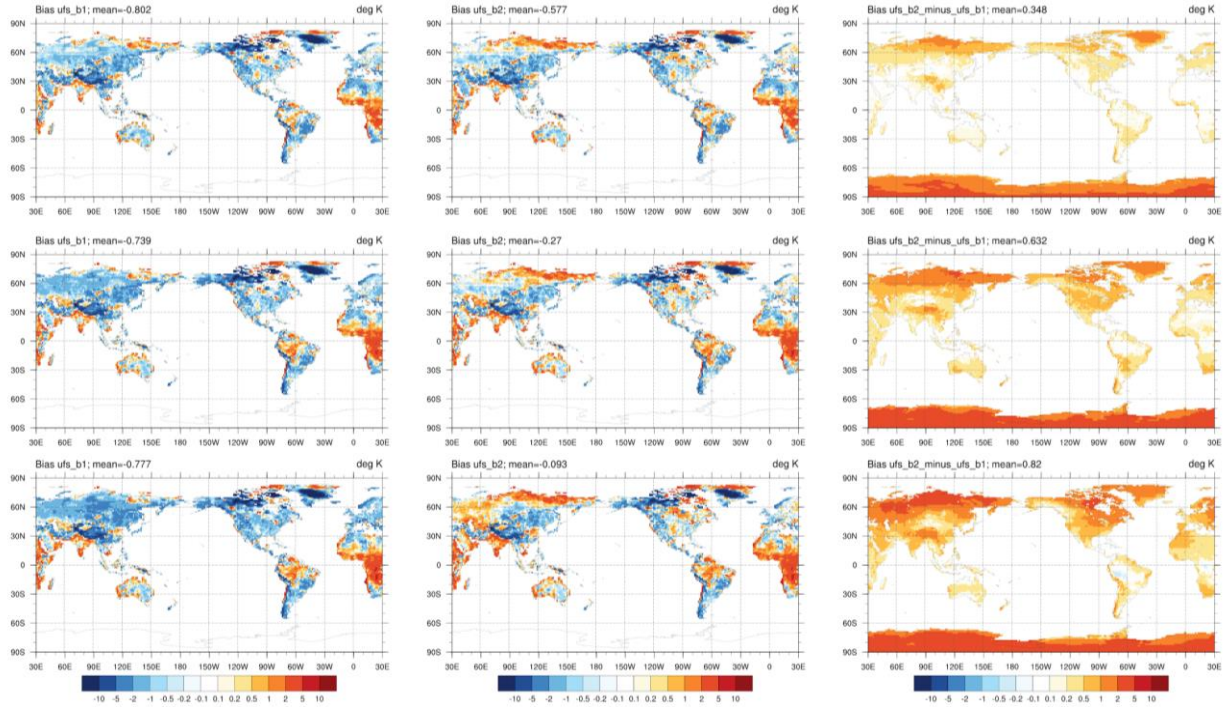

Figure 6.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in P1 (left column) and P2 (middle column); Difference (P2 minus P1) in 2-m daily maximum air temperature (right column).

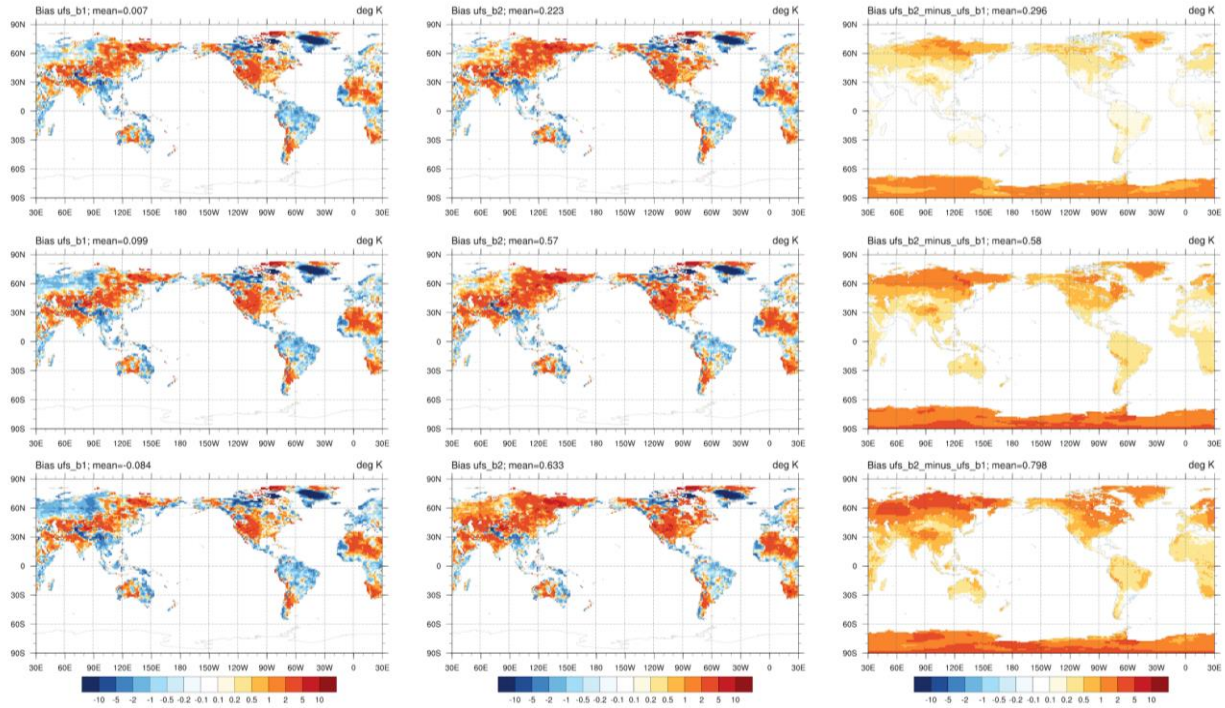

Figure 6.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in P1 (left column) and P2 (middle column); Difference (P2 minus P1) in 2-m daily minimum air temperature (right column).

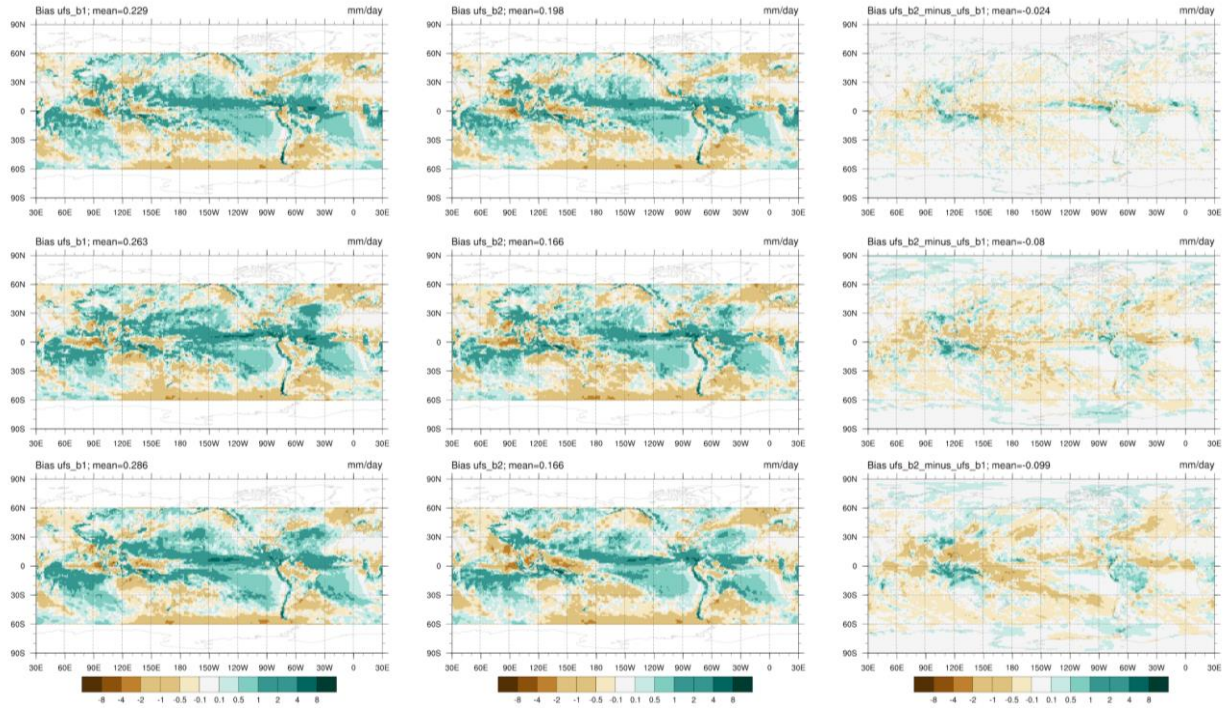

Figure 6.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in P1 (left column) and P2 (middle column); Difference (P2 minus P1) in precipitation rate (right column).

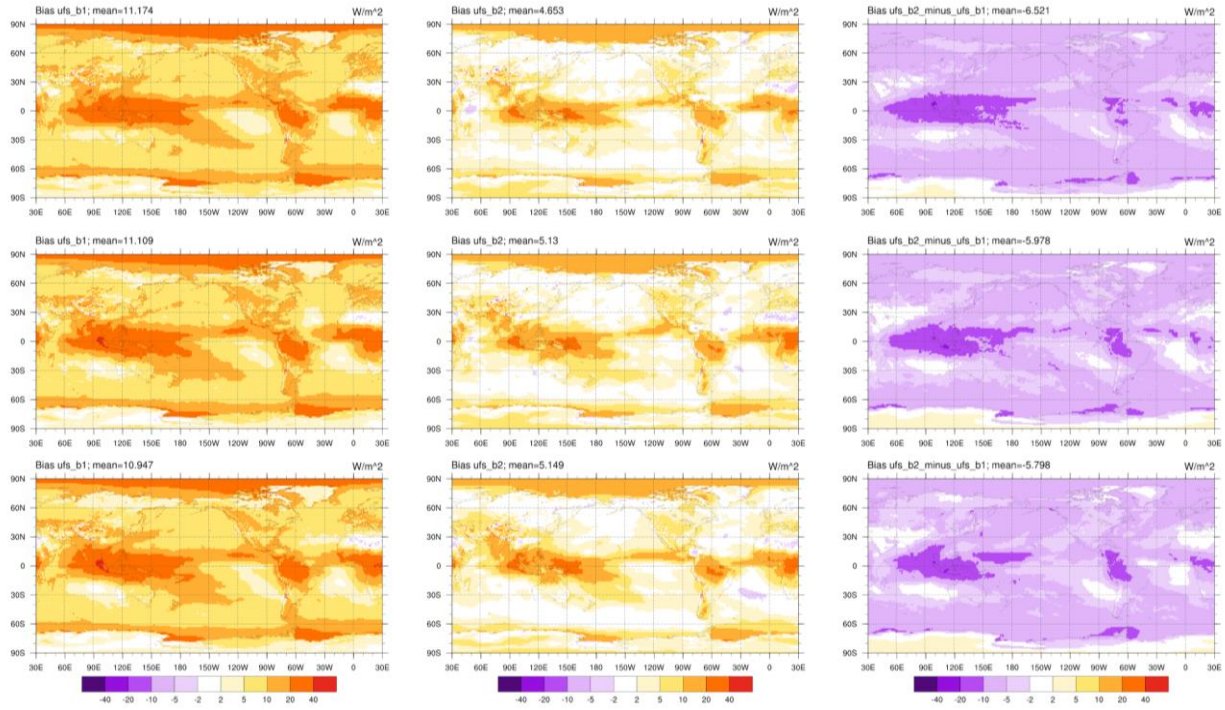

Figure 6.8: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in P1 (left column) and P2 (middle column); Difference (P2 minus P1) in OLR (right column).

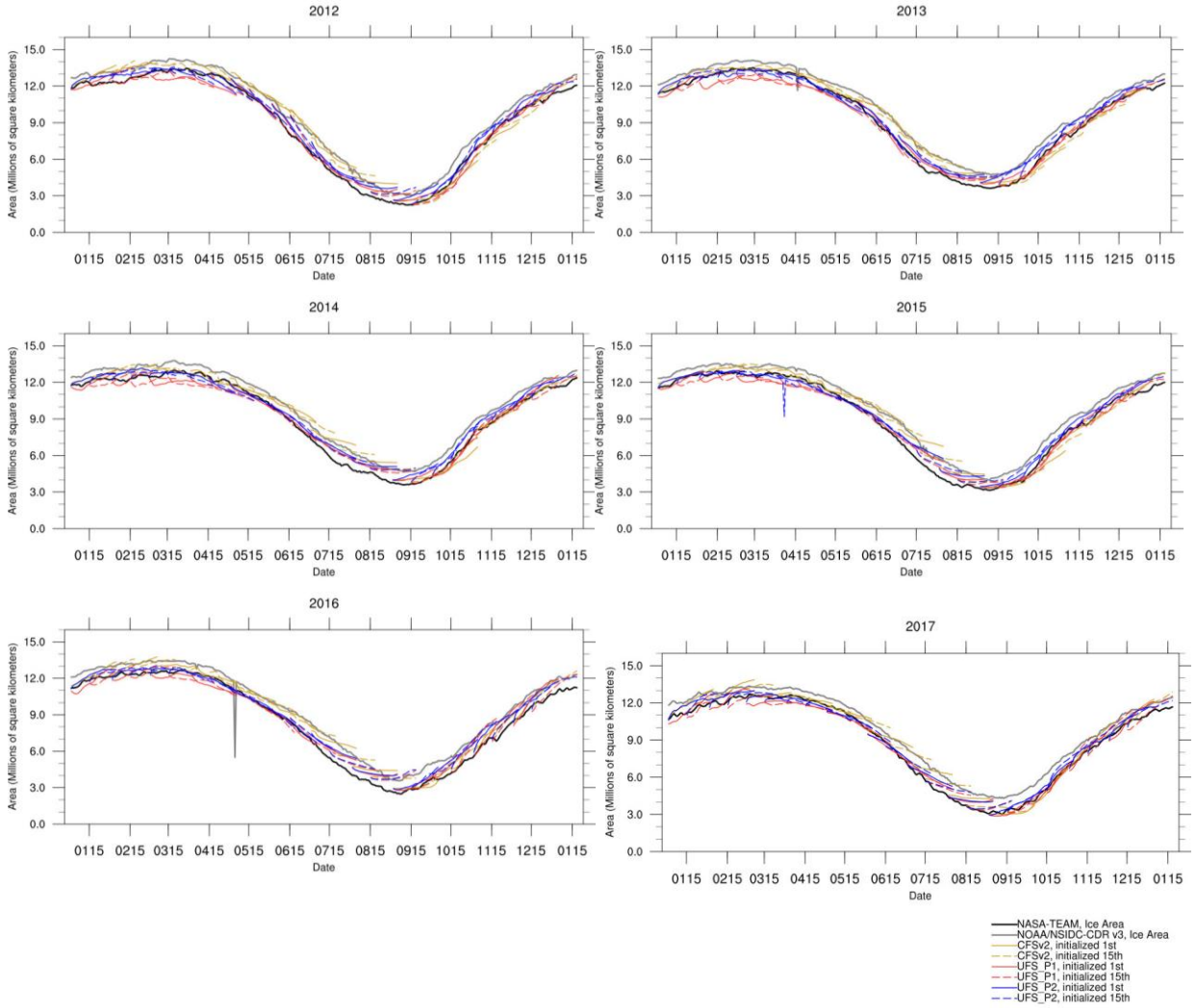

Figure 6.9: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow), P1 (red), and P2 (blue). Thick lines correspond to the observational estimate derived from ice concentrations from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date.

P3.1 Figures

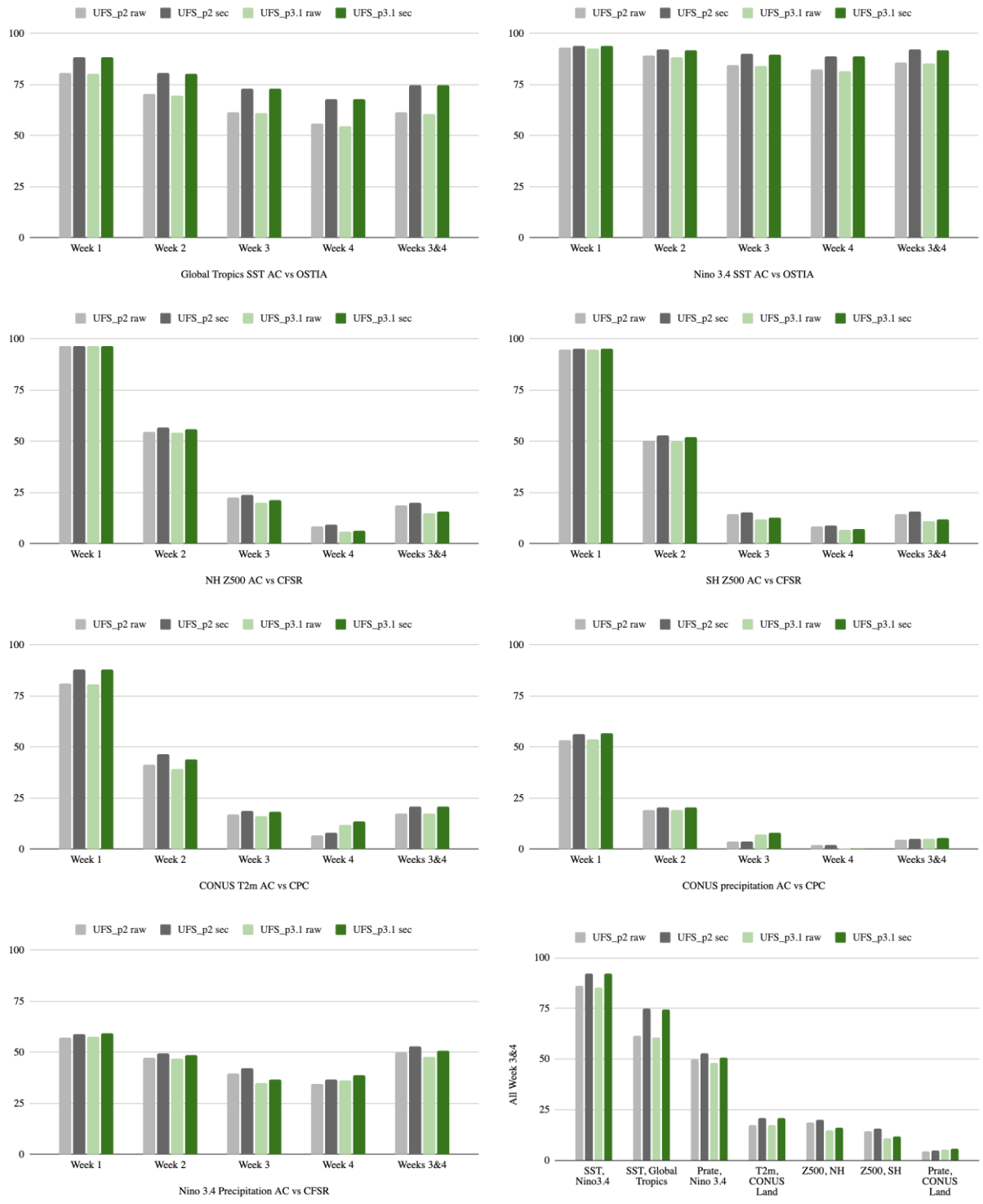

Figure 7.1: Anomaly correlation scores by forecast week of P2 (RAW-pale gray, SEC-dark gray) and P3.1 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of the anomaly correlation scores for Weeks 3&4.

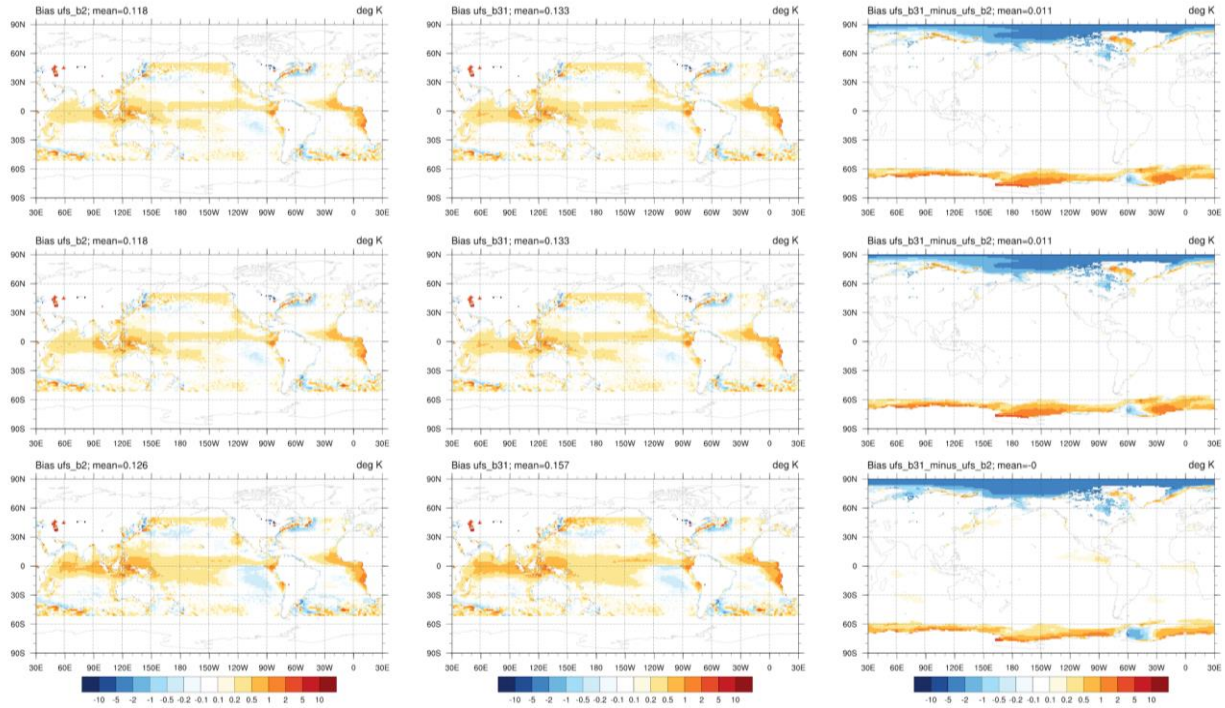

Figure 7.2: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean SST bias wrt OSTIA from 50°S to 50°N in P2 (left column) and P3.1 (middle column); Difference (P3.1 minus P2) in surface temperatures over ocean grid points, including sea ice (right column).

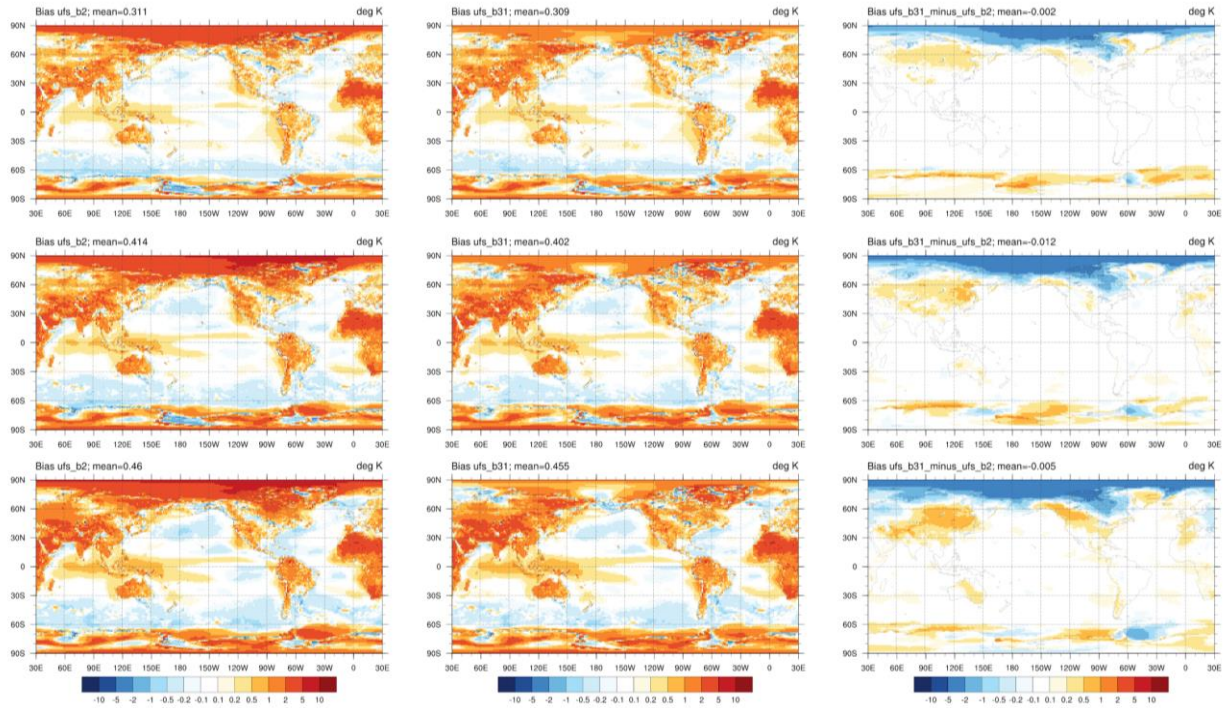

Figure 7.3: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in P2 (left column) and P3.1 (middle column); Difference (P3.1 minus P2) in 2-m air temperature (right column).

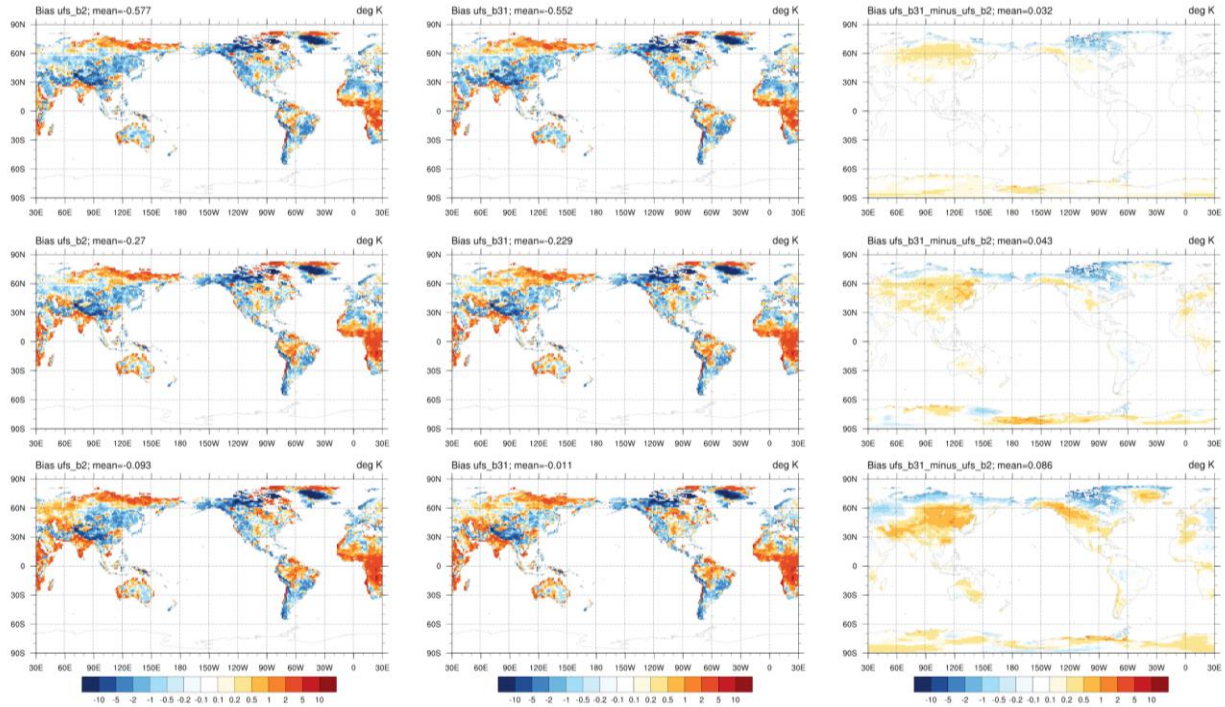

Figure 7.4: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in P2 (left column) and P3.1 (middle column); Difference (P3.1 minus P2) in 2-m daily maximum air temperature (right column).

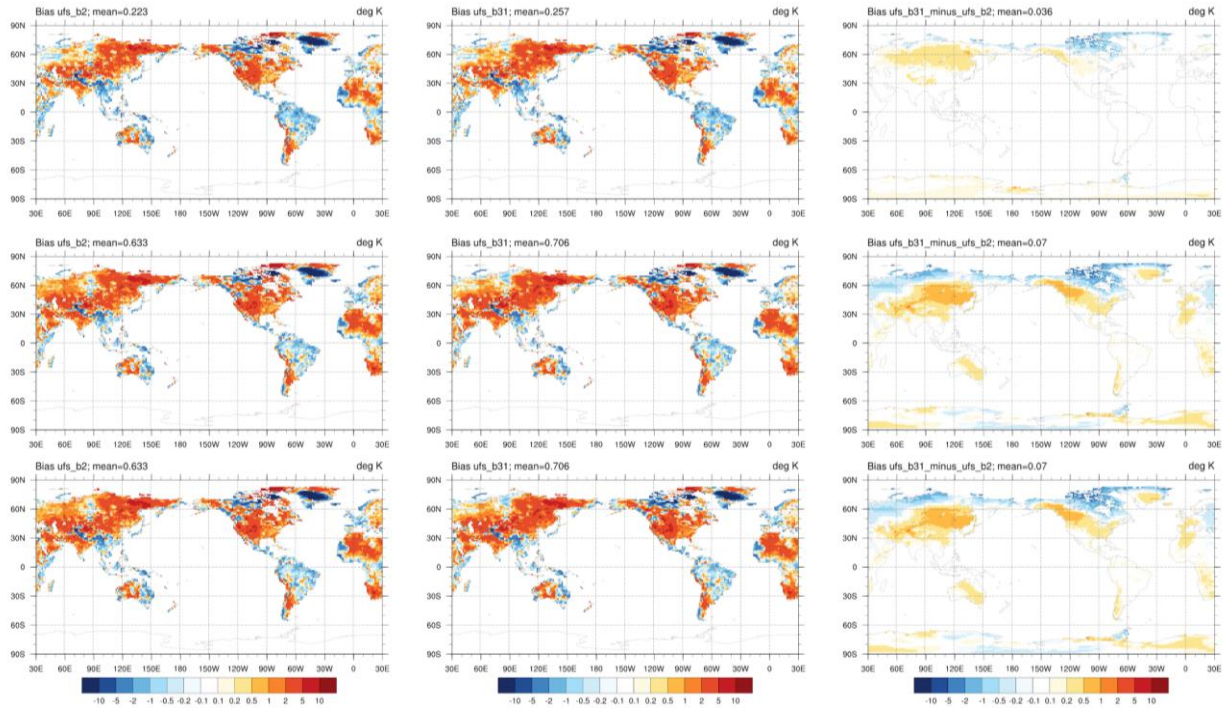

Figure 7.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in P2 (left column) and P3.1 (middle column); Difference (P3.1 minus P2) in 2-m daily minimum air temperature (right column).

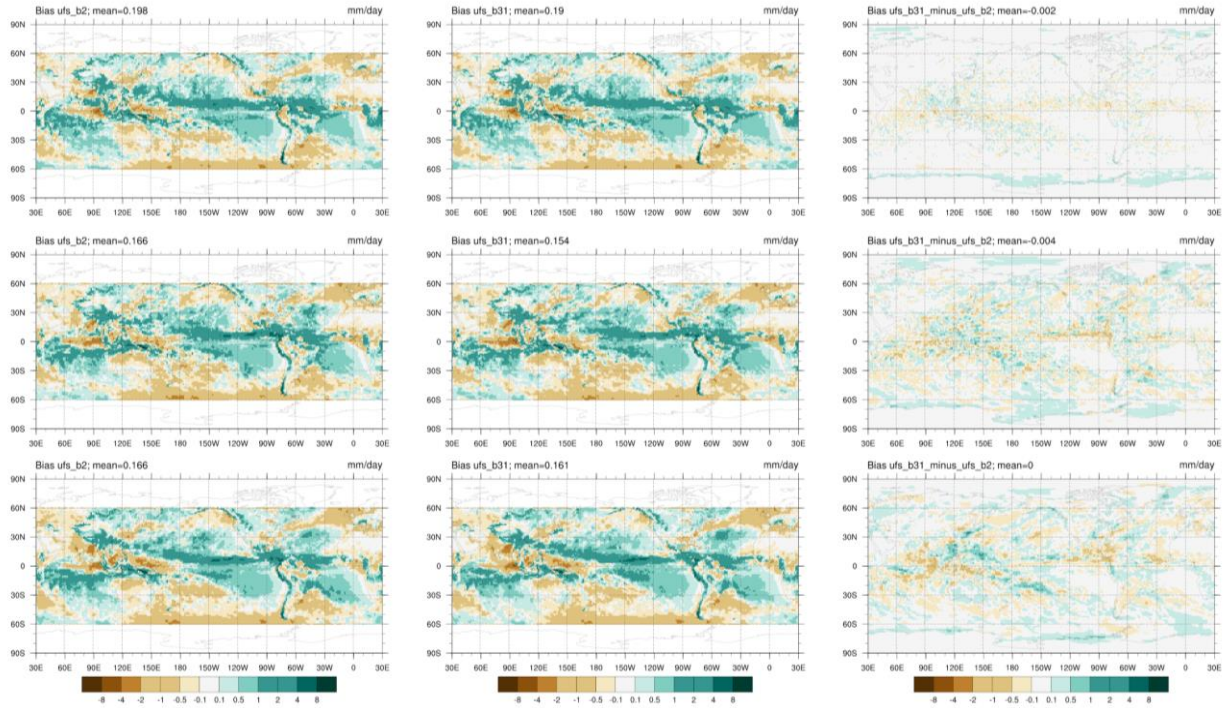

Figure 7.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in P2 (left column) and P3.1 (middle column); Difference (P3.1 minus P2) in precipitation rate (right column).

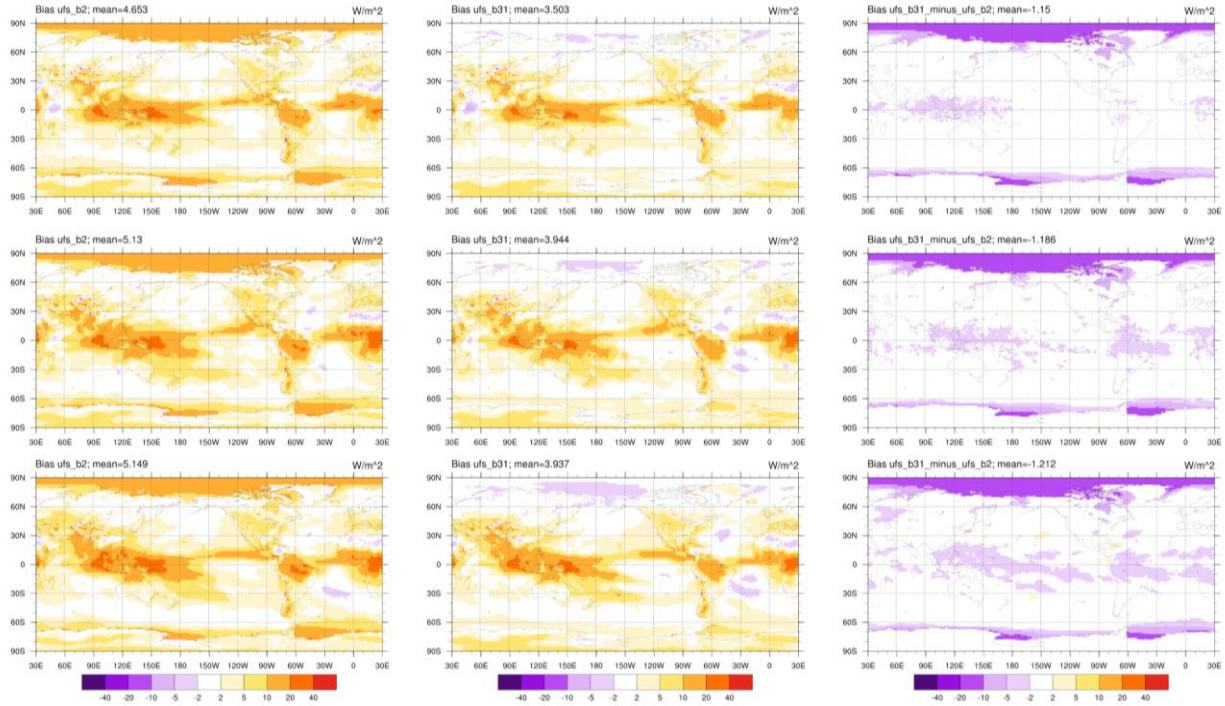

Figure 7.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in P2 (left column) and P3.1 (middle column); Difference (P3.1 minus P2) in OLR (right column).

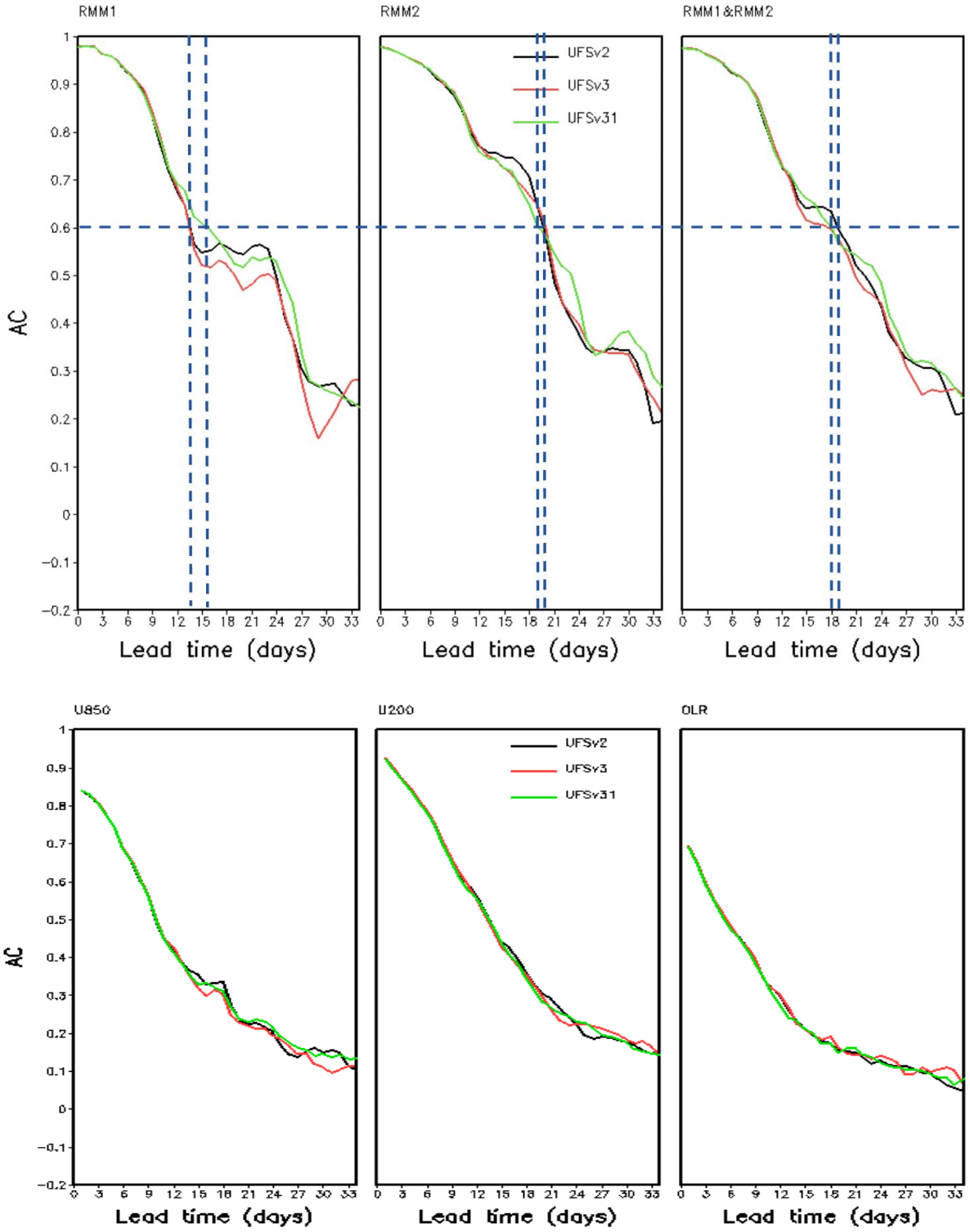

Figure 7.8: Top: AC of MJO RMM1 (left), RMM2 (middle) and combined RMM1&RMM2 (right). Bottom: AC of equatorial filtered zonal wind at 850mb (left) and 200mb (middle), and equatorial filtered OLR (right) in P2 (black line), P3 (red line; note that this intermediate prototype is beside the scope of the present paper) and P3.1 (green line).

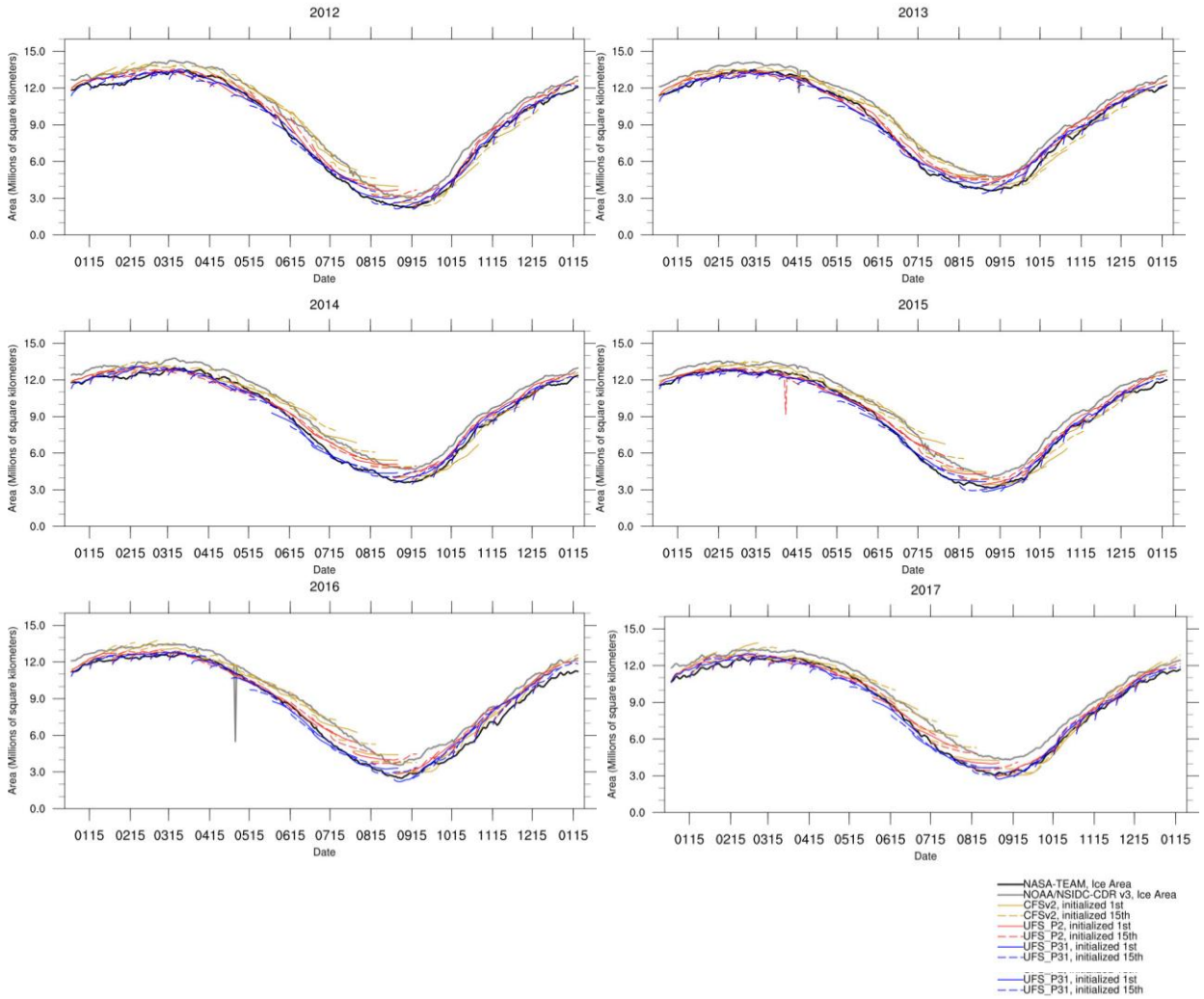

Figure 7.9a: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow), P2 (red), and P3.1 (blue). Thick lines correspond to the observational estimate derived from ice concentrations from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date.

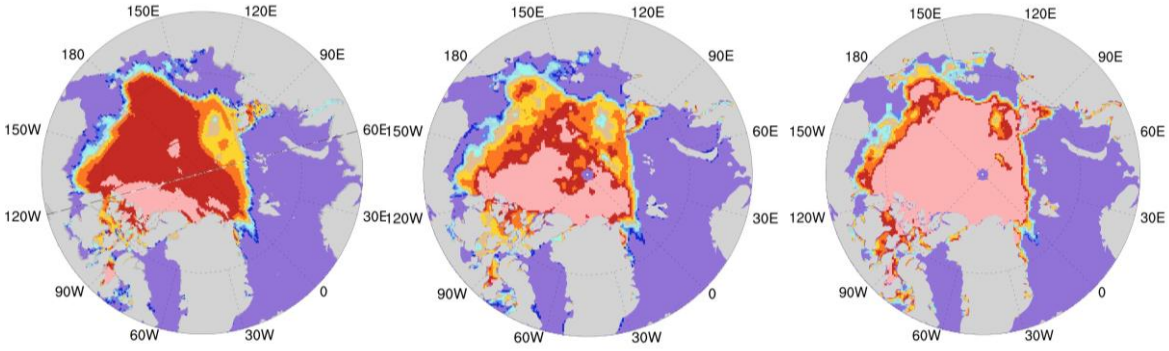

Figure 7.9b: Example of model sea ice concentration on day 1 of integration (left), and the sea ice concentration for the same date from NASA-Team (middle) and NOAA-CDR (right)

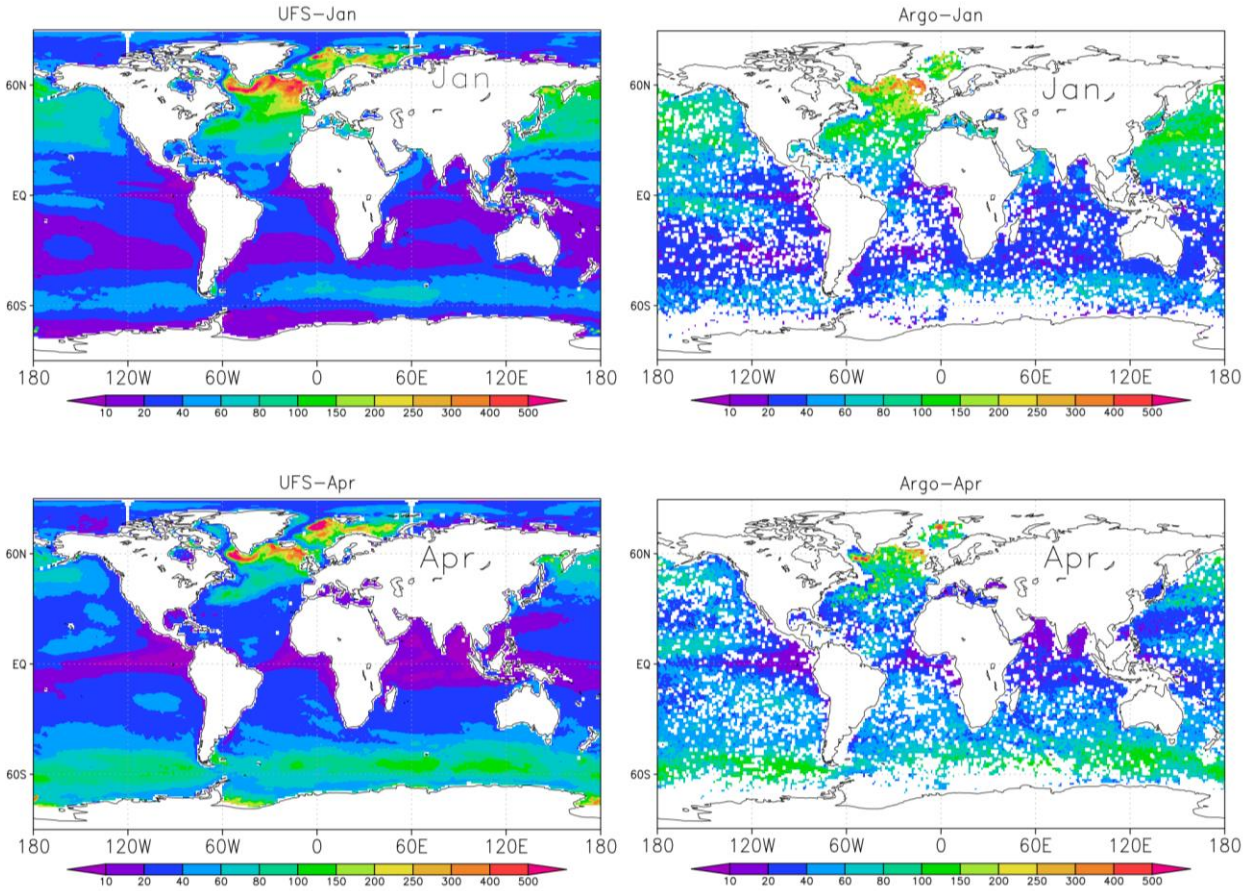

Figure 7.10: January and April mean mixed layer depth (left panel: P3.1, right panel: Argo observational data).

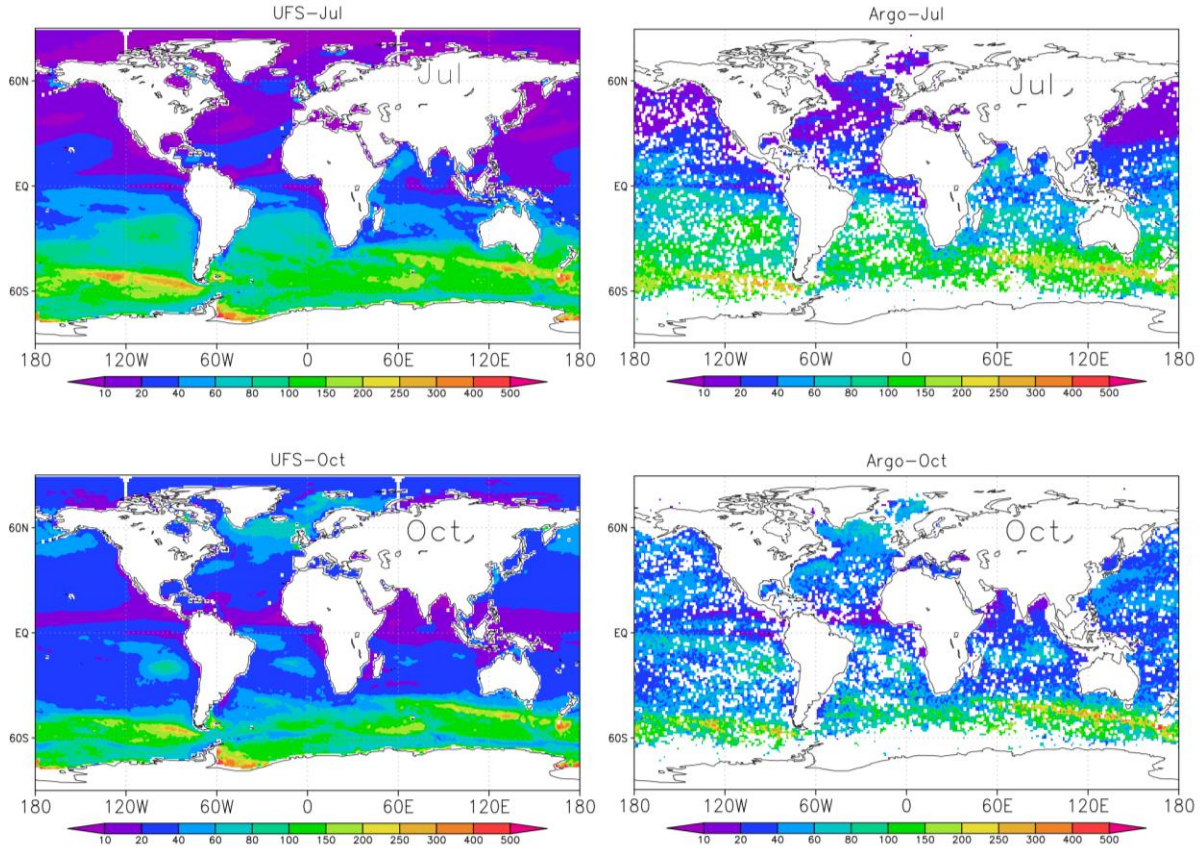

Figure 7.11: Same as Fig. 7.10, but for July and October.

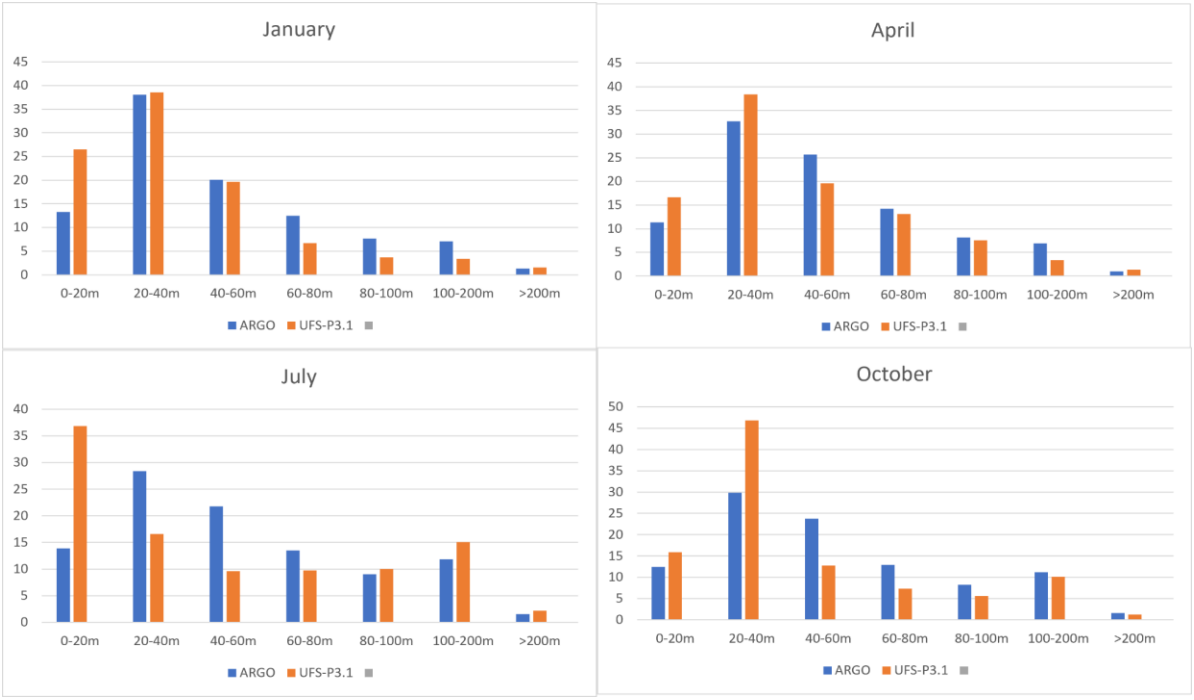

Figure 7.12: Probability distribution function of MLD for January, April, July and October between P3.1 and Argo observational data. The vertical axis shows the percentage of MLD values falling in a given binned range.

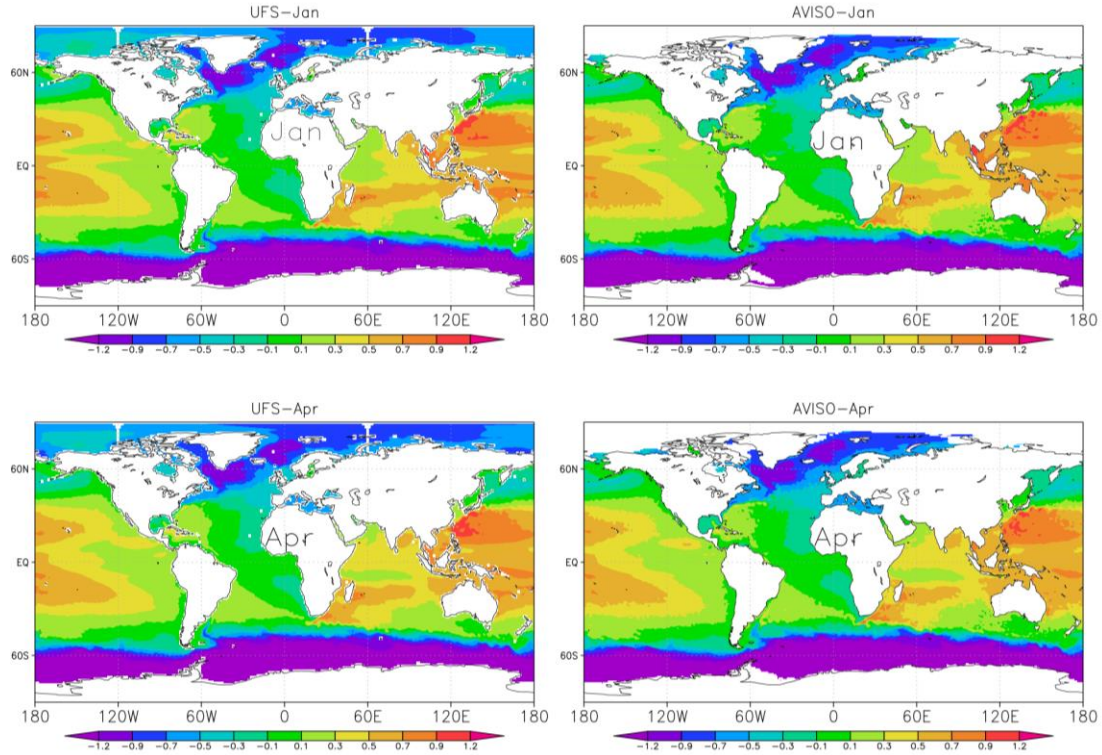

Figure 7.13: January and April mean sea surface height (left panel: P3.1, right panel: Aviso). Unit: meter. Note the global mean values are removed due to the difference in the geoid definition between MOM6 and Aviso.

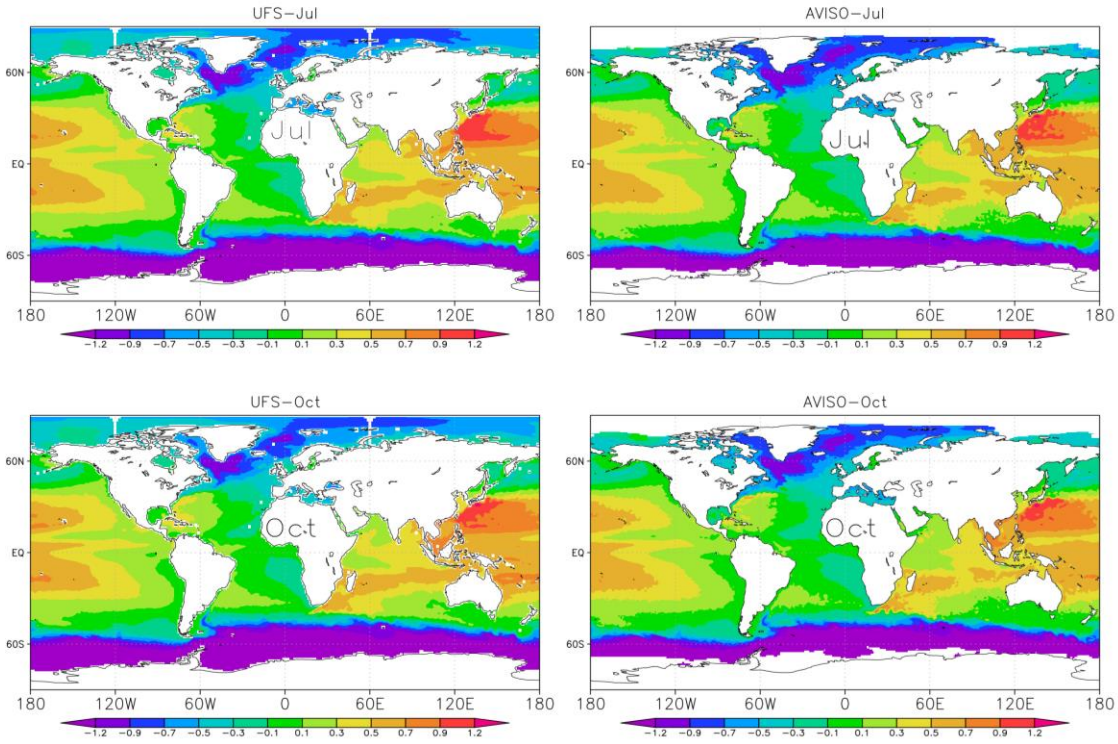

Figure 7.14: Same as Fig. 7.13, but for July and October.

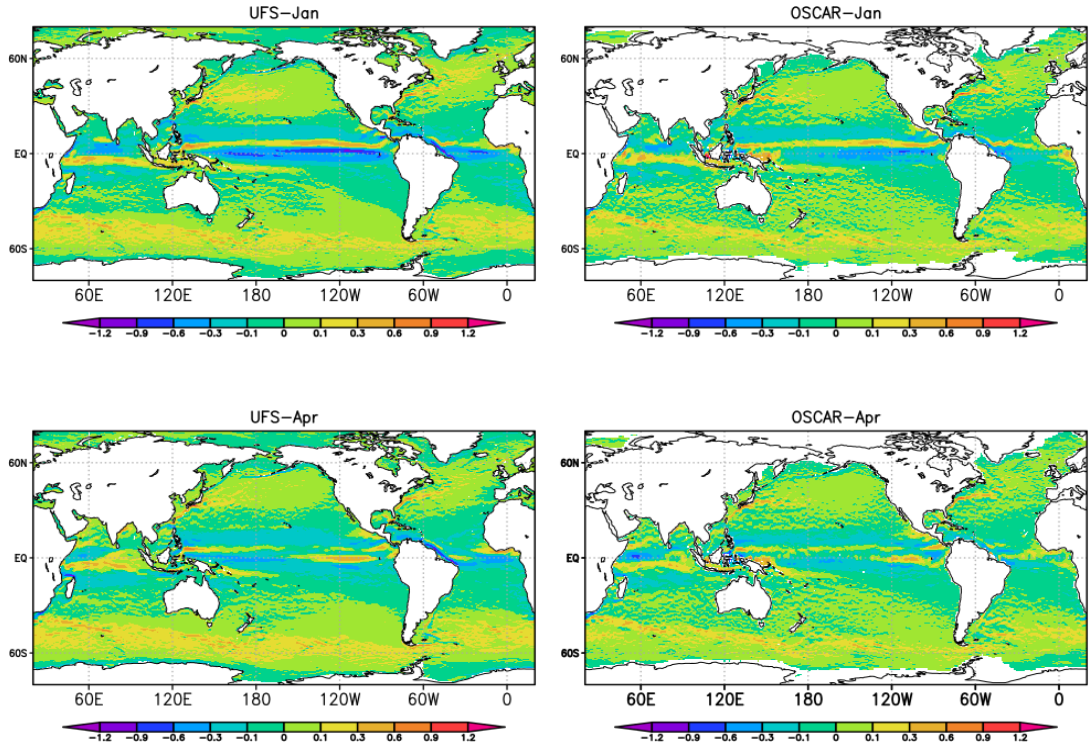

Figure 7.15: January and April mean ocean surface zonal currents. Left panel: P3.1, right panel: OSCAR. Unit is m/s.

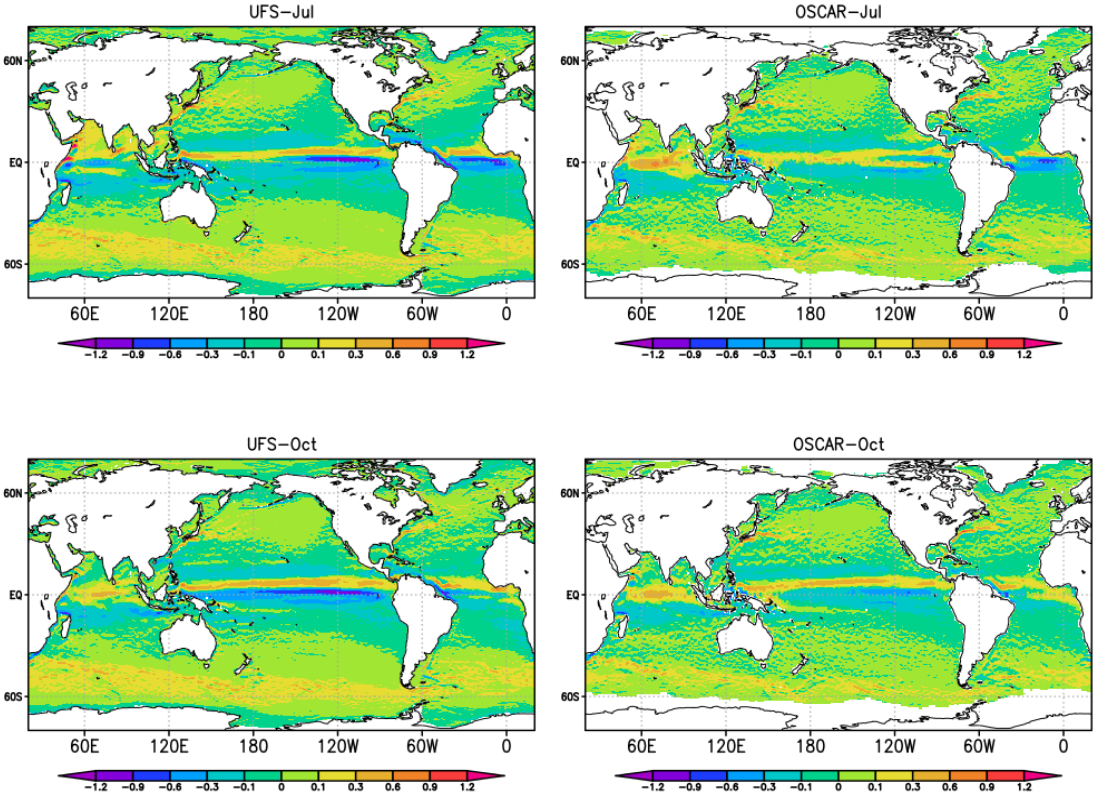

Figure 7.16: Same as Fig. 7.15, but for July and October.

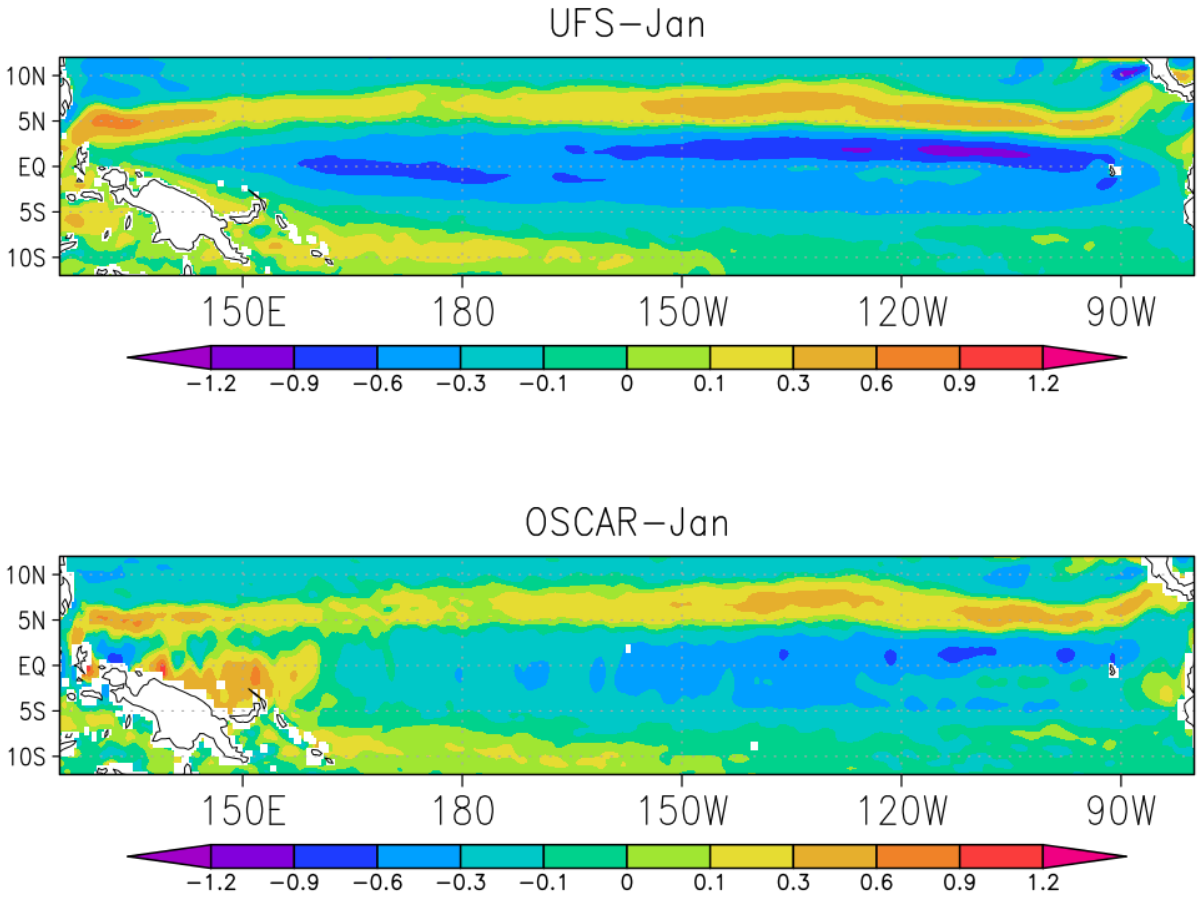

Figure 7.17: Same as Fig. 7.15, but zoomed in over the equatorial Pacific to highlight the stratification of North/South Equatorial Currents and Equatorial Countercurrent.

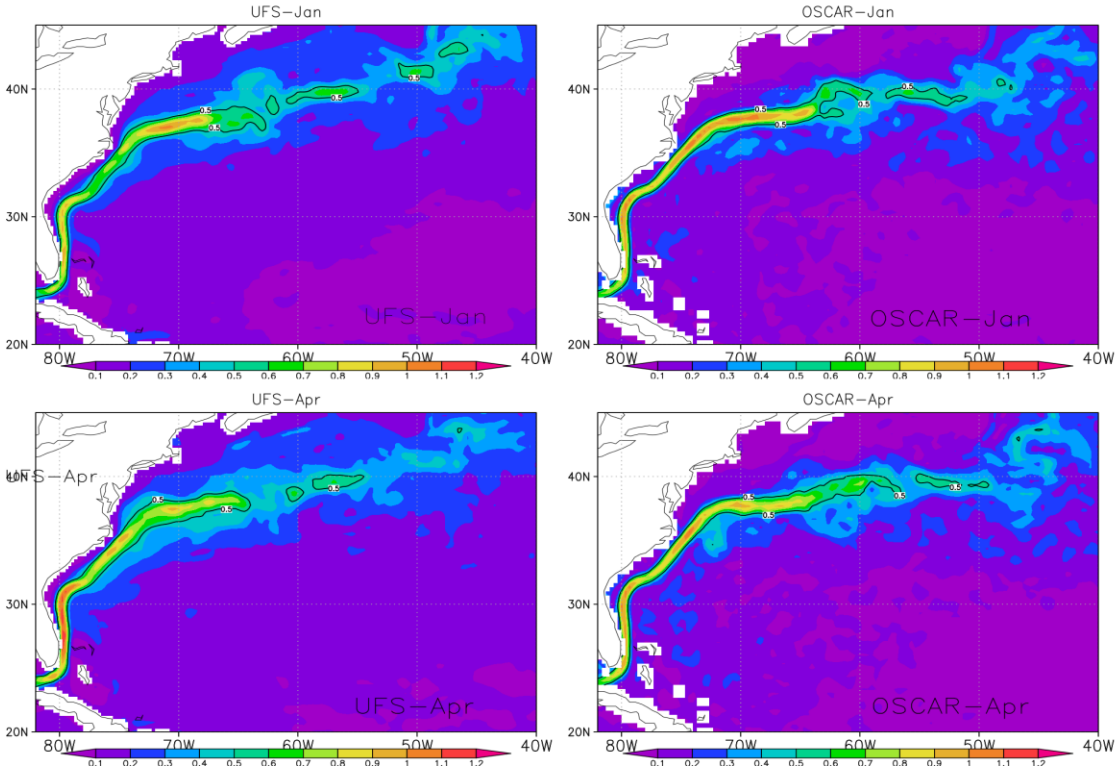

Figure 7.18: January and April mean ocean currents magnitude in Gulf stream region. Left panel: P3.1, right panel: OSCAR. Unit is m/s.

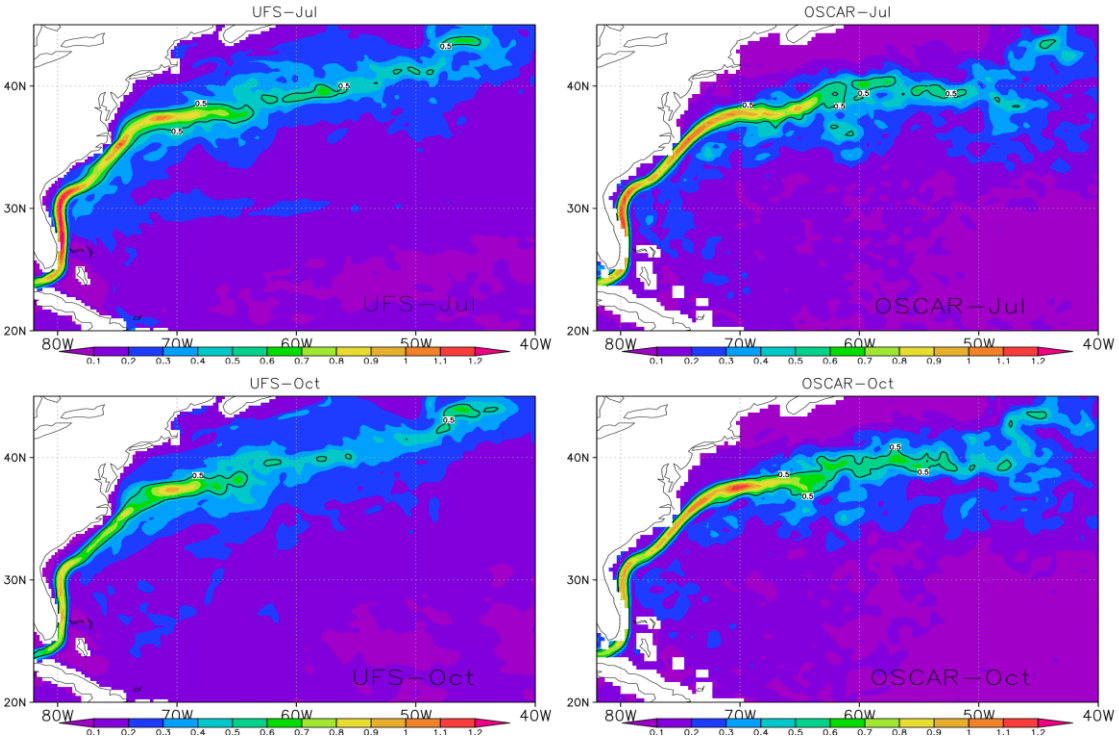

Figure 7.19: Same as Figure 7.18, but for July and October.

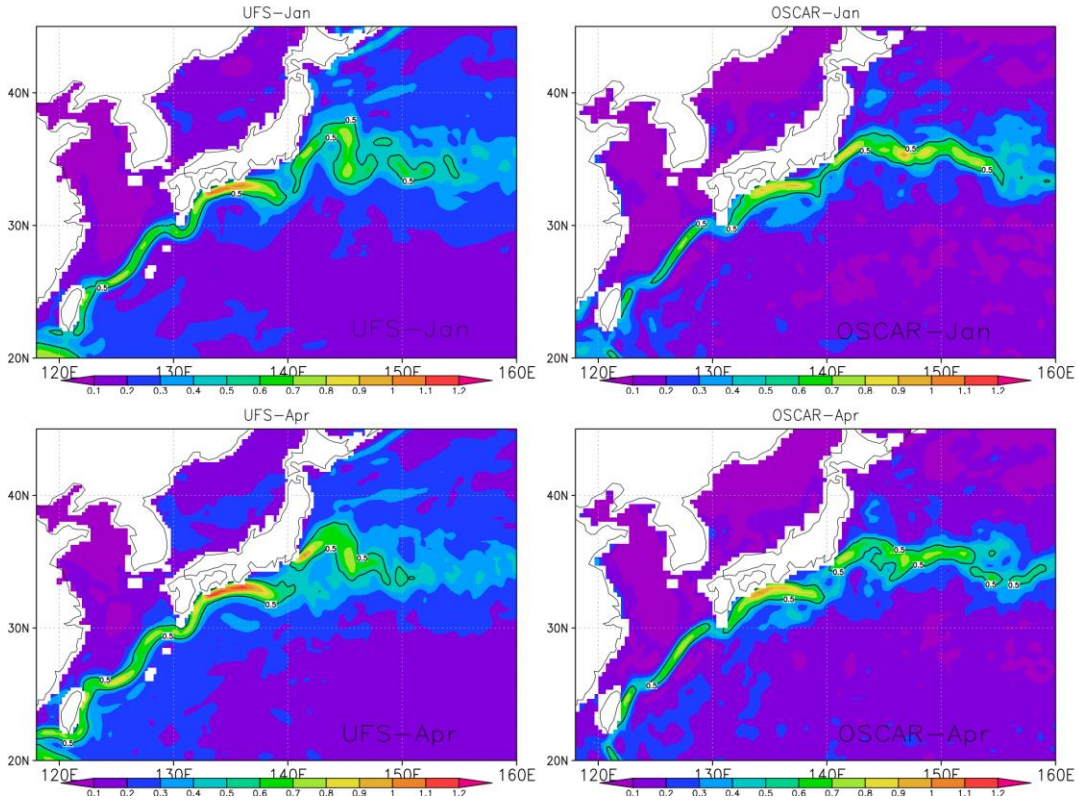

Figure 7.20: January and April mean ocean currents magnitude in Kuroshio region. Left panel: P3.1, right panel: OSCAR. Unit is m/s.

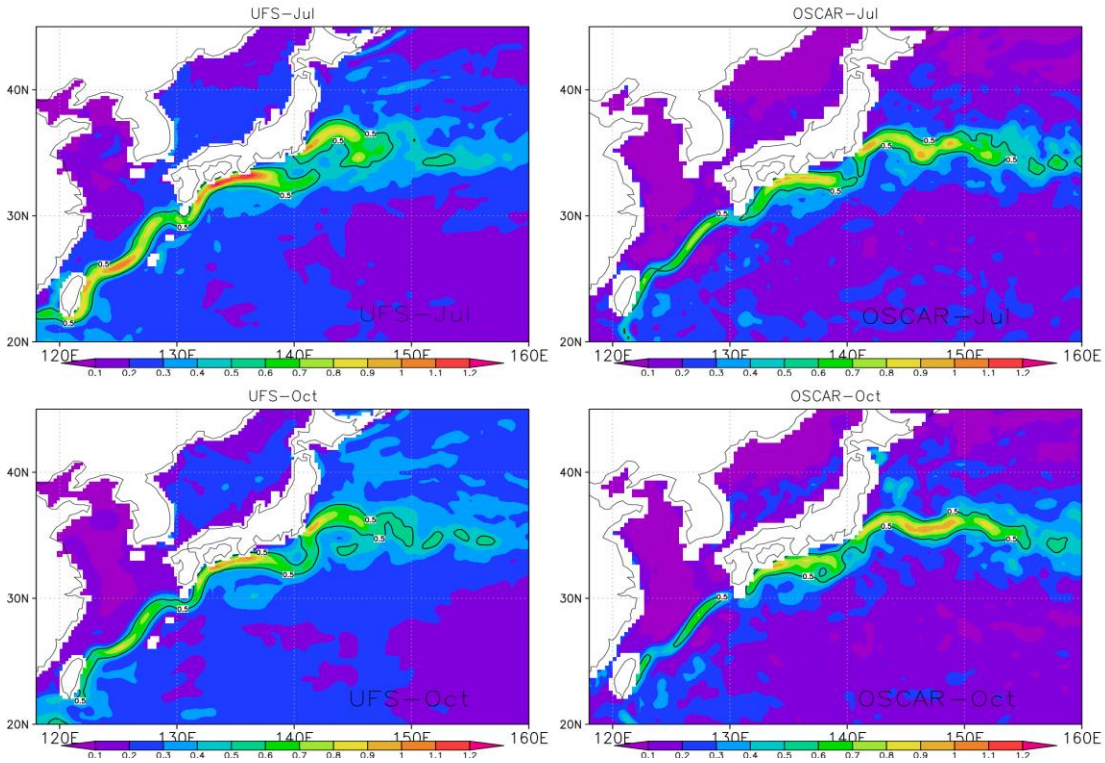

Figure 7.21: Same as Fig. 7.20, but for July and October.

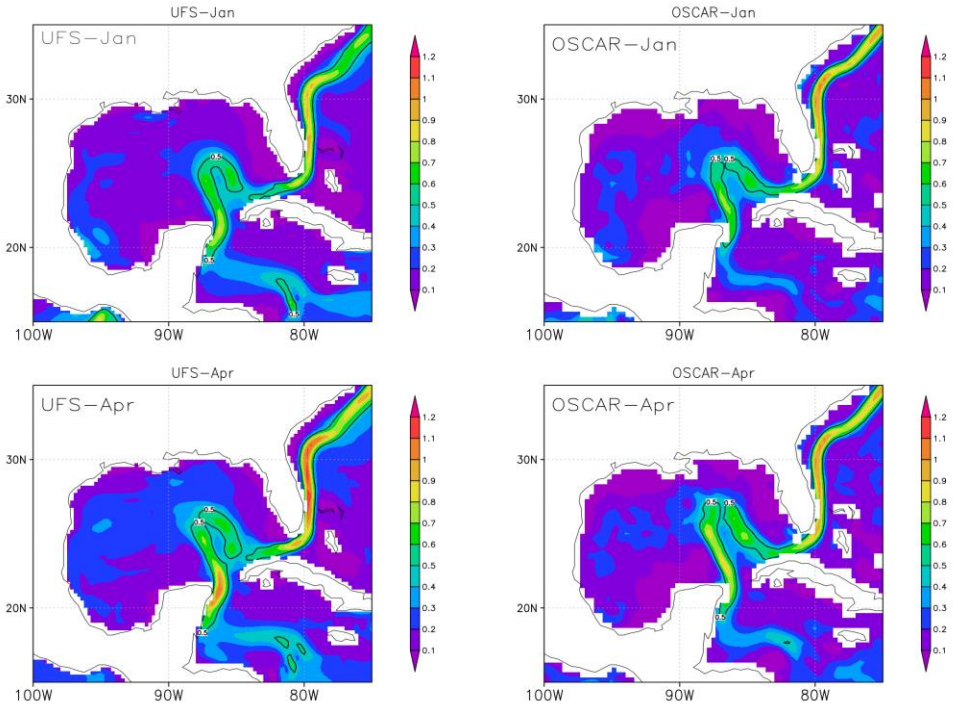

Figure 7.22: January and April mean ocean currents magnitude in Loop current region. Left panel: P3.1, right panel: OSCAR. Unit is m/s.

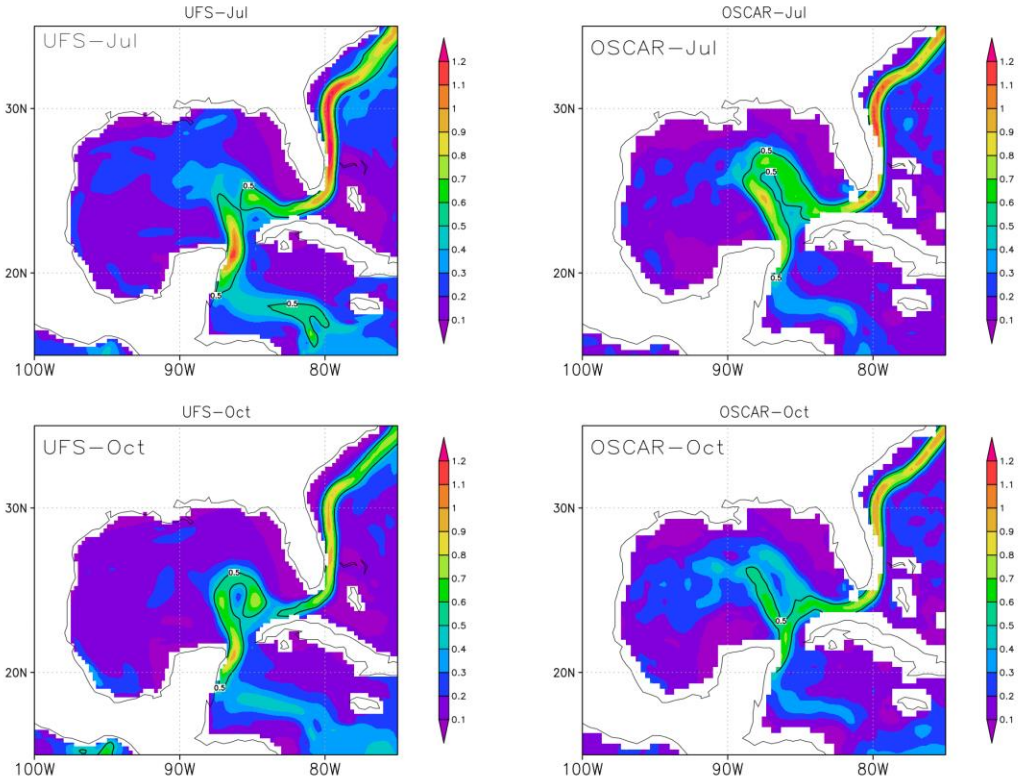

Figure 7.23: Same as Fig. 7.22, but for July and October.

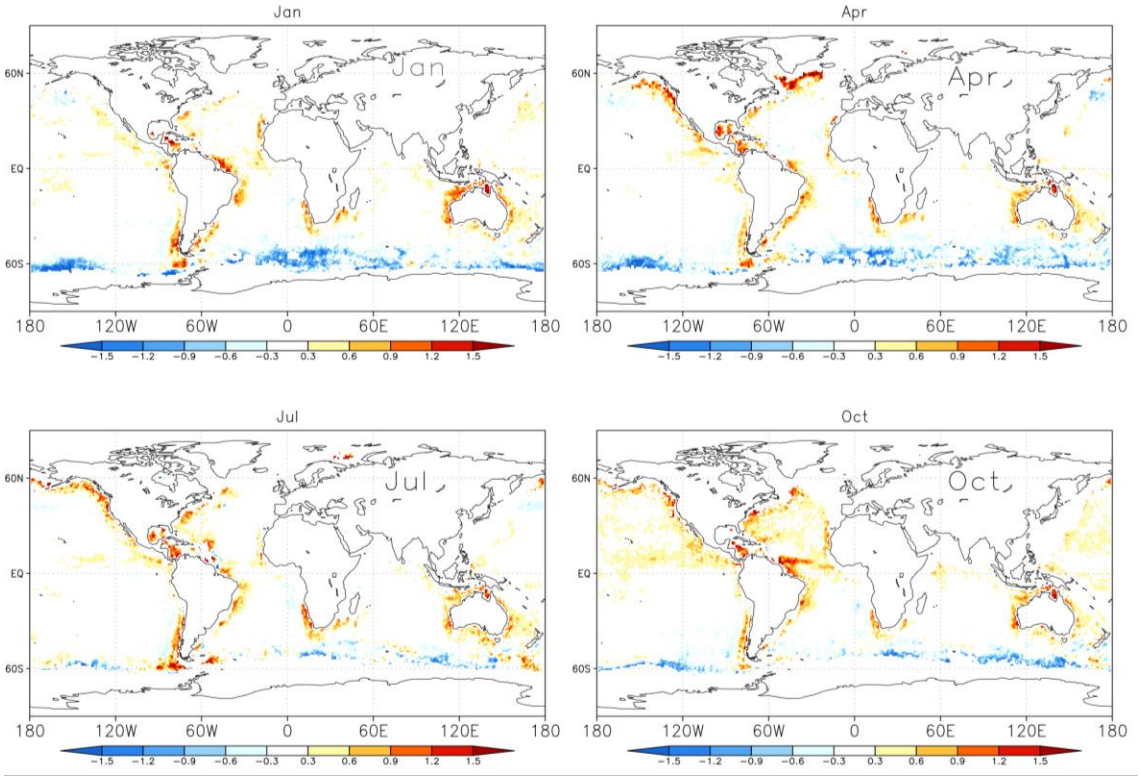

Figure 7.24: Differences of Sea Surface Salinity (SSS) between P3.1 and SMOS for January, April, July and October means. Unit is psu.

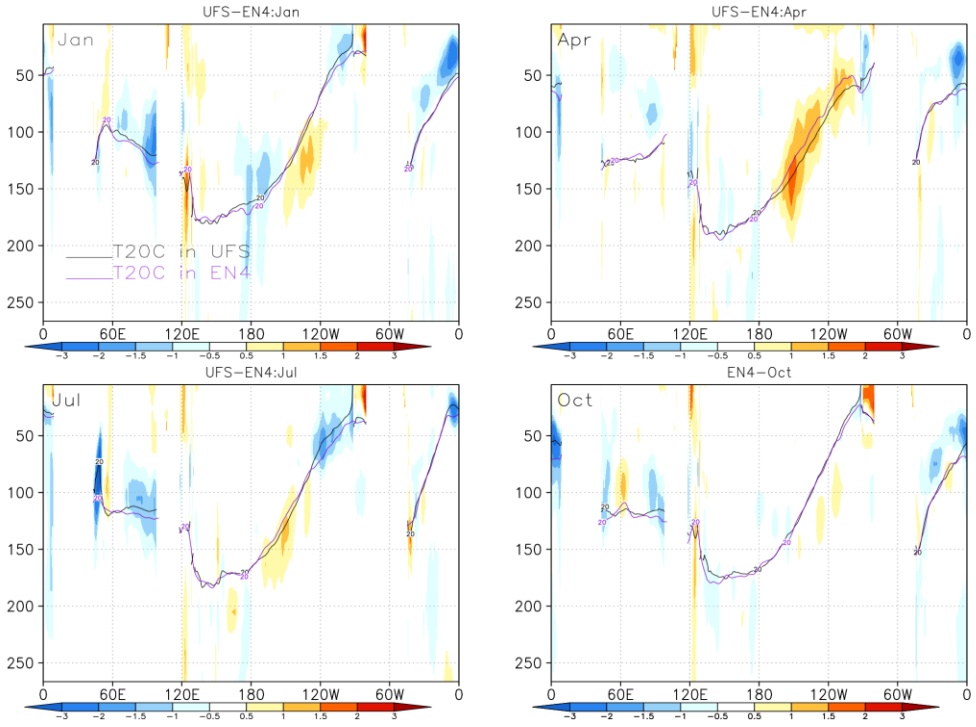

Figure 7.25: Vertical cross section of temperature at equator for January, April, July and October means. Black/purple lines are 20°C isothermal depths in UFS and EN4. Yellow and blue shaded areas are the temperature difference between P3.1 and EN4 from surface to 300 meters depth. The vertical axis shows depth in meters.

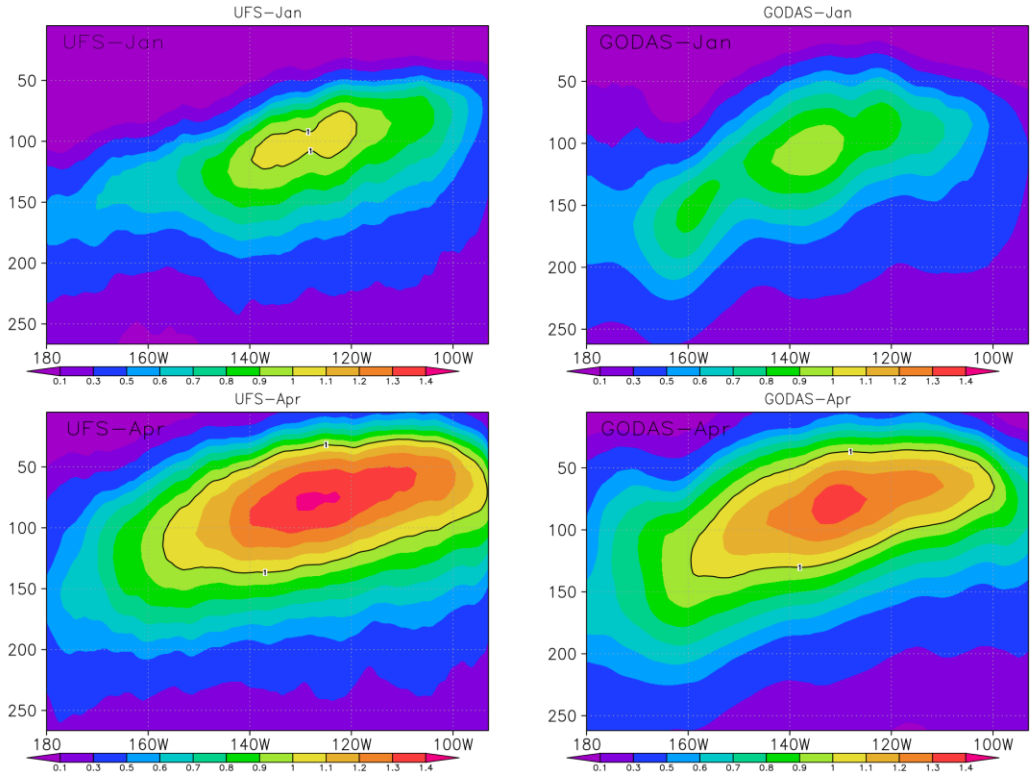

Figure 7.26: Vertical cross section of ocean zonal currents at the equator in the Eastern Pacific for January and April means. The vertical axis shows depth in meters. Left panel: P3.1, right panel: GODAS.

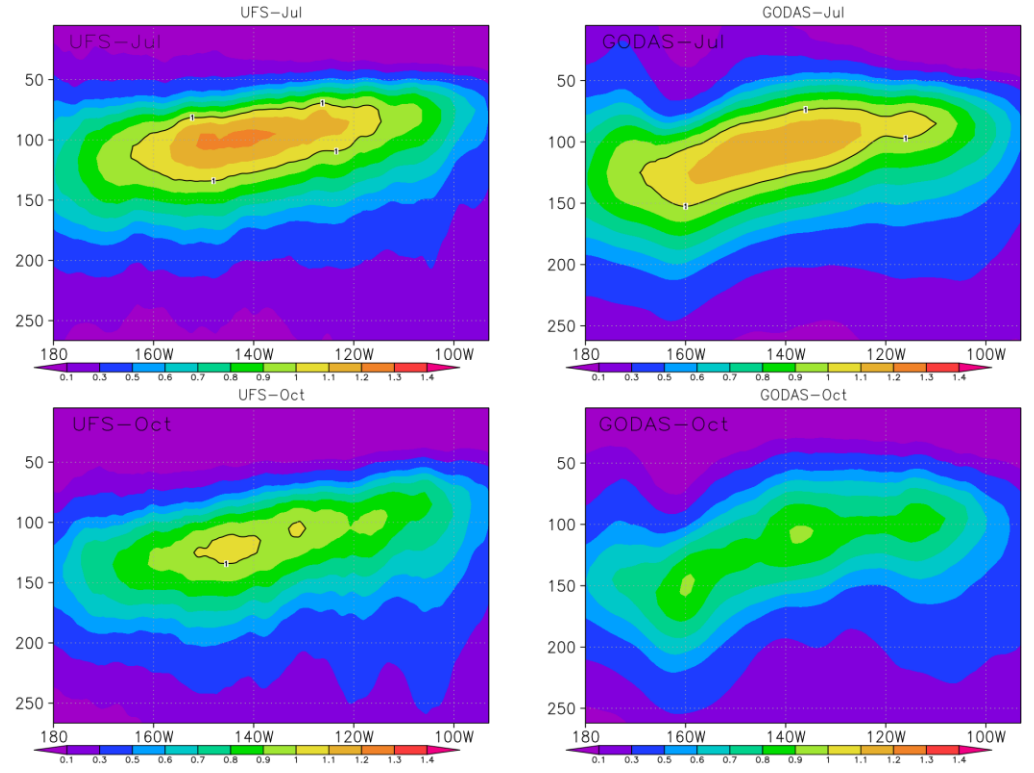

Figure 7.27: Same as Fig. 7.26, but for July and October means.

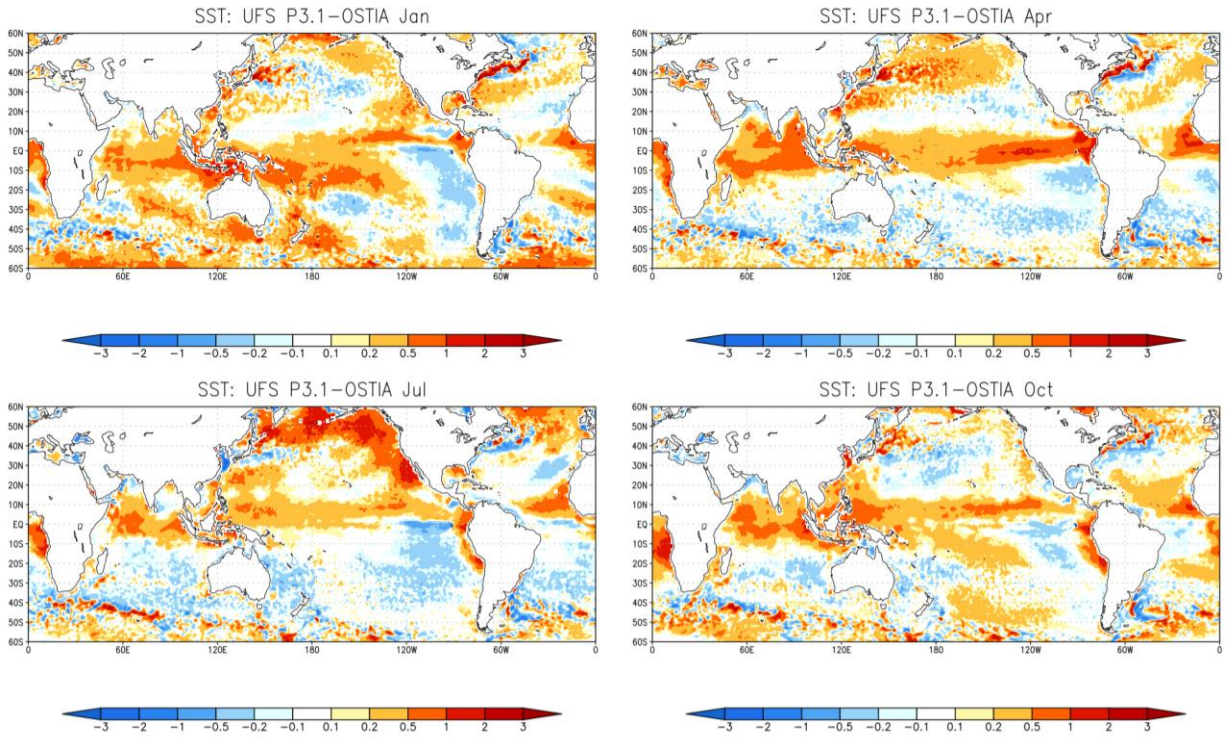

Figure 7.28: January, April, July and October mean SST differences between P3.1 and OSTIA.

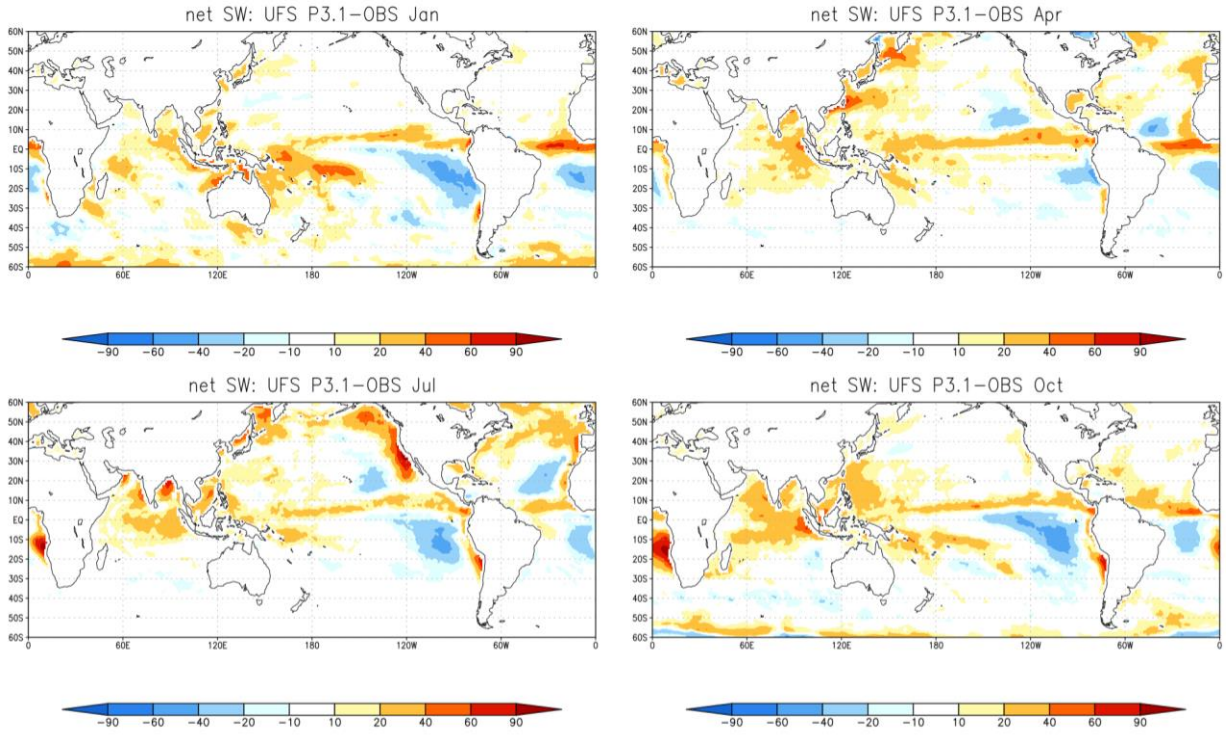

Figure 7.29: January, April, July and October mean net shortwave radiation differences between P3.1 and CERS EBAF.

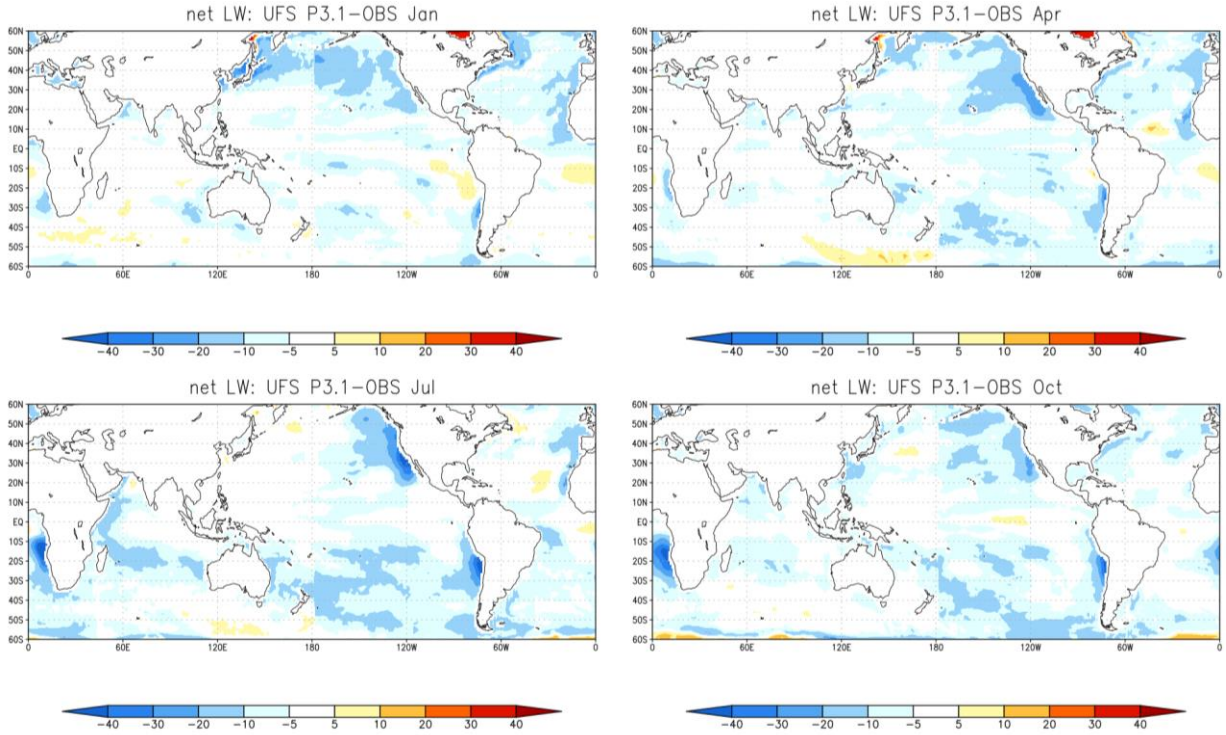

Figure 7.30: January, April, July and October mean net longwave radiation differences between P3.1 and CERES EBAF.

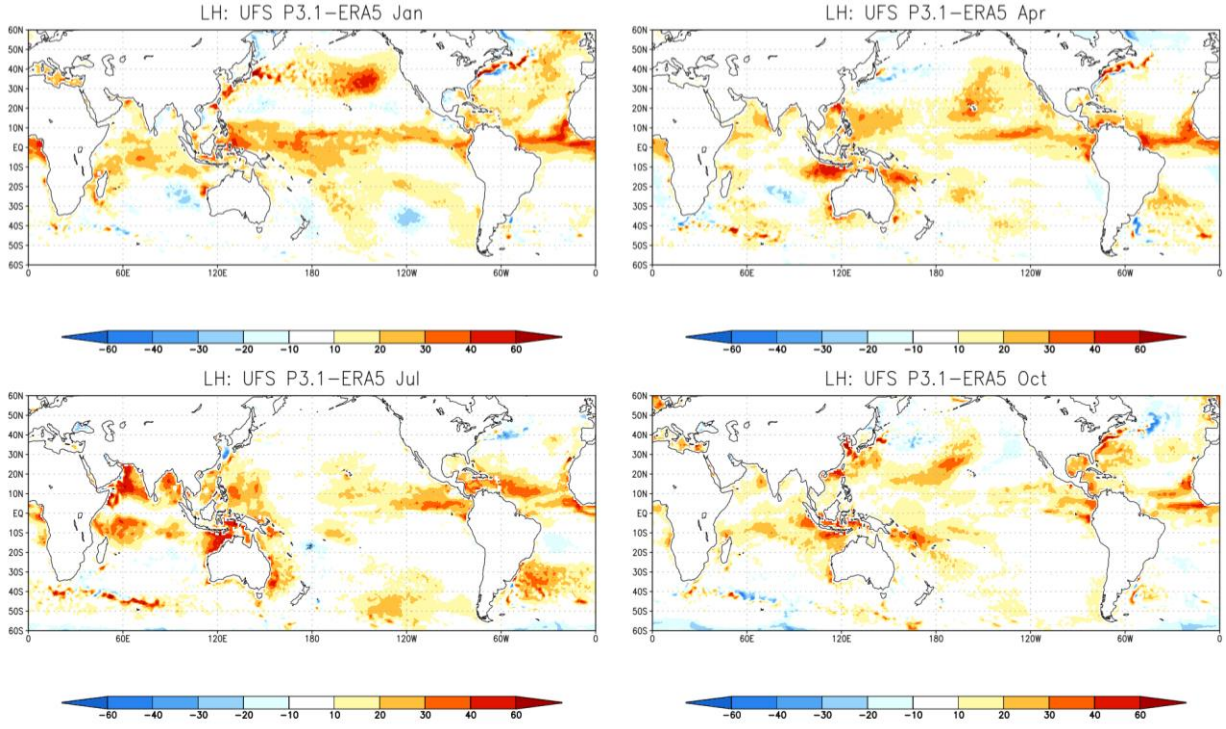

Figure 7.31: January, April, July and October mean latent heat differences between P3.1 and ERA5.

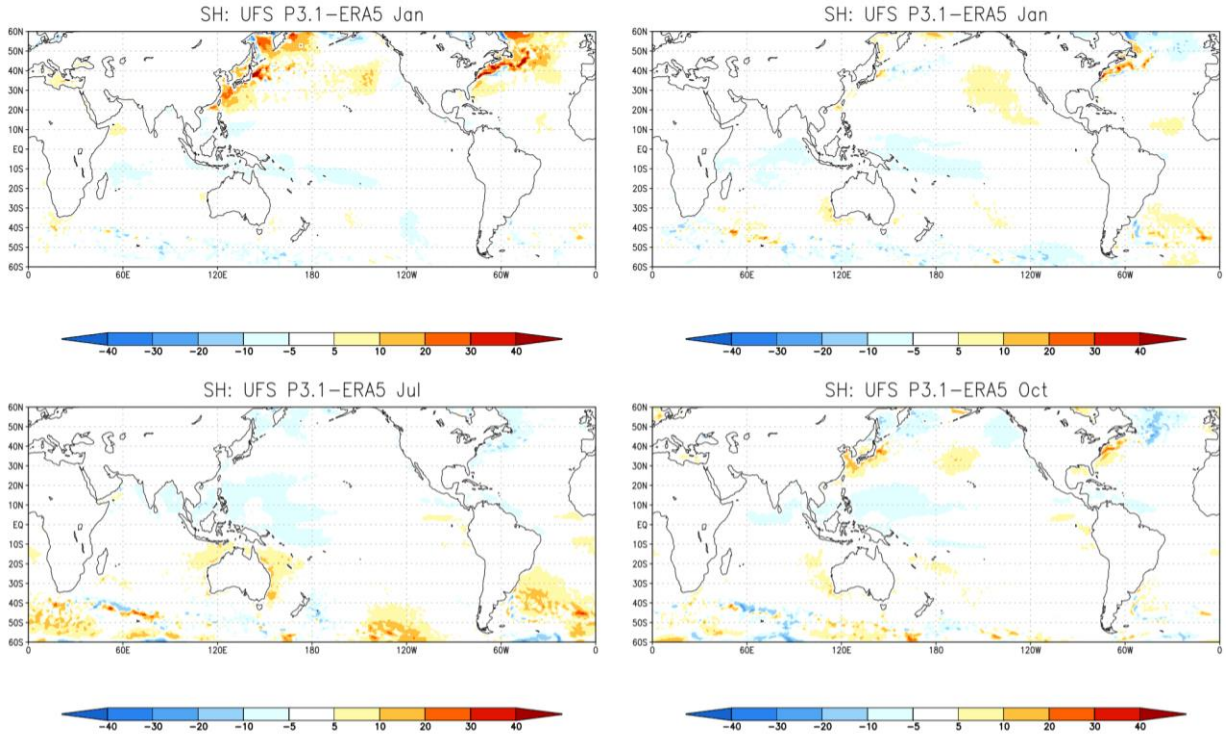

Figure 7.32: January, April, July and October mean sensible heat differences between P3.1 and ERA5.

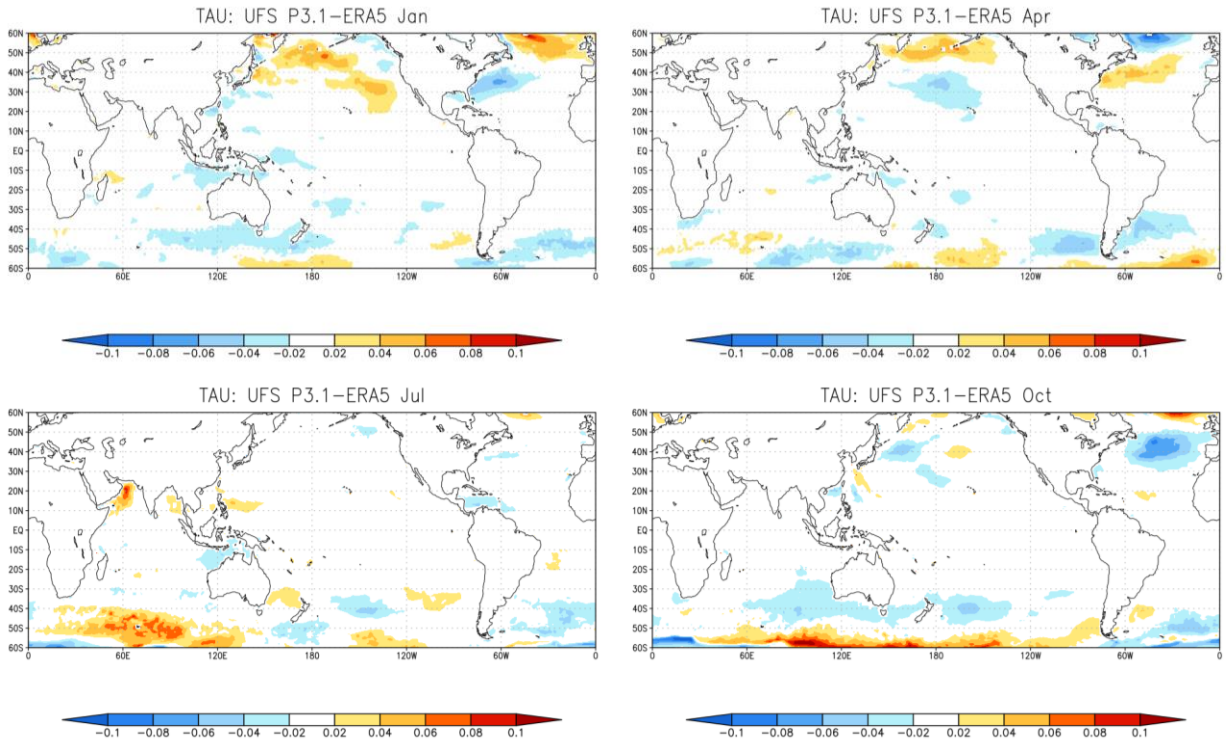

Figure 7.33: January, April, July and October mean zonal wind stress differences between UFS and ERA5.

P4 Figures

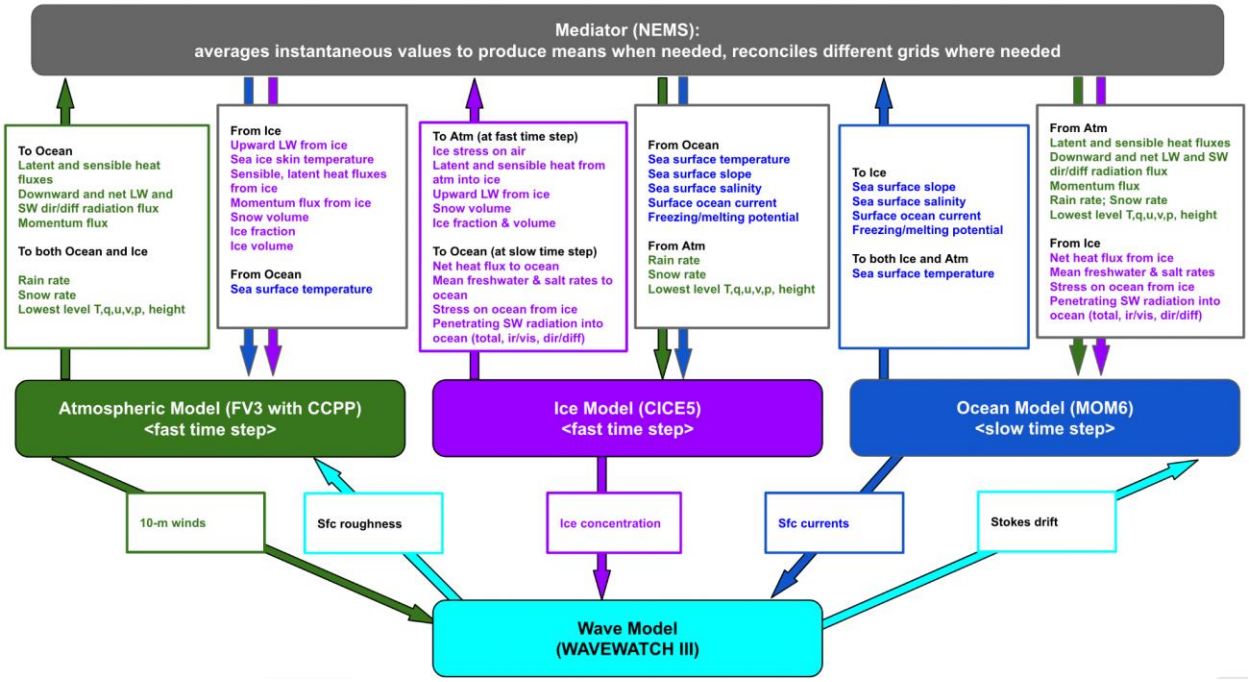

Figure 8.1: Schematic diagram of the coupling fields between the various components for P4.

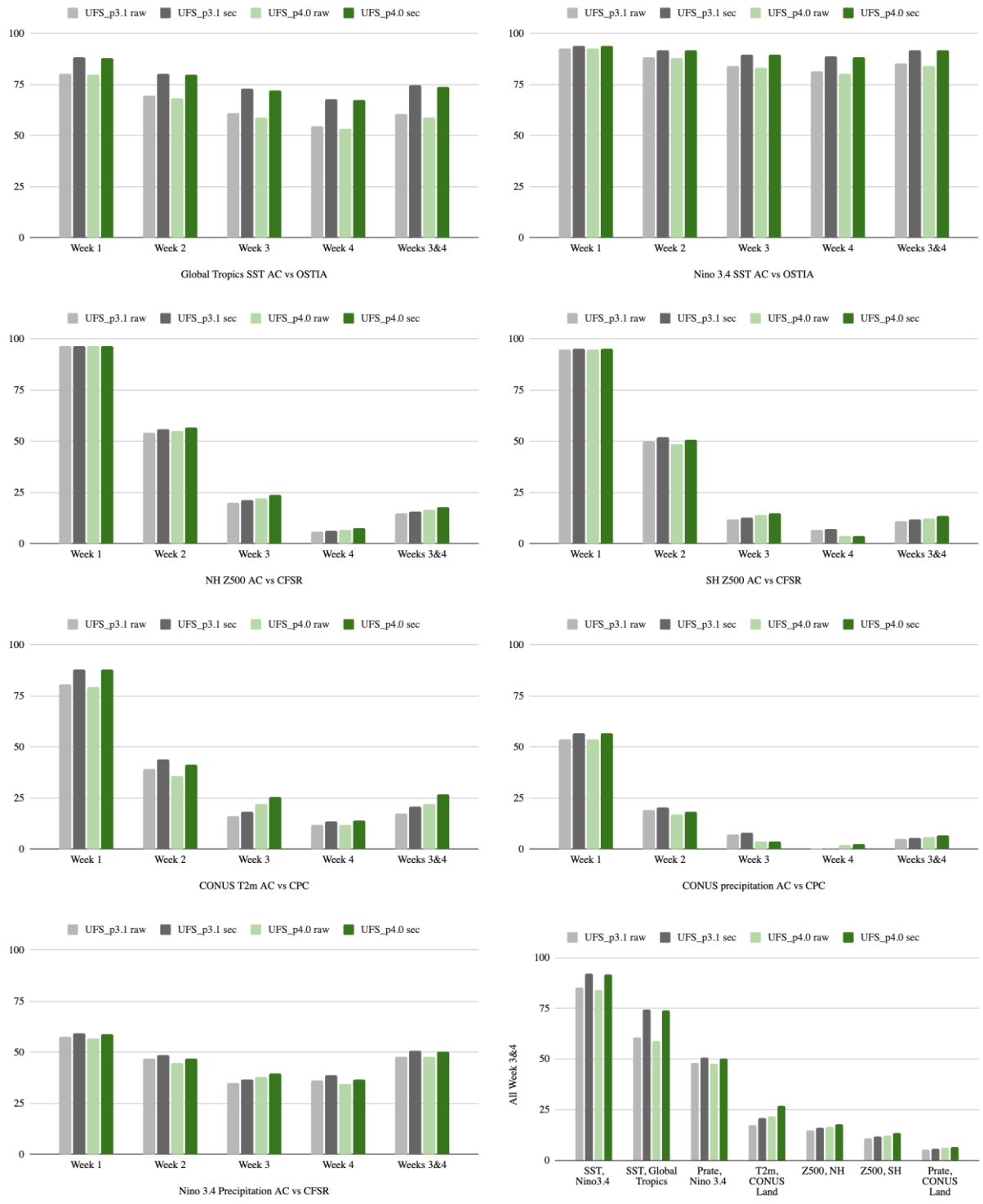

Figure 8.2: Anomaly correlation scores by forecast week of P3.1 (RAW-pale gray, SEC-dark gray) and P4 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of the anomaly correlation scores for Weeks 3&4.

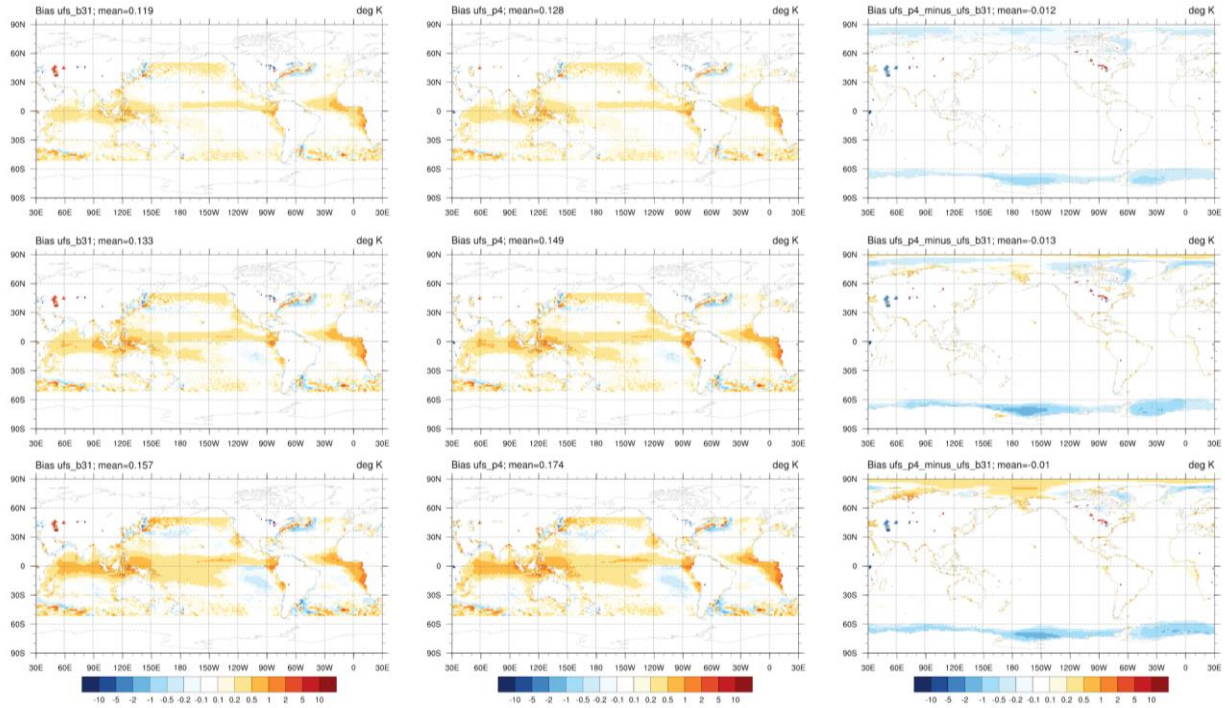

Figure 8.3: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean SST bias wrt OSTIA from 50°S to 50°N in P3.1 (left column) and P4 (middle column); Difference (P4 minus P3.1) in surface temperatures over ocean grid points, including sea ice (right column).

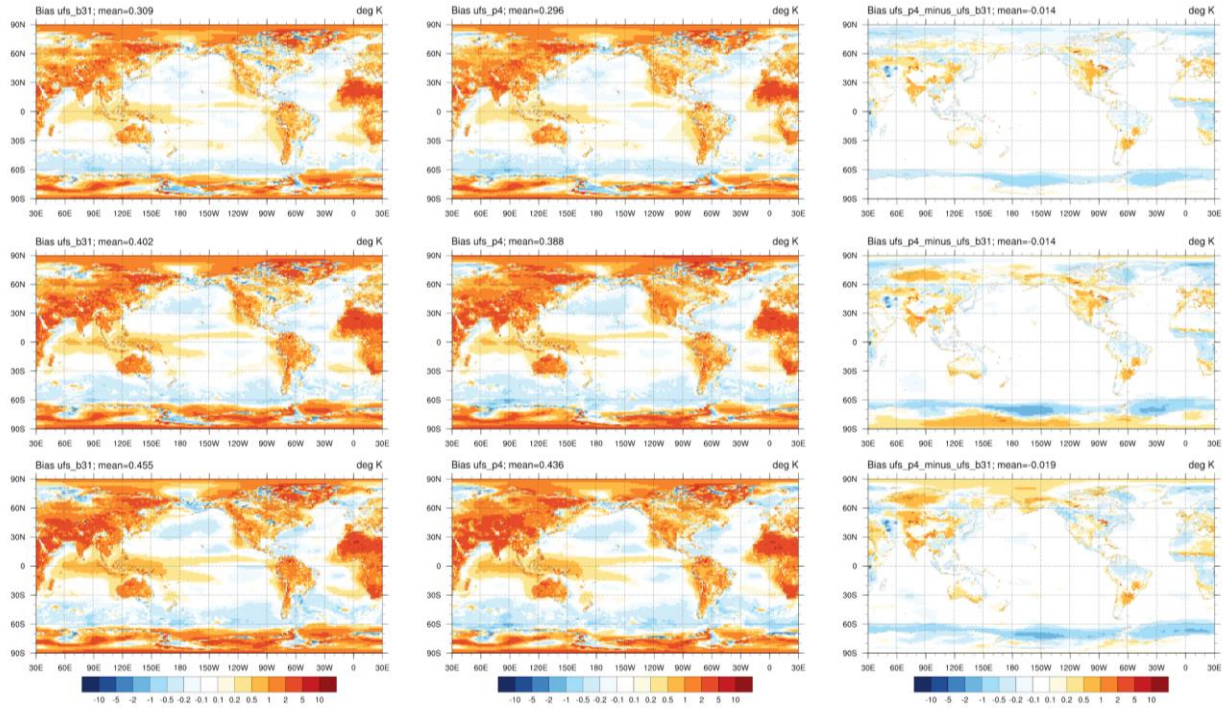

Figure 8.4: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in P3.1 (left column) and P4 (middle column); Difference (P3.1 minus P4) in 2-m air temperature (right column).

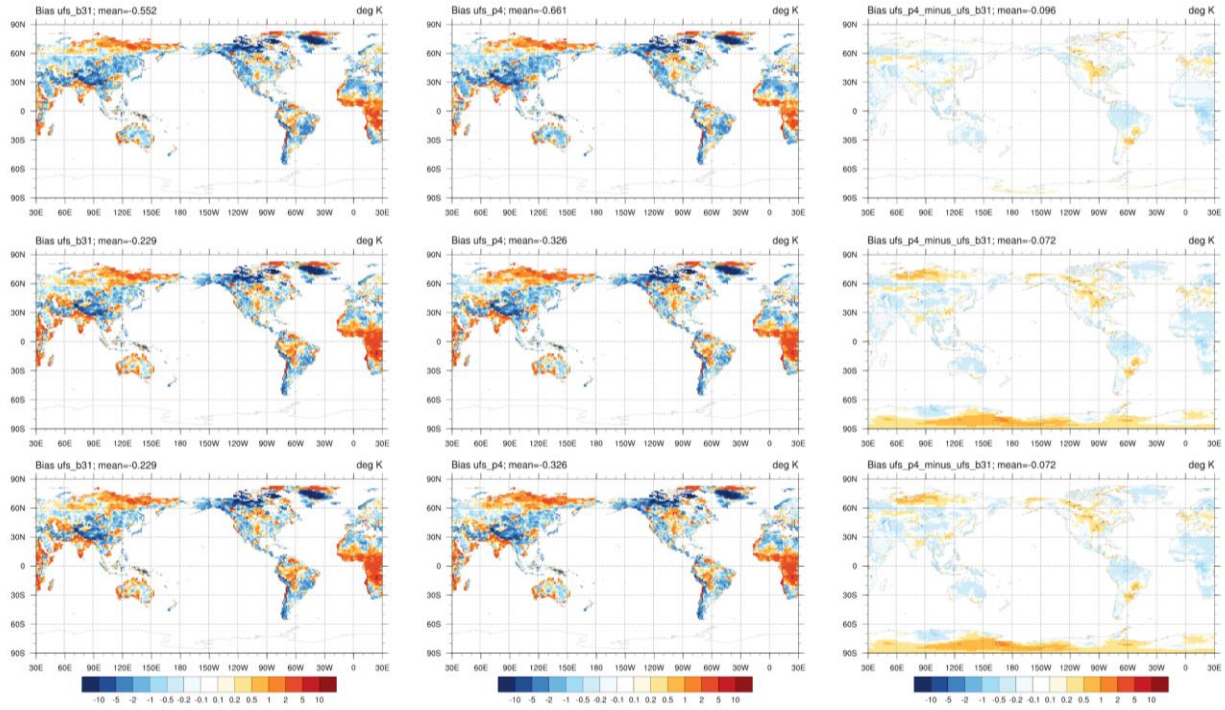

Figure 8.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in P3.1 (left column) and P4 (middle column); Difference (P4 minus P3.1) in 2-m daily maximum air temperature (right column).

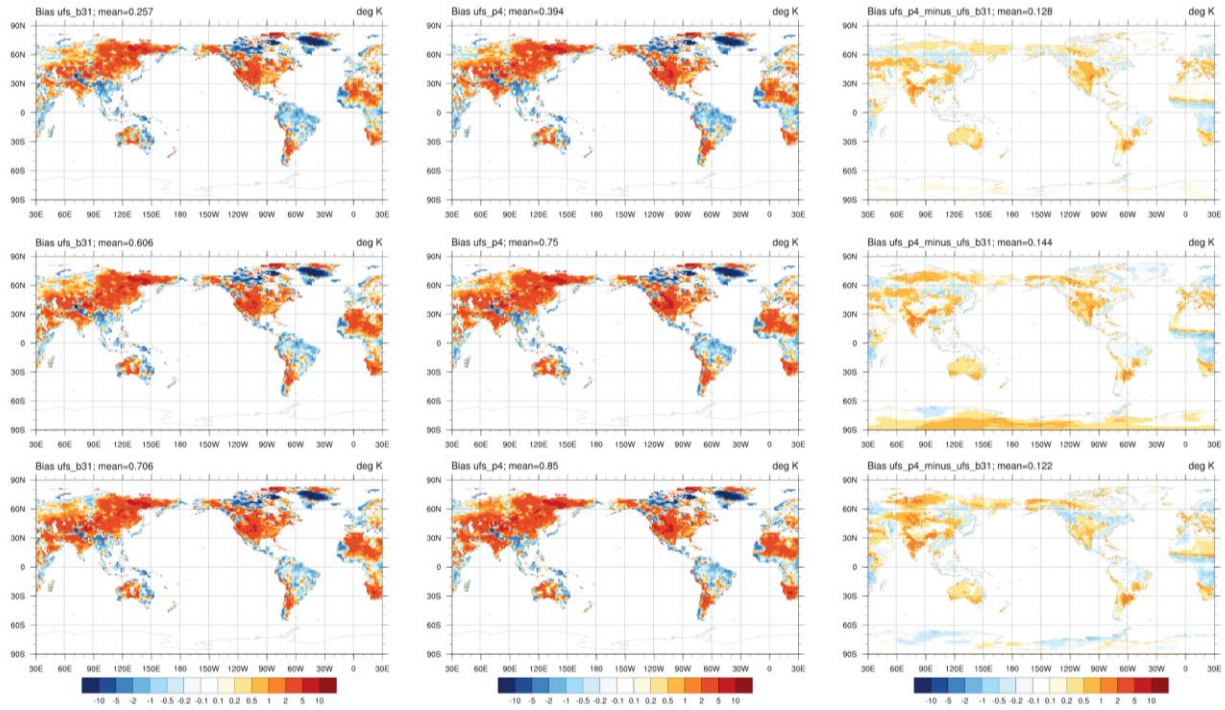

Figure 8.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in P3.1 (left column) and P4 (middle column); Difference (P4 minus P3.1) in 2-m daily minimum air temperature (right column).

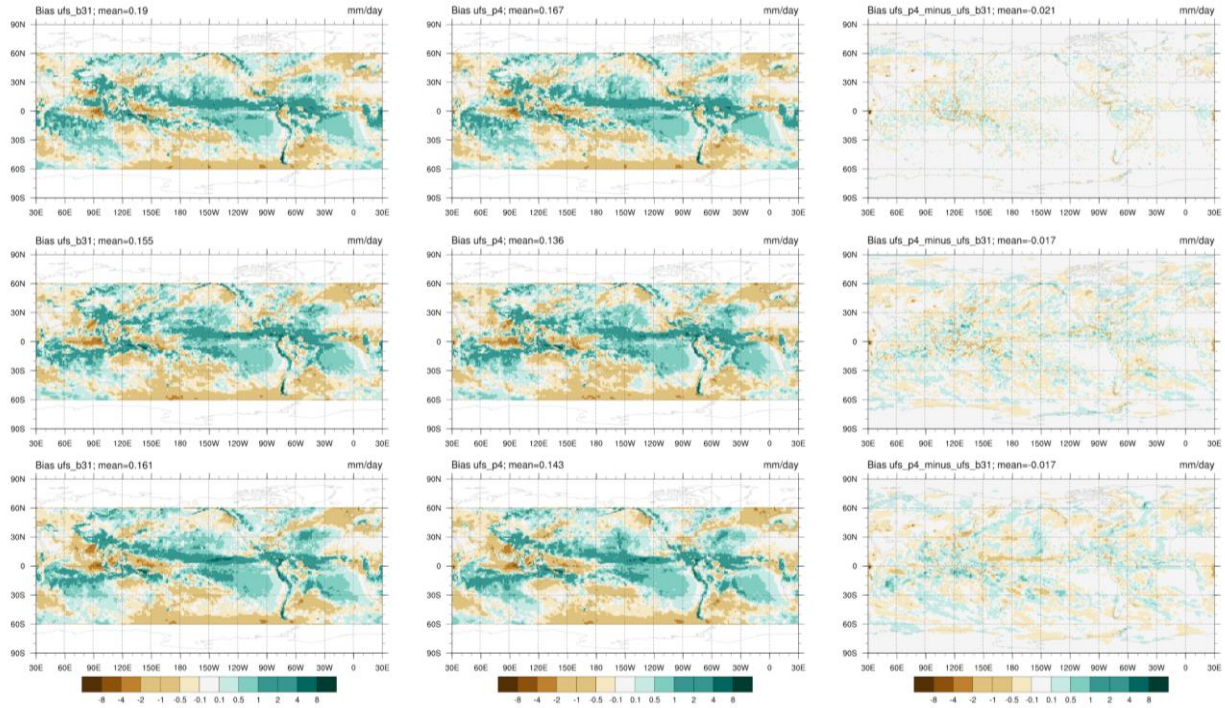

Figure 8.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in P3.1 (left column) and P4 (middle column); Difference (P4 minus P3.1) in precipitation rate (right column).

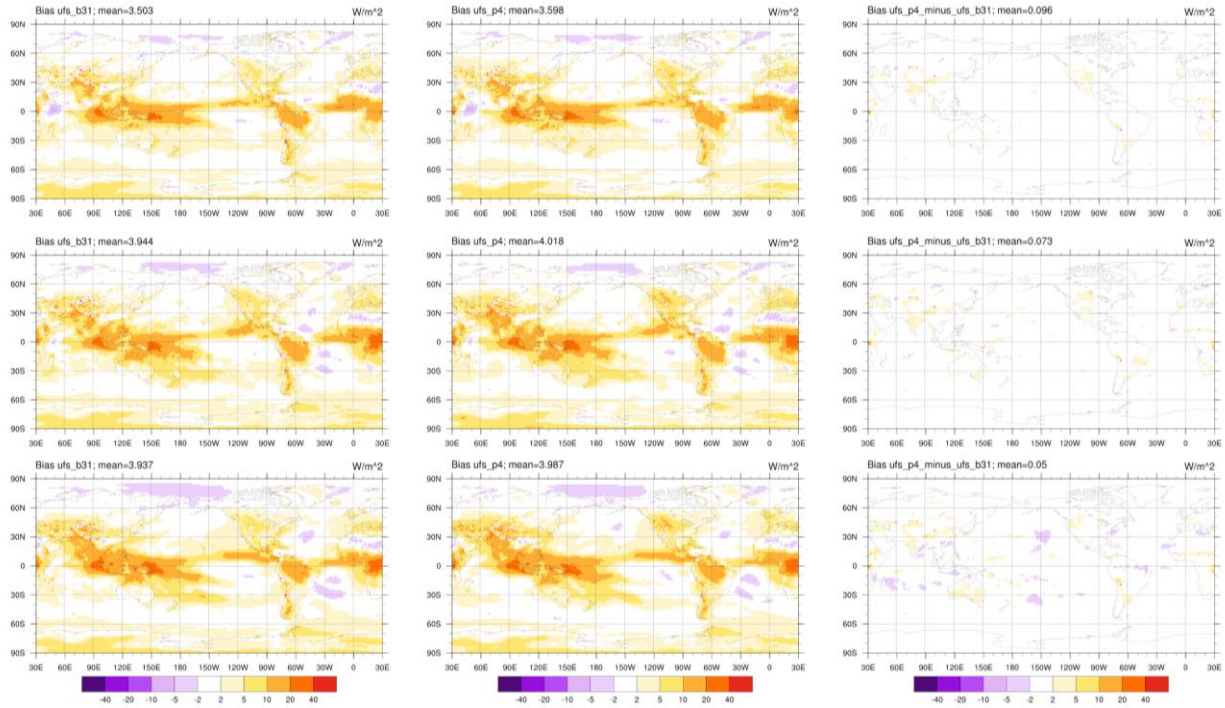

Figure 8.8: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in P3.1 (left column) and P4 (middle column); Difference (P4 minus P3.1) in OLR (right column).

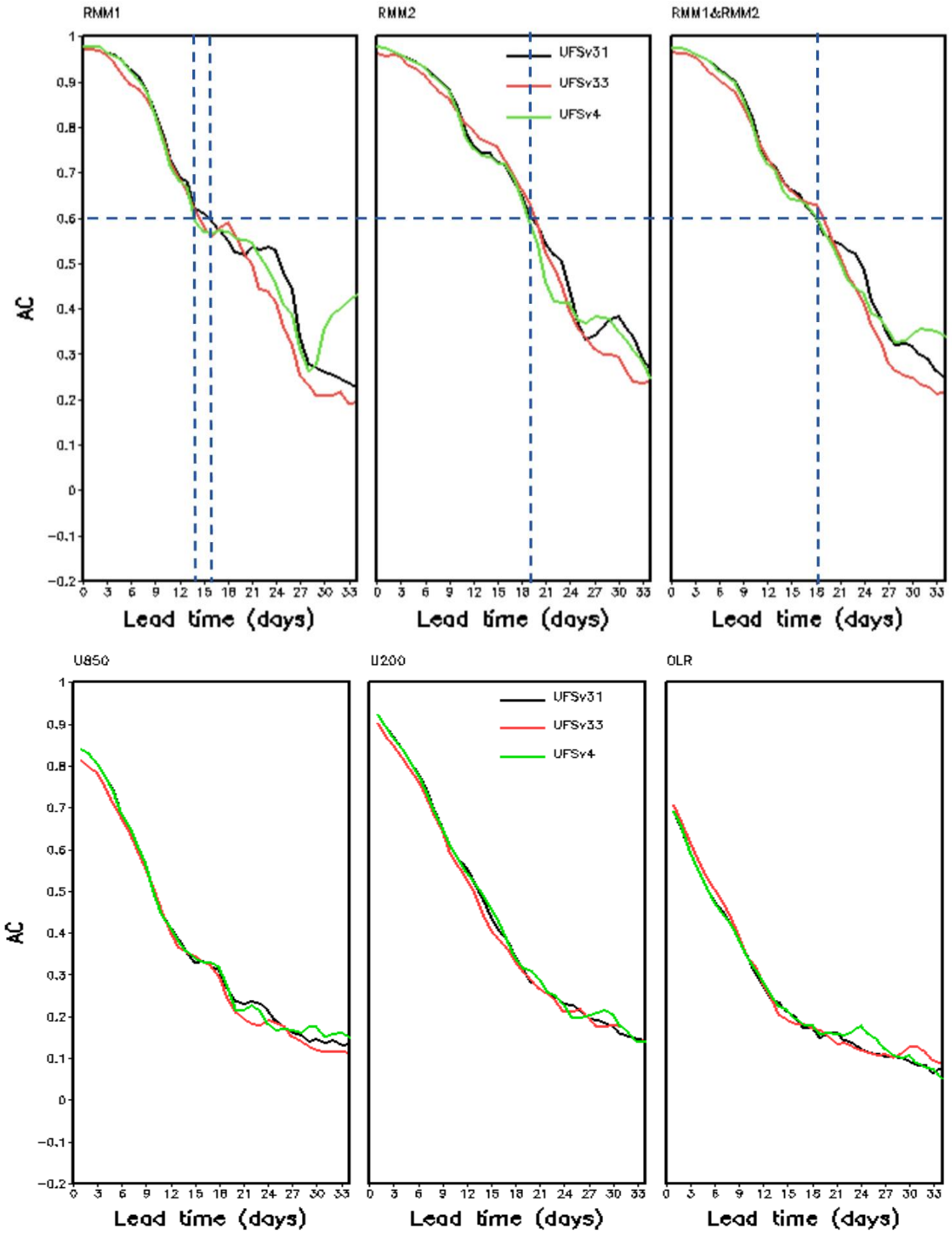

Figure 8.9: Top: AC of MJO RMM1 (left), RMM2 (middle) and combined RMM1&RMM2 (right). Bottom: AC of equatorial filtered zonal wind at 850mb (left) and 200mb (middle), and equatorial filtered OLR (right) in P3.1 (black line), P3.3 (red line; note that this intermediate prototype is beside the scope of the present paper), and P4 (green line).

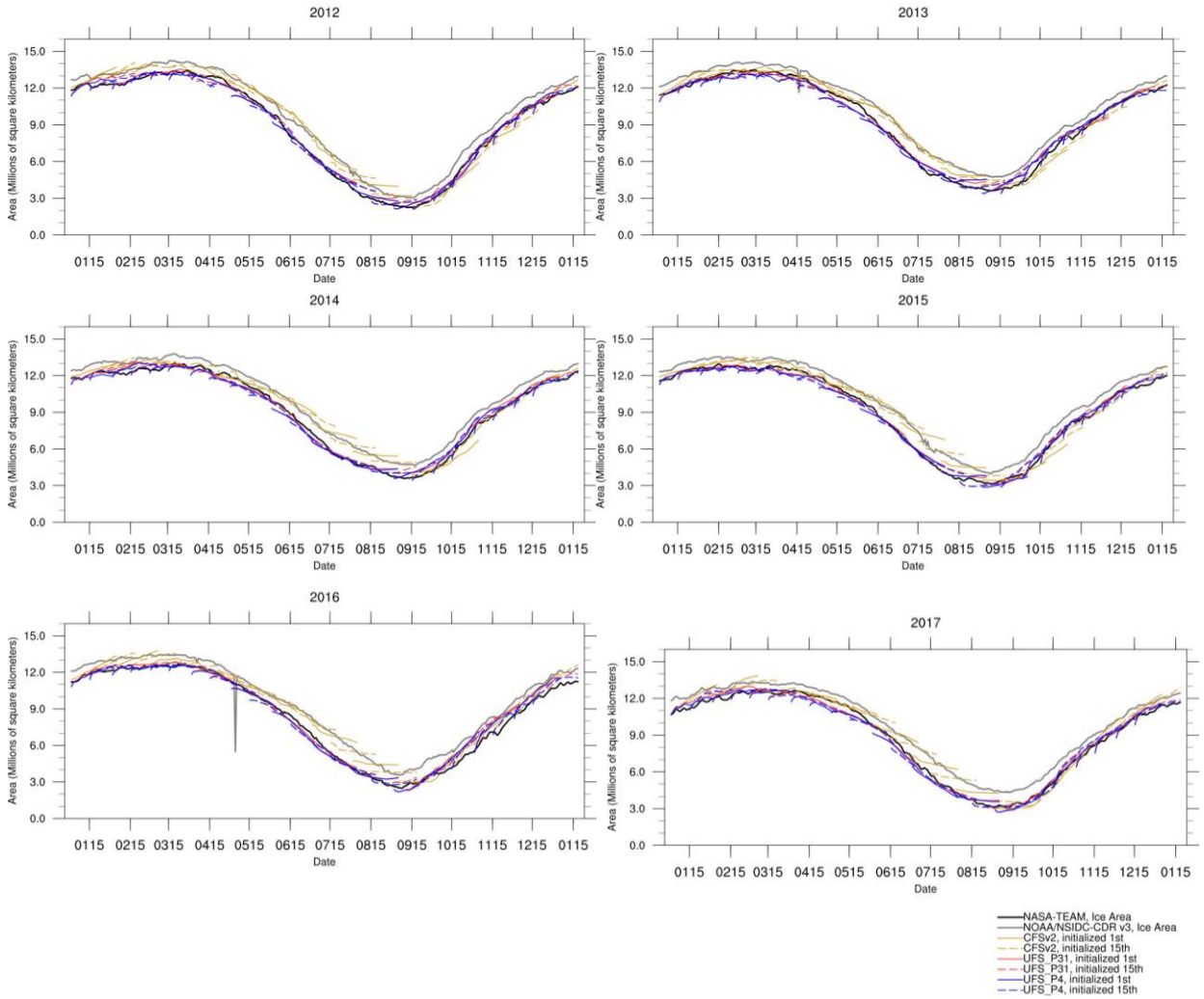

Figure 8.10: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow), P3.1 (red), and P4 (blue). Thick lines correspond to the observational estimate derived from ice concentrations from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date.

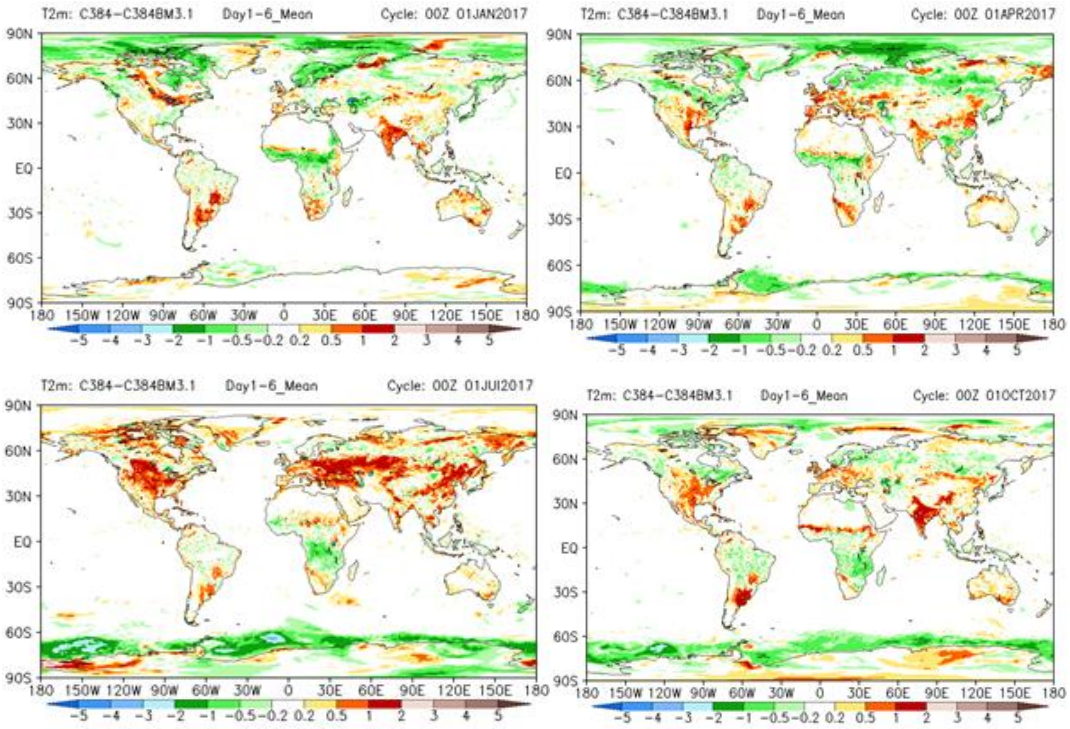

Figure 8.11: Mean day 1-6 difference (P4 minus P3.1) of 2-m air temperature in forecasts initialized on the 1st day of the month in January (top left), April (top right), July (bottom left) and October (bottom right) of 2017.

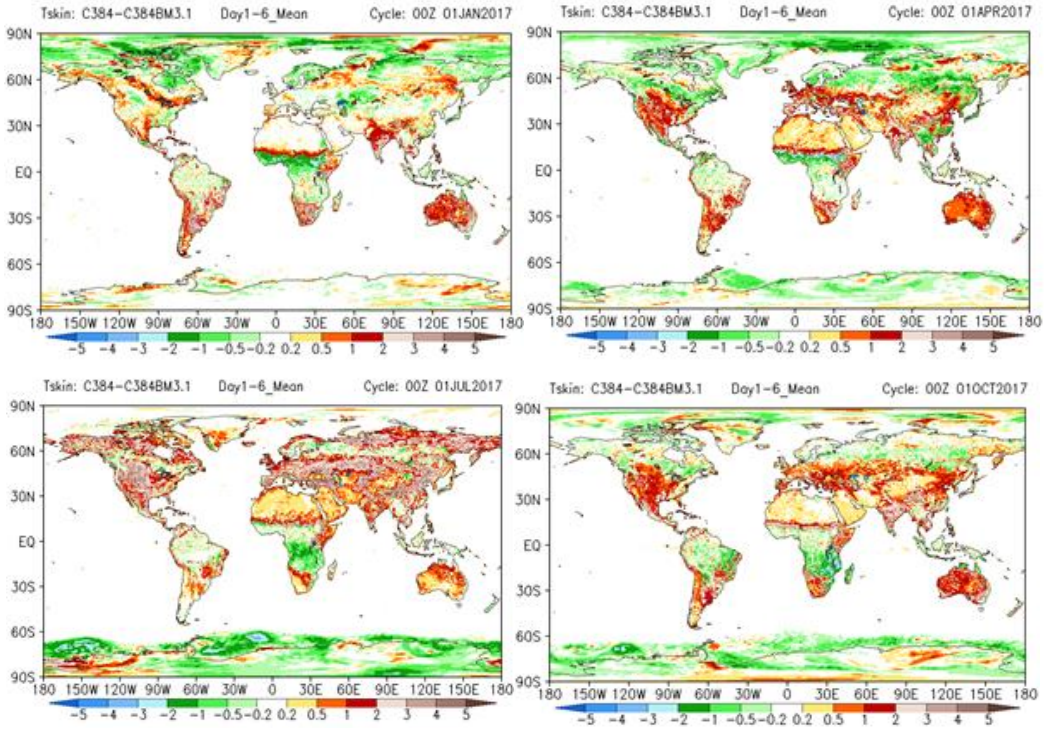

Figure 8.12: Mean day 1-6 difference (P4 minus P3.1) of surface skin temperature in forecasts initialized on the 1st day of the month in January (top left), April (top right), July (bottom left) and October (bottom right) of 2017.

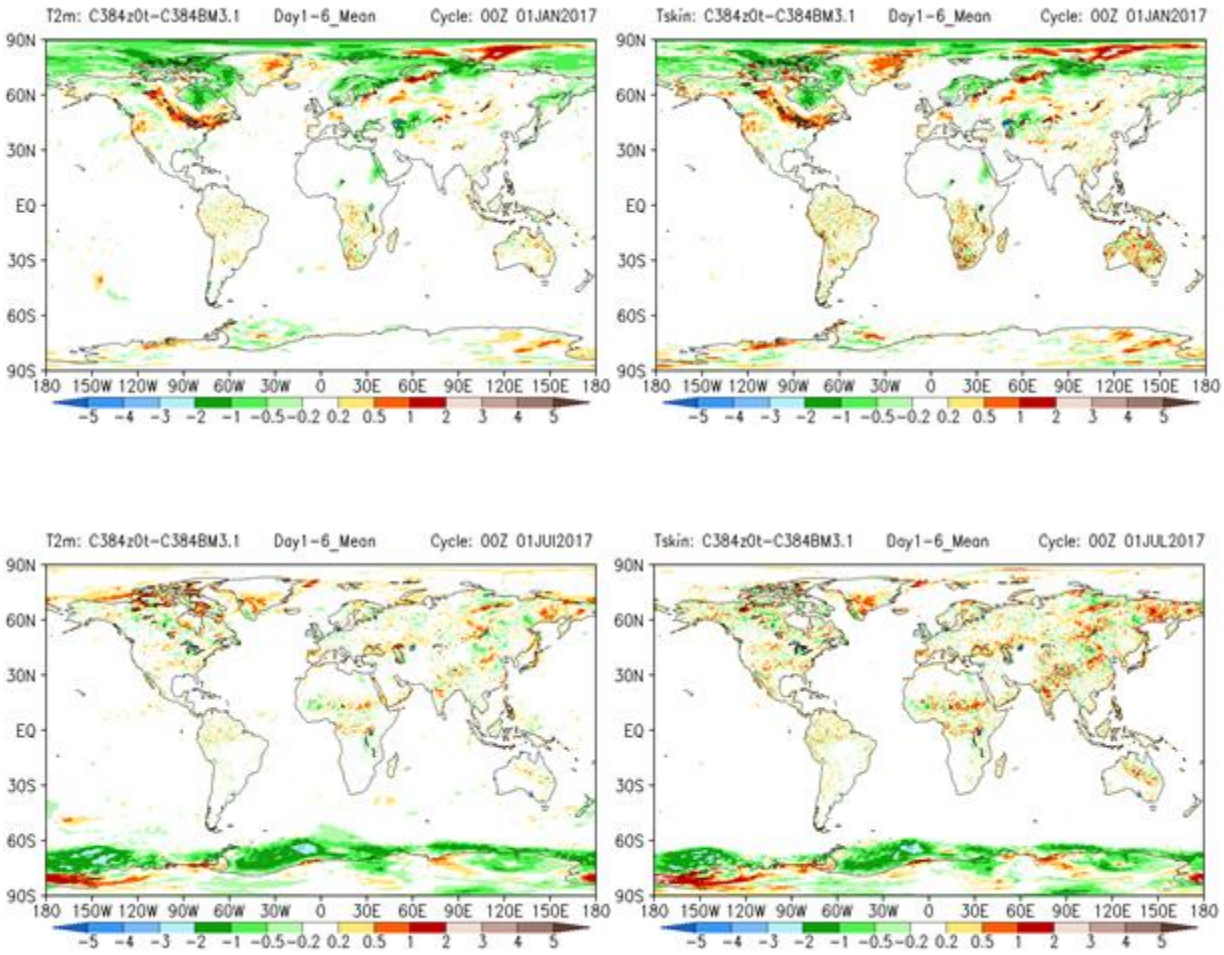

Figure 8.13: Mean day 1-6 difference (P4 EXP minus P3.1) of 2-m air temperature (left) and surface skin temperature (right) for forecasts initialized on the 1st day of the month in January (top) and July (bottom) of 2017.

P5 Figures

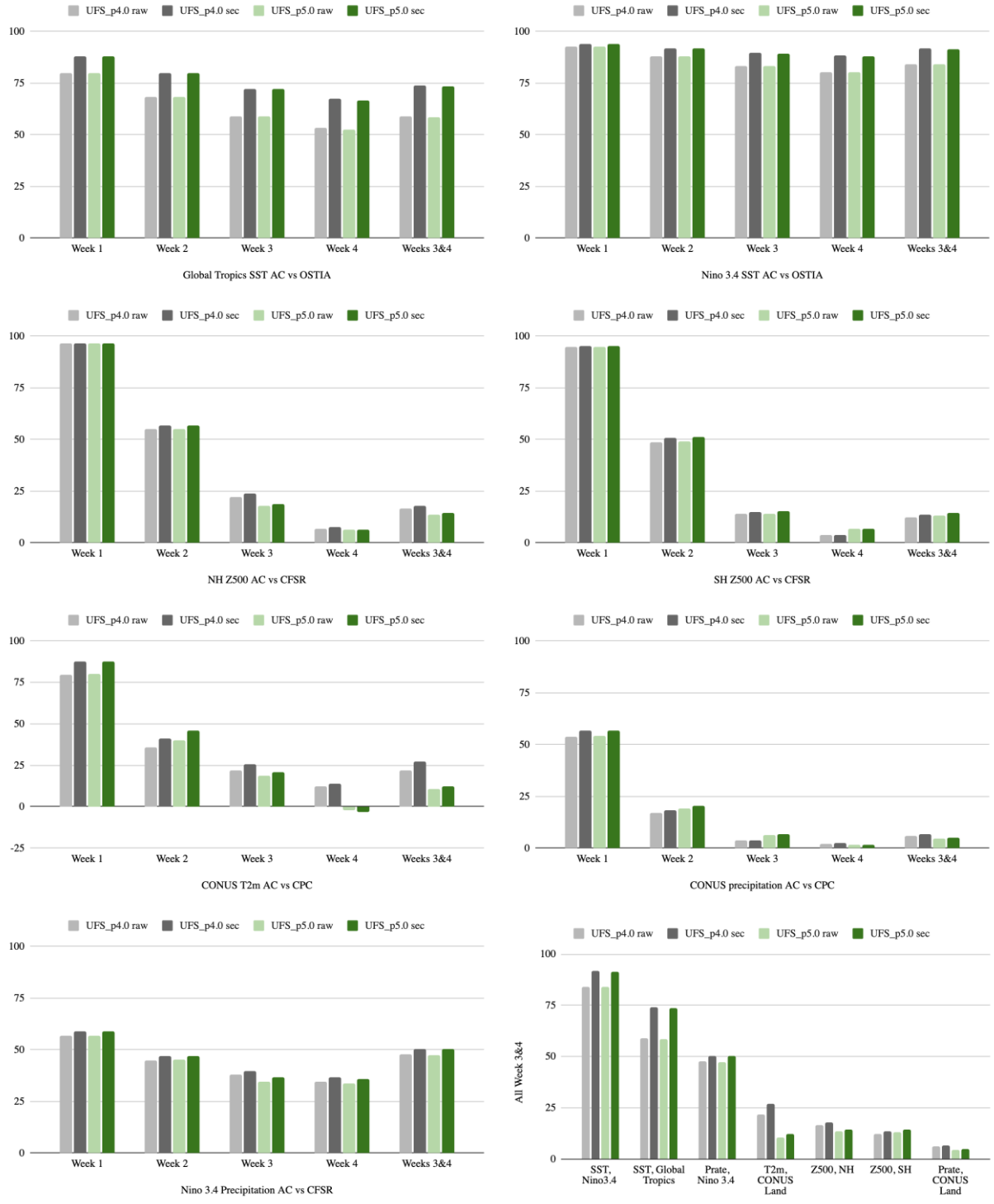

Figure 9.1: Anomaly correlation scores by forecast week of P4 (RAW-pale gray, SEC-dark gray) and P5 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of the anomaly correlation scores for Weeks 3&4.

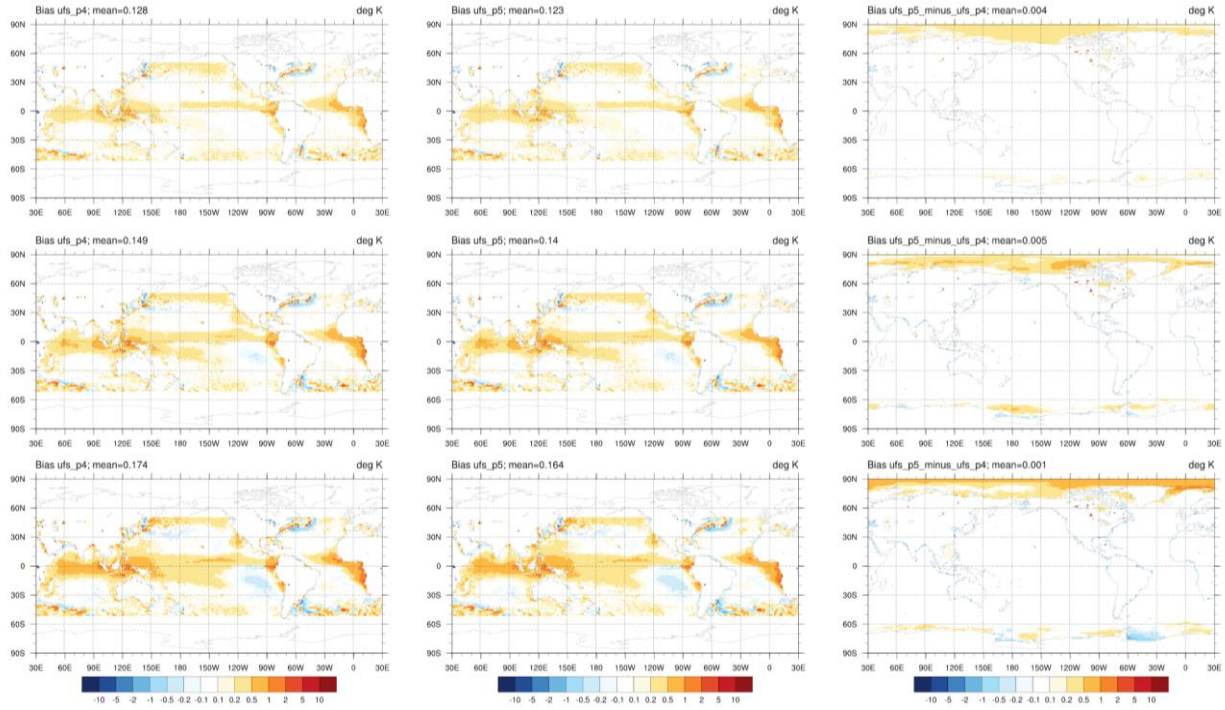

Figure 9.2: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean SST bias wrt OSTIA from 50°S to 50°N in P4 (left column) and P5 (middle column); Difference (P5 minus P4) in surface temperatures over ocean grid points, including sea ice (right column).

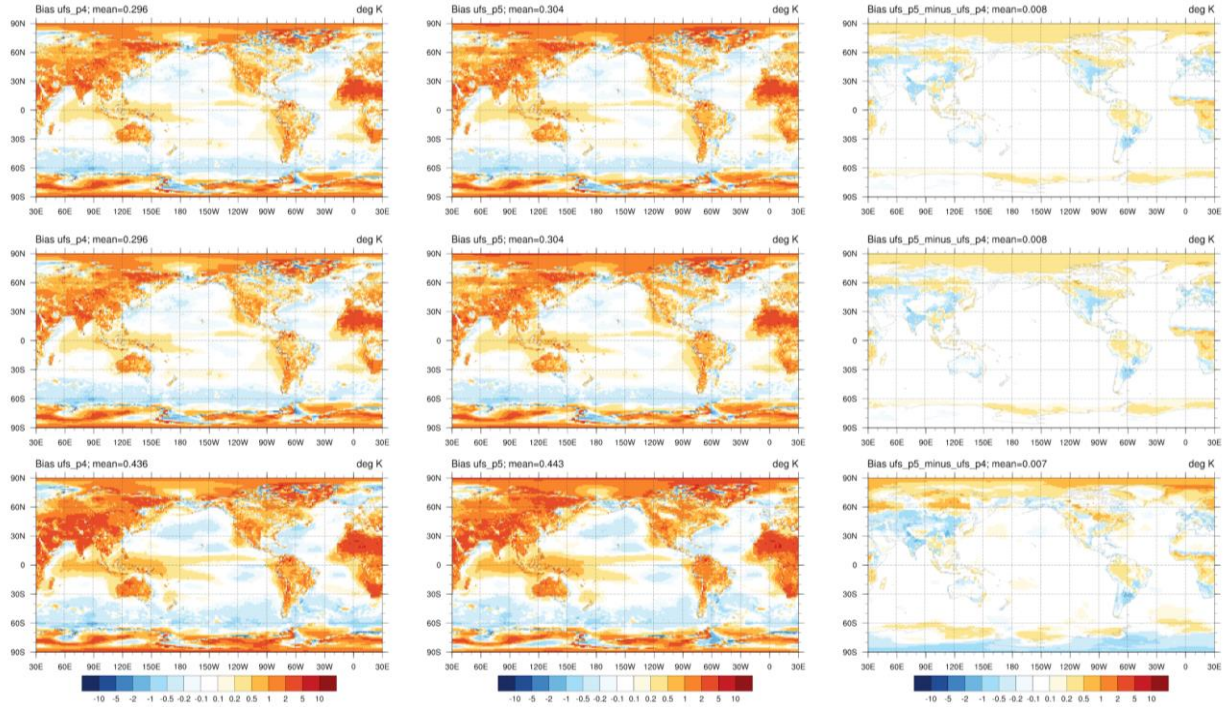

Figure 9.3: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in P4 (left column) and P5 (middle column); Difference (P4 minus P5) in 2-m air temperature (right column).

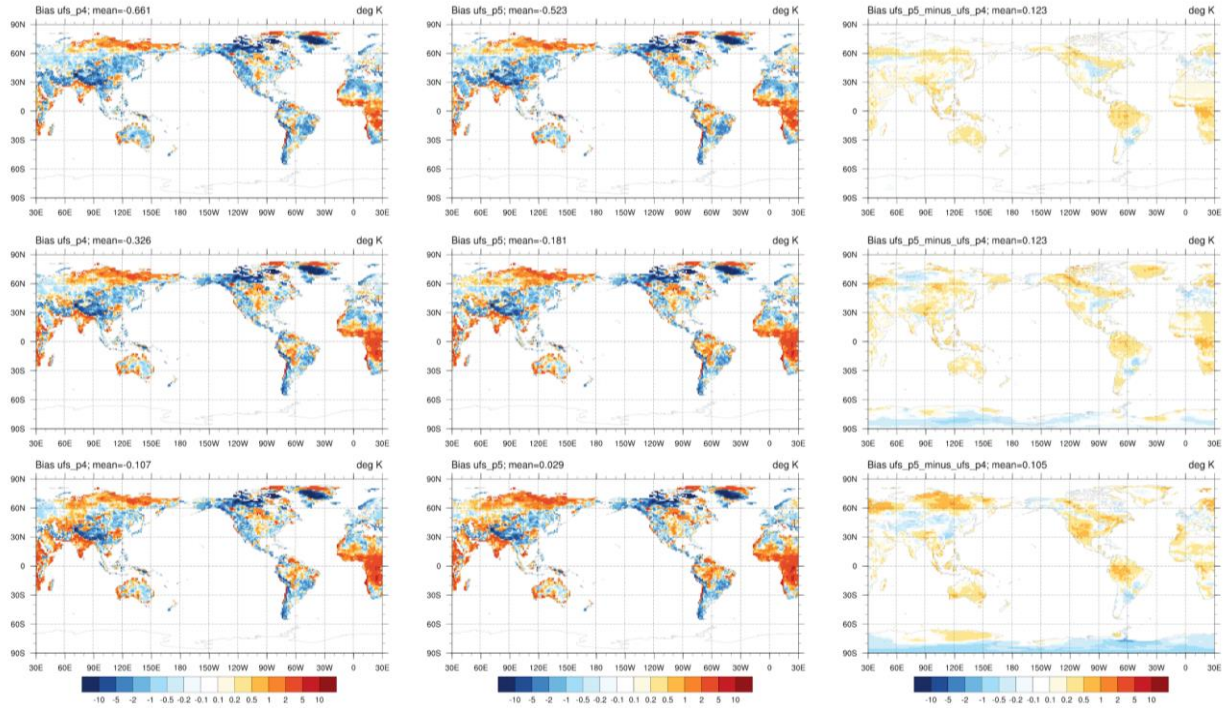

Figure 9.4: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in P4 (left column) and P5 (middle column); Difference (P5 minus P4) in 2-m daily maximum air temperature (right column).

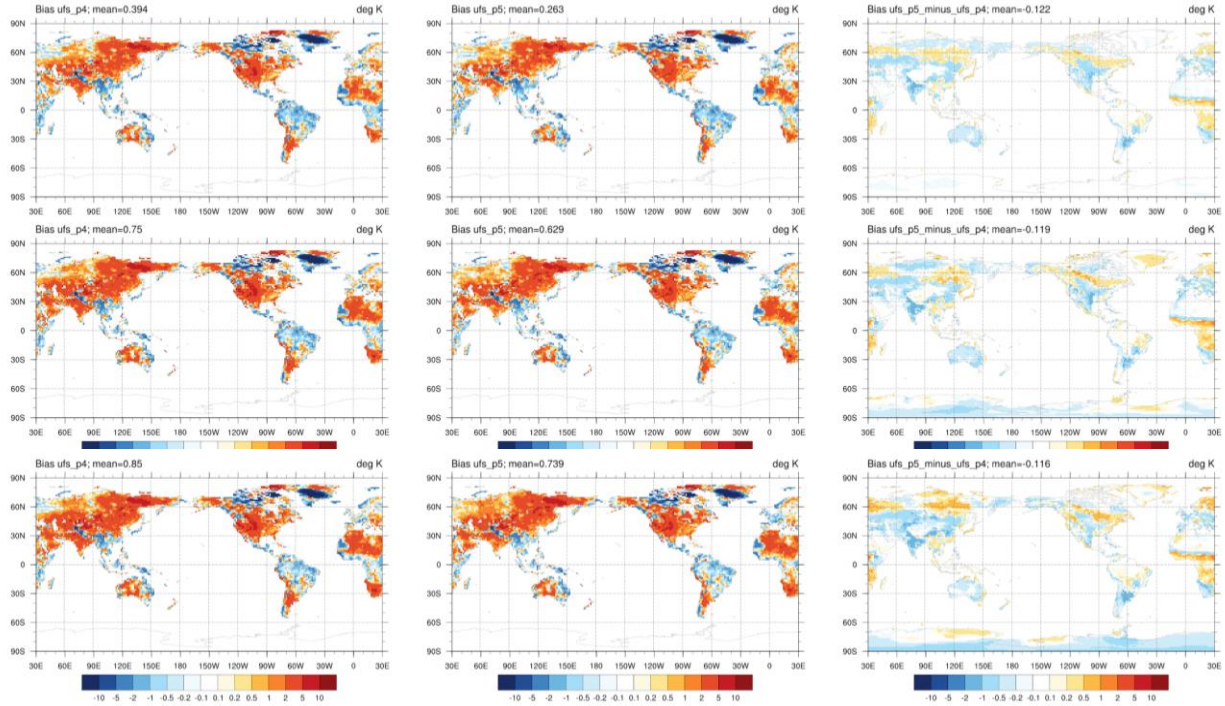

Figure 9.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in P4 (left column) and P5 (middle column); Difference (P5 minus P4) in 2-m daily minimum air temperature (right column).

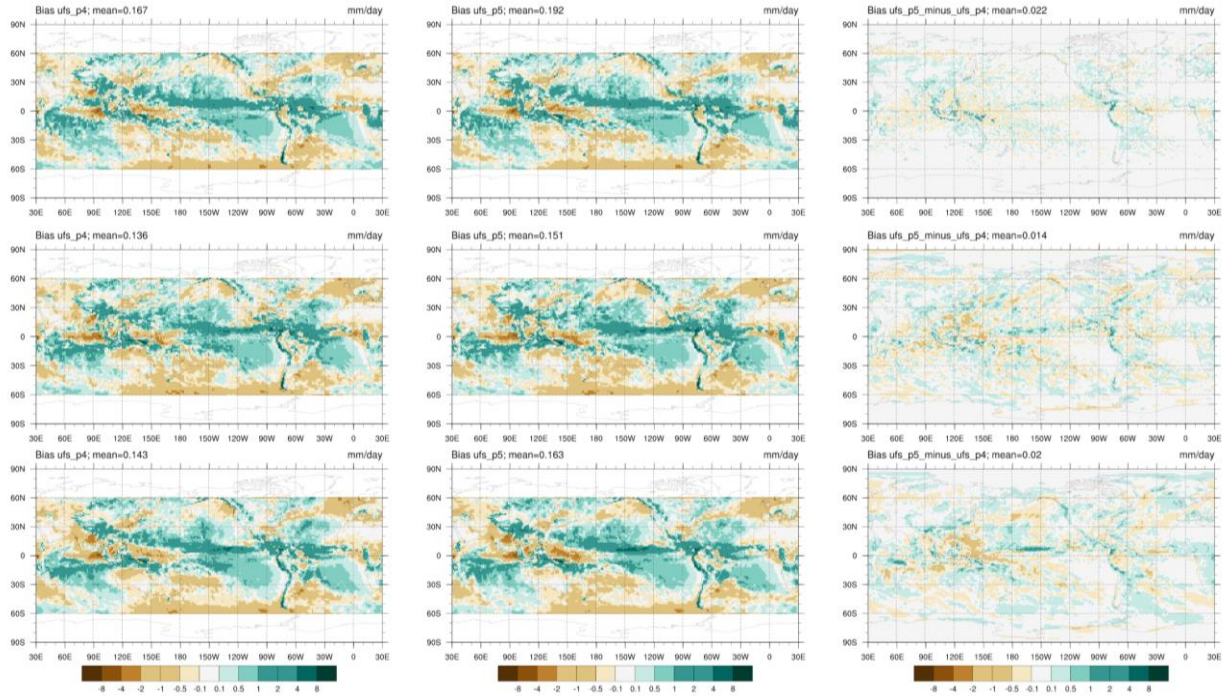

Figure 9.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in P4 (left column) and P5 (middle column); Difference (P5 minus P4) in precipitation rate (right column).

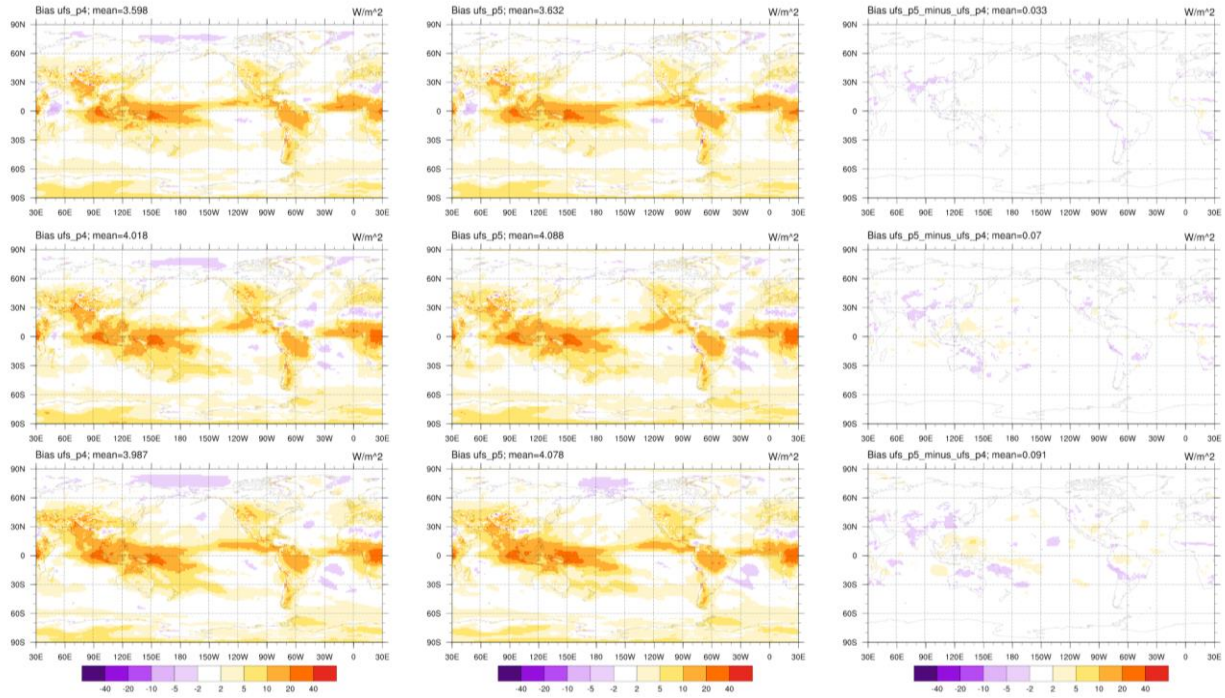

Figure 9.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in P4 (left column) and P5 (middle column); Difference (P5 minus P4) in OLR (right column).

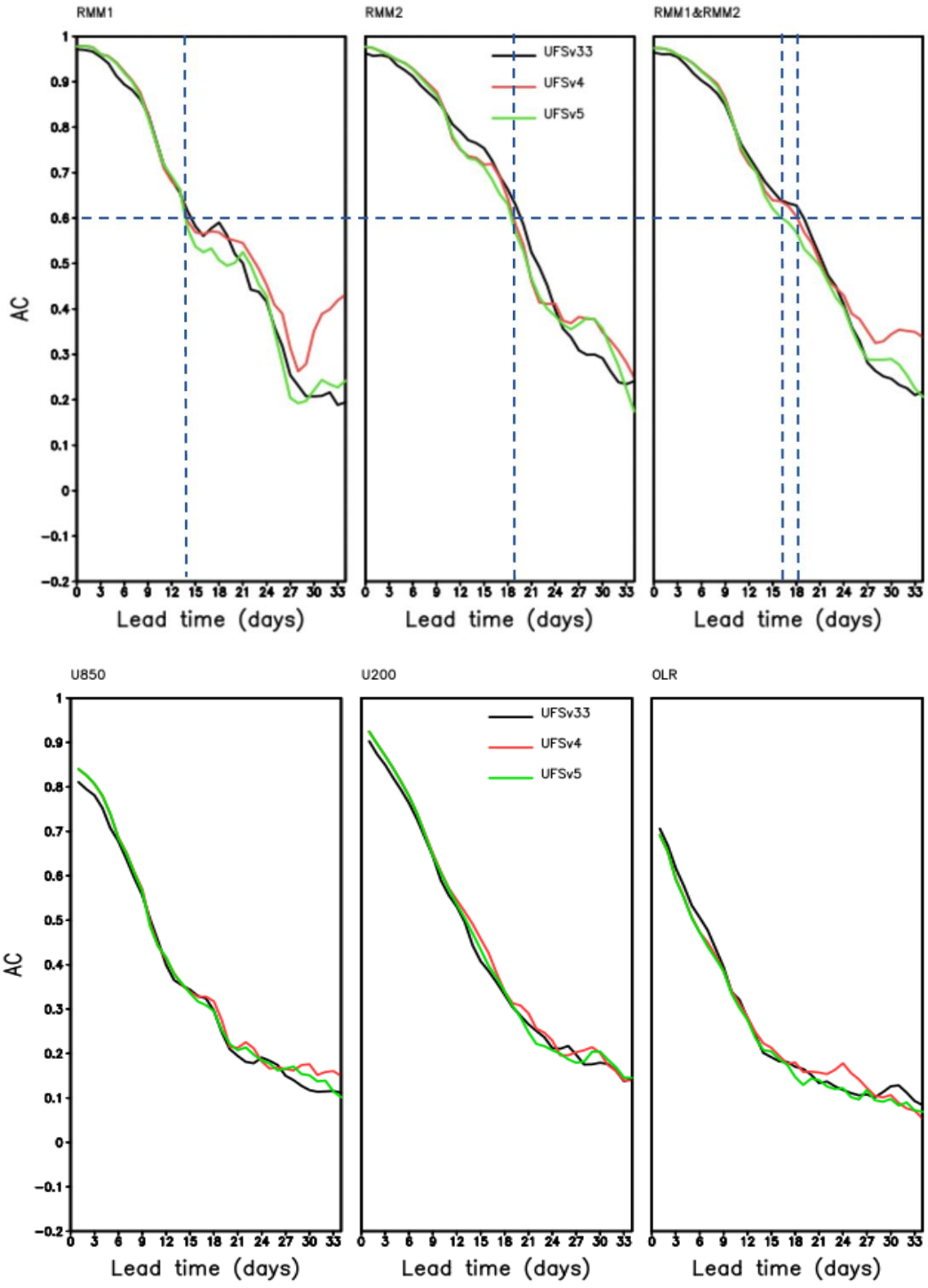

Figure 9.8: Top: AC of MJO RMM1 (left), RMM2 (middle) and combined RMM1&RMM2 (right). Bottom: AC of equatorial filtered zonal wind at 850mb (left) and 200mb (middle), and equatorial filtered OLR (right) in P3.3 (black line; note that this intermediate prototype is beside the scope of the present paper), P4 (red line), and P5 (green line).

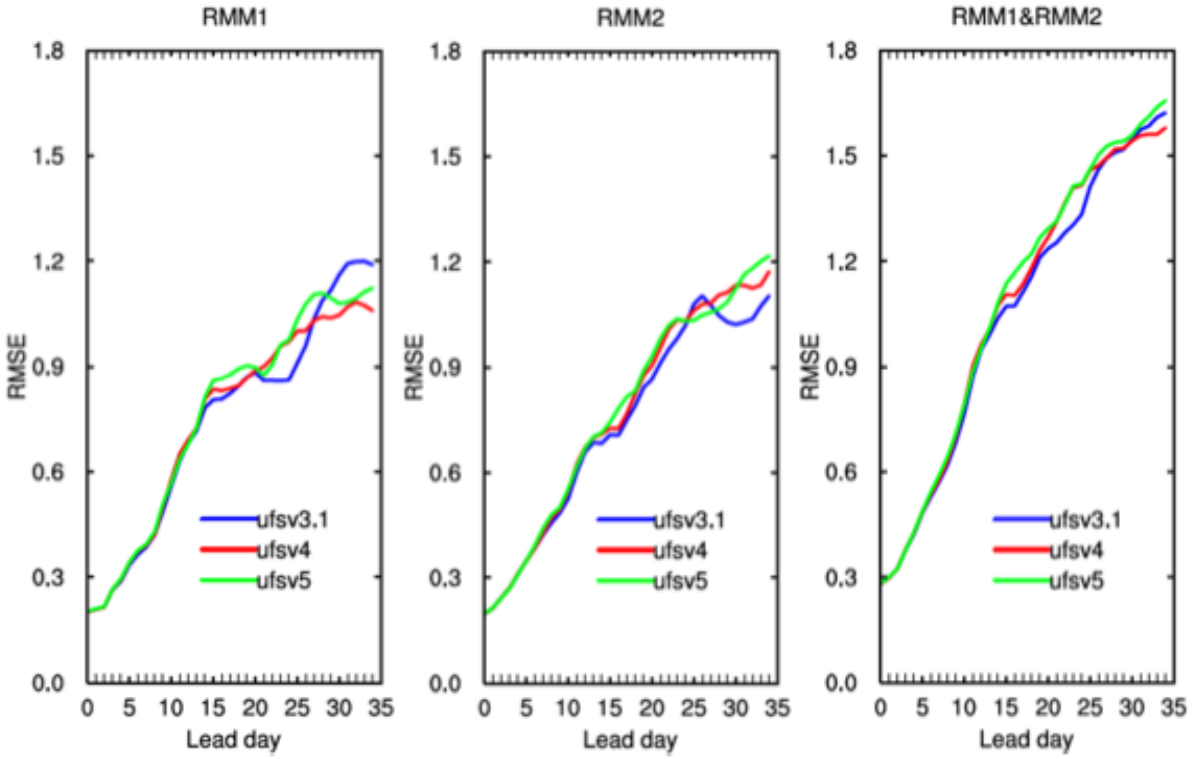

Figure 9.9: RMSE of MJO RMM1, RMM2 and RMM1&2 for P3.1(blue), P4 (red), and P5 (green)

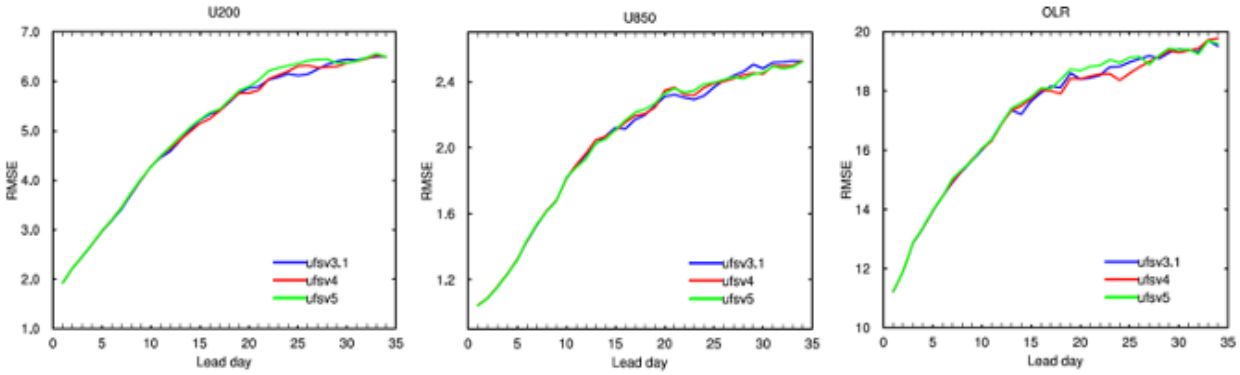

Figure 9.10: RMSE of MJO components (U200, U850 and OLR) for P3.1 (blue), P4 (red), and P5 (green)

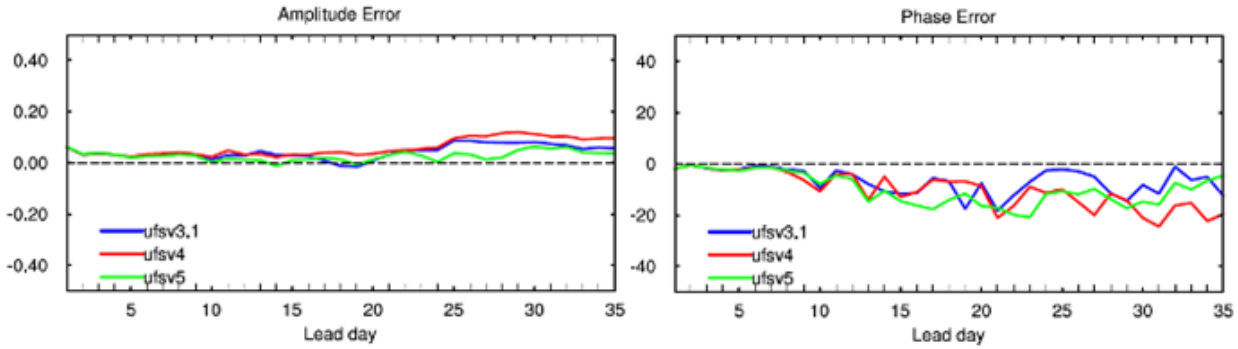

Figure 9.11: Error of MJO amplitude (left) and phase (right) for for P3.1 (blue), P4 (red), and P5 (green)

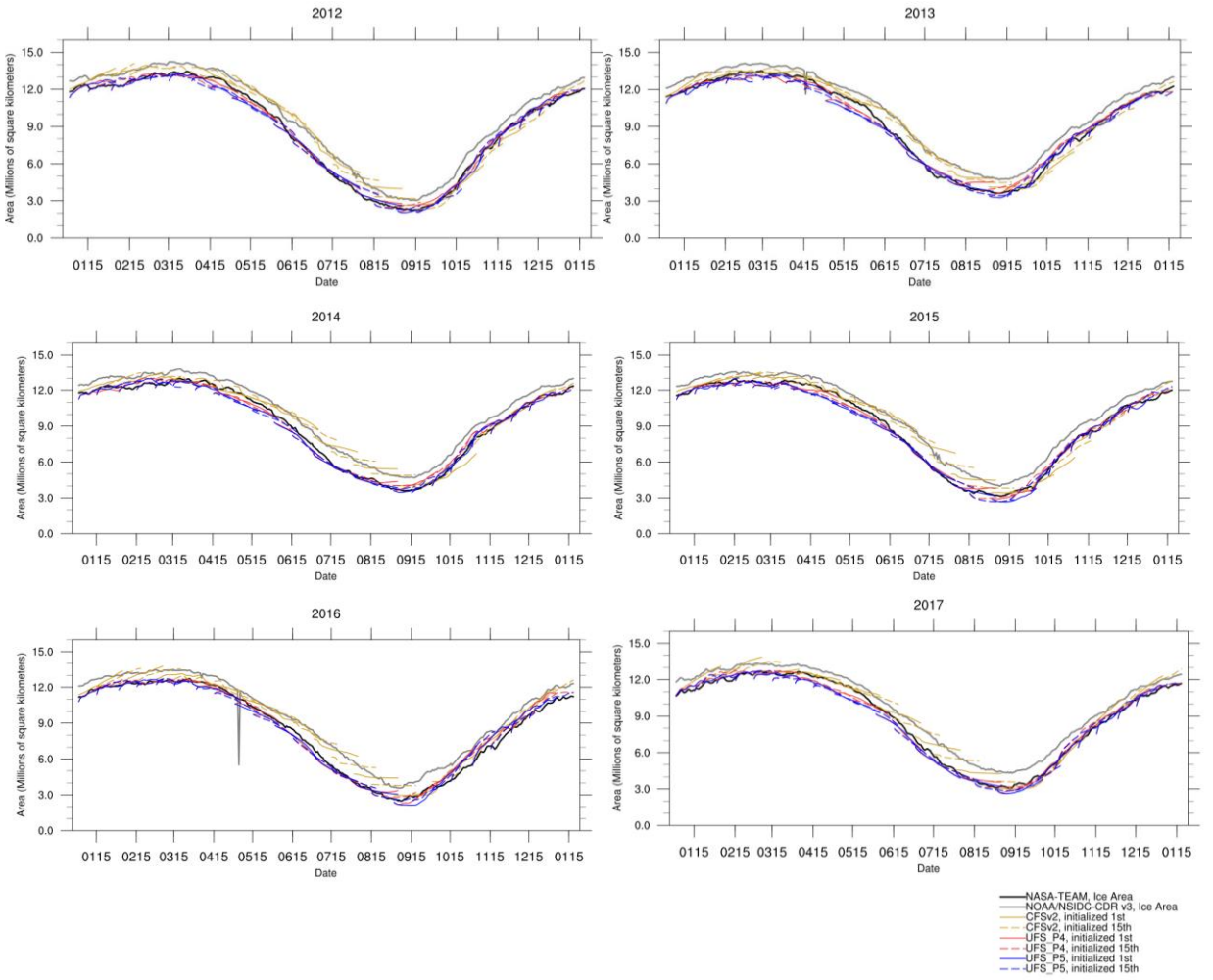

Figure 9.12: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow), P4 (red), and P5 (blue). Thick lines correspond to the observational estimate derived from ice concentrations from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date.

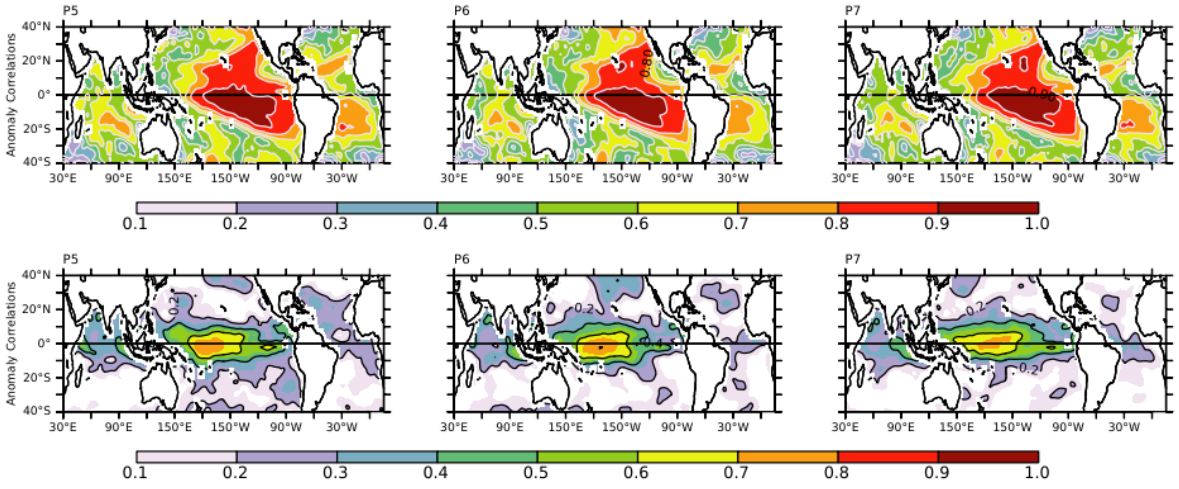

Figure 9.13: Anomaly correlation coefficient of Week 3&4 averaged SST forecast in P5, P6, and P7 with OSTIA (top row). The same, but for zonal wind stress forecast with ERA5 (bottom row).

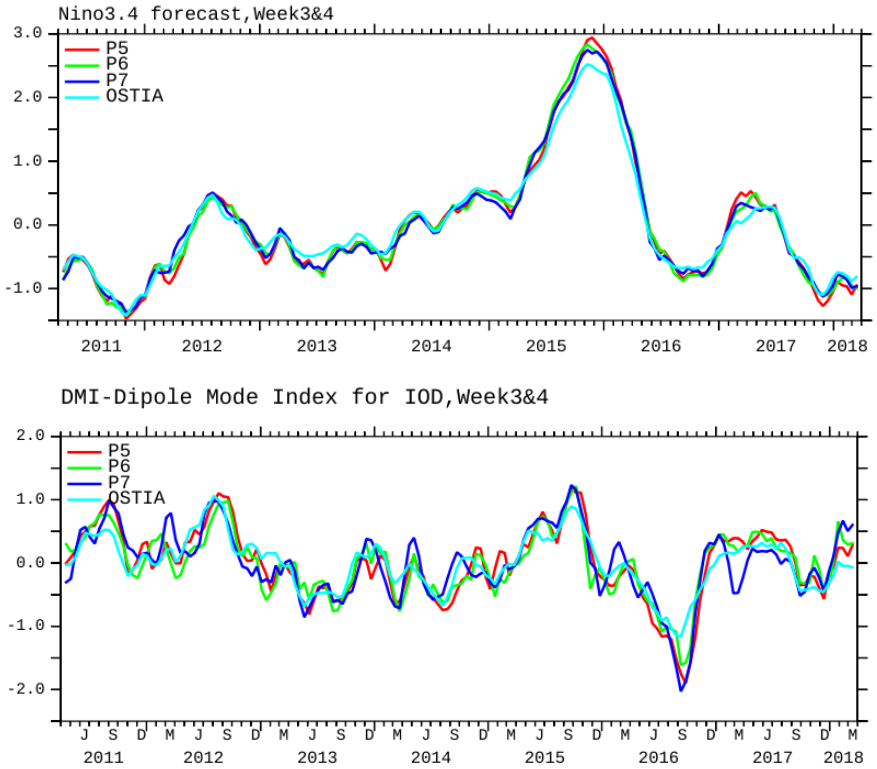

Figure 9.14: Week 3&4 forecast of Niño 3.4 index (top) and Dipole Model Index (DMI) for the period Apr, 2011–Mar, 2018 in P5, P6, P7 and concurrent time series from OSTIA.

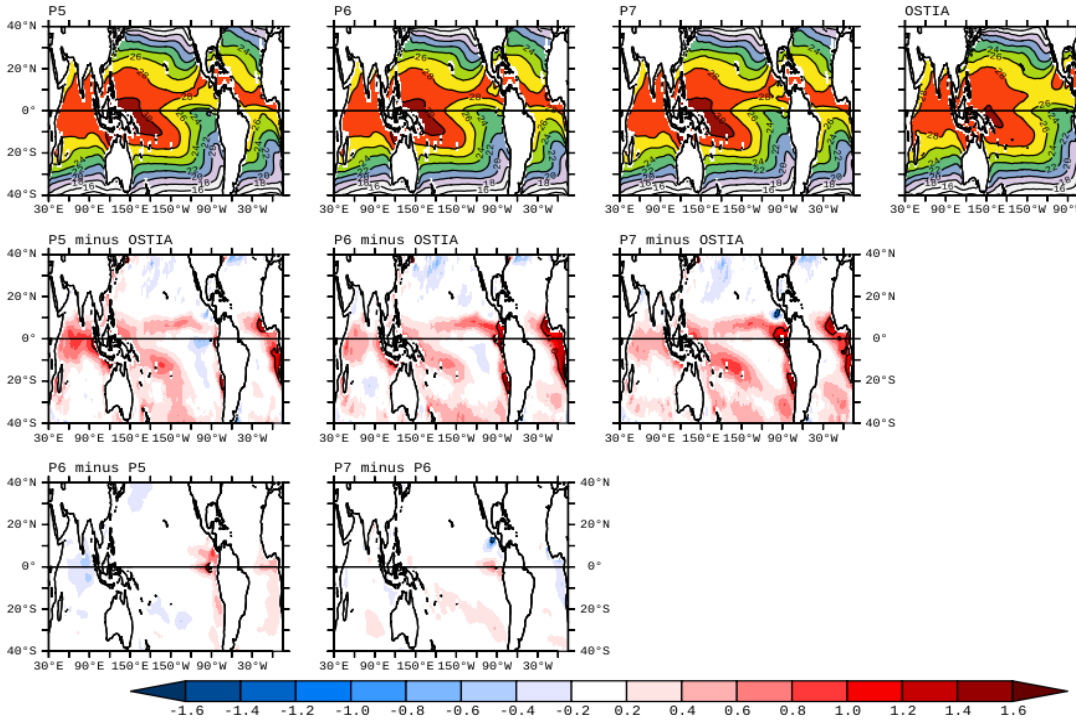

Figure 9.15: Week 3&4 average forecast of SST averaged for October, November, December (OND) from P5, P6, P7 and corresponding SST from OSTIA (top). The forecast biases in individual prototypes are shown in the middle row in the same order. The changes in SST forecast from P5 to P6 and P6 to P7 are shown in the last row.

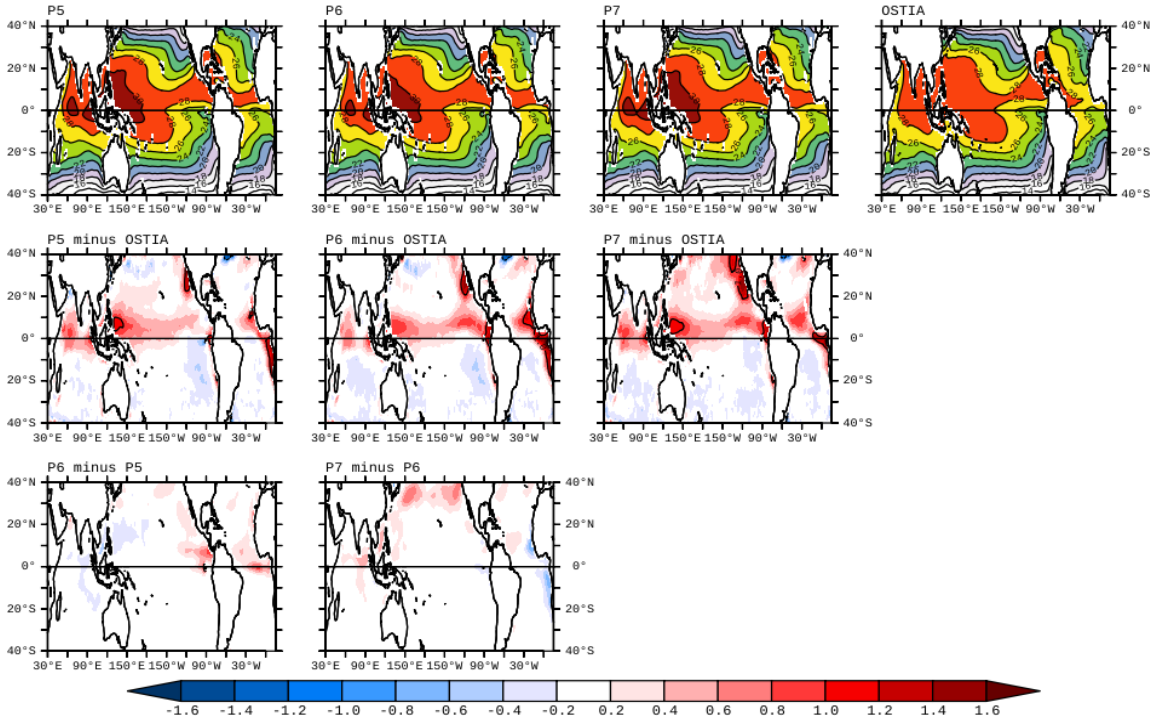

Figure 9.16: Week 3&4 average forecast of SST averaged for May, June, July (MJJ) in P5, P6, P7 and corresponding SST in OSTIA (top). The forecast biases in individual prototypes are shown in the middle row. The changes in SST forecast from P5 to P6 and P6 to P7 are shown in the last row.

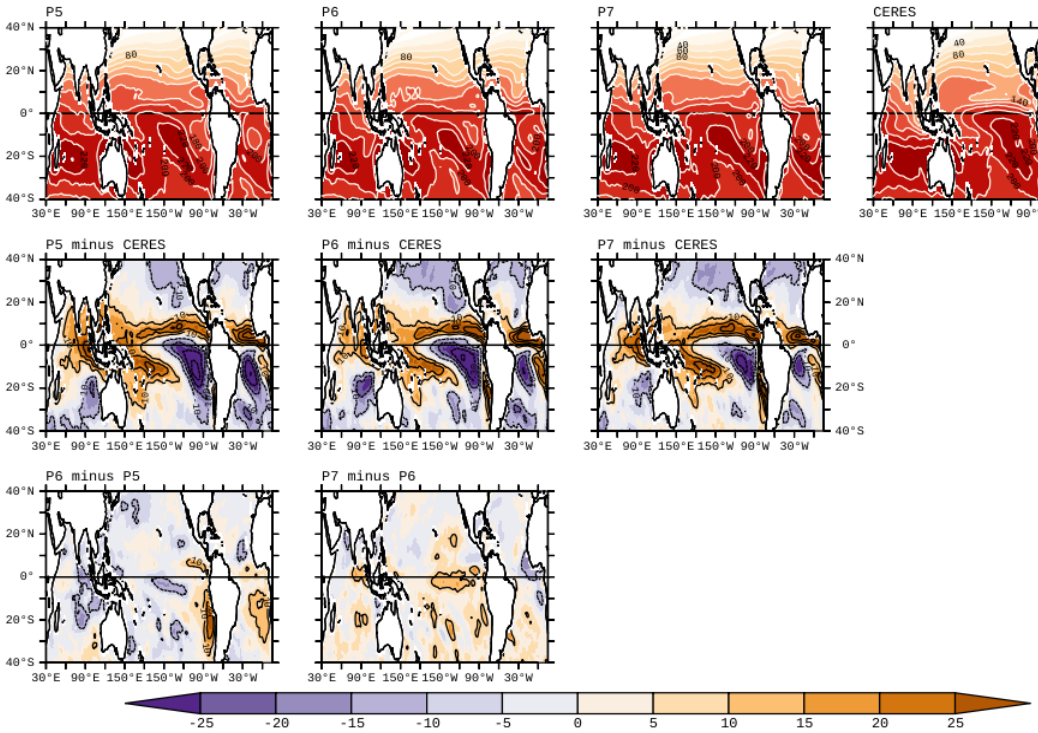

Figure 9.17: Week 3&4 average forecast of the net surface radiative flux (Watt/m^2) averaged for OND in P5, P6, P7 and corresponding observation from CERES-SYN1° (top). The forecast biases in individual prototypes are shown in the middle row. The changes in forecast of net radiative flux from P5 to P6 and P6 to P7 are shown in the bottom row.

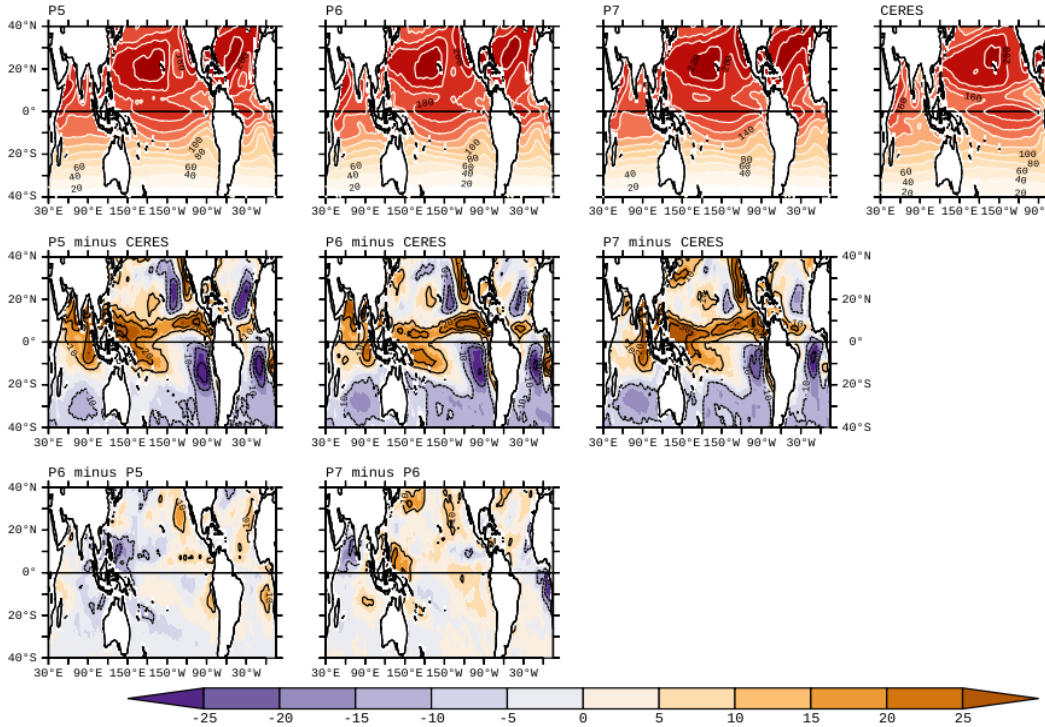

Figure 9.18: Week 3&4 average forecast of the net surface radiative flux (Watt/m^2) averaged for MJJ in P5, P6, P7 and corresponding observation from CERES-SYN1° (top). The forecast biases in individual prototypes are shown in the middle row. The changes in forecast of net radiative flux from P5 to P6 and P6 to P7 are shown in the bottom row.

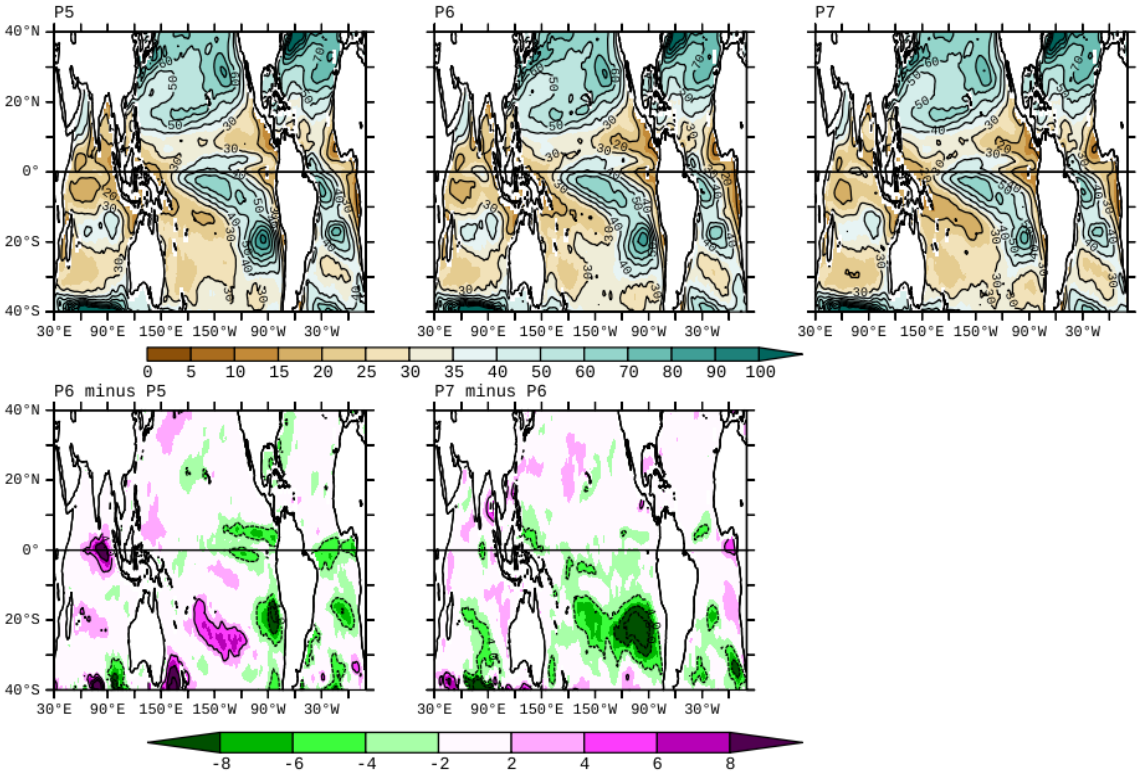

Figure 9.19: Week 3&4 average forecast of MLD averaged for OND in P5, P6, P7 (top). The changes in MLD forecast from P5 to P6 and P6 to P7 are shown in the bottom row.

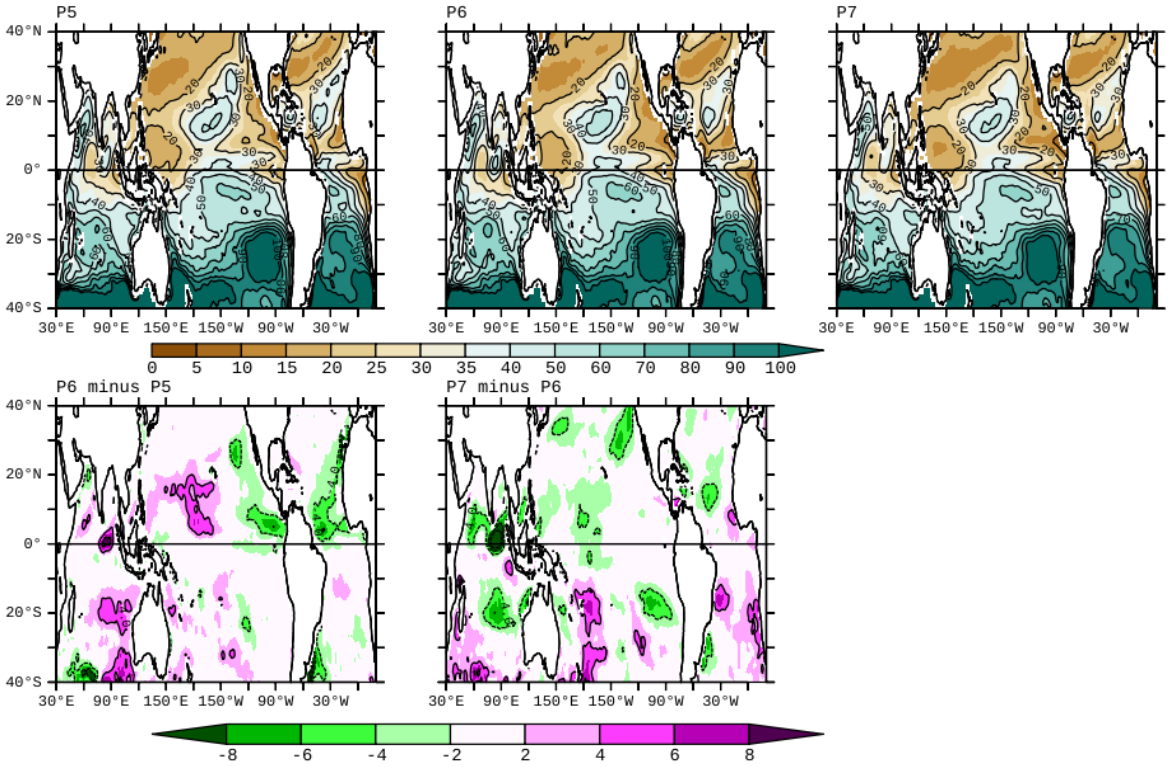

Figure 9.20: Week 3&4 average forecast of MLD averaged for MJJ in P5, P6, P7 (top). The changes in MLD forecast from P5 to P6 and P6 to P7 are shown in the bottom row.

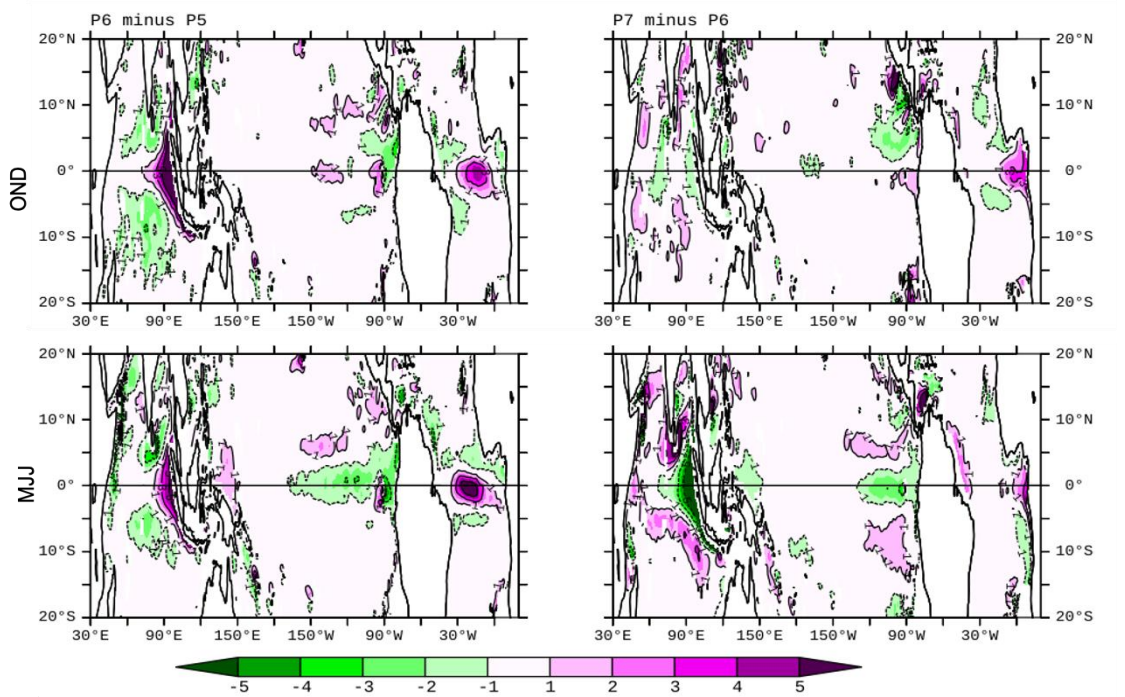

Figure 9.21: Changes in week 3&4 average forecast of Thermocline depth (D20) from P5 to P6 (left) and P6 to P7 (right) averaged for OND (top) and MJJ (bottom).

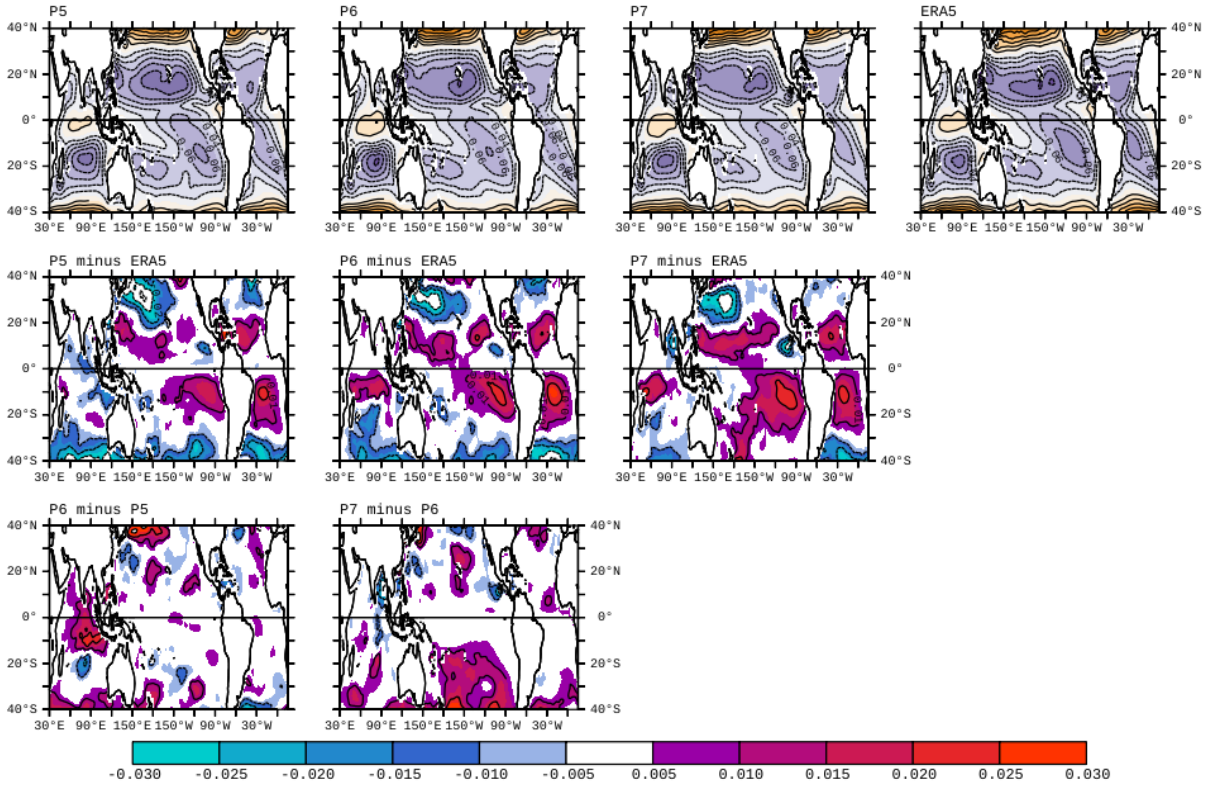

Figure 9.22: Week 3&4 average forecast of zonal wind stress (τ_x) averaged for OND in P5, P6, P7 and the corresponding observation from ERA5 (top). Forecast biases in individual prototypes are shown in the middle row. The changes in zonal wind stress forecast from P5 to P6 and P6 to P7 are shown in the bottom row.

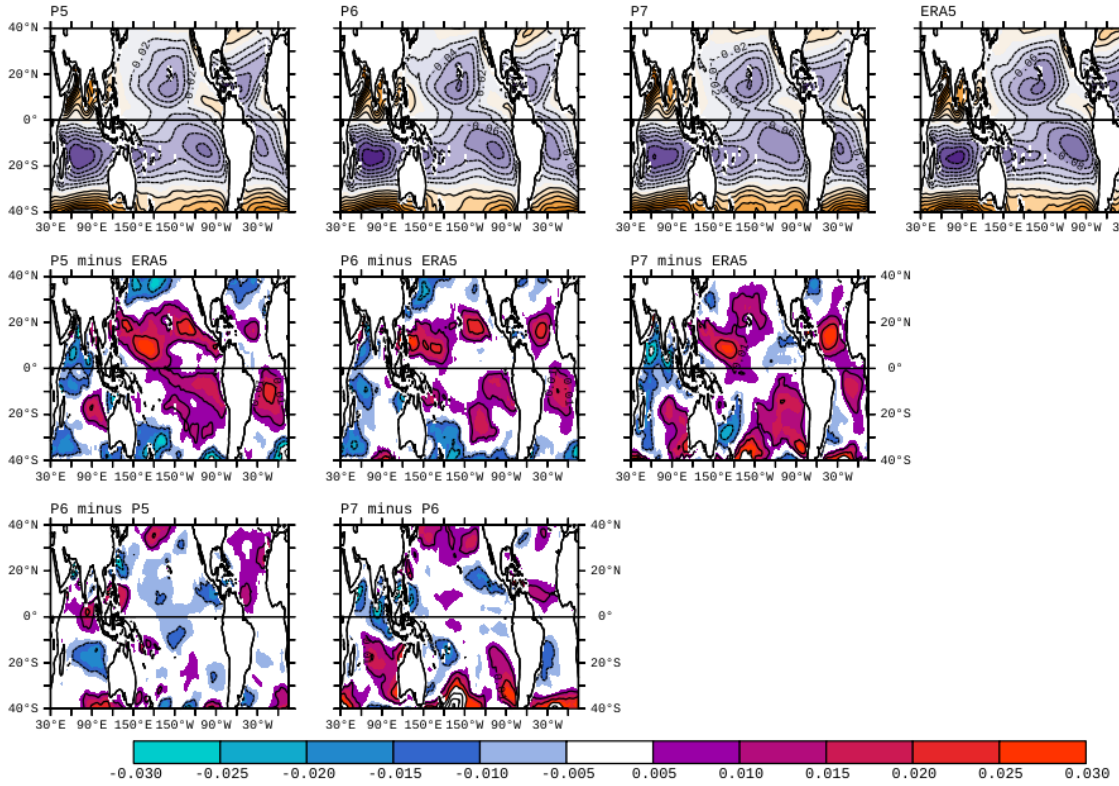

Figure 9.23: Week 3&4 average forecast of zonal wind stress (τ_x) averaged for MJJ in P5, P6, P7 and the corresponding observation from ERA5 (top). Forecast biases in individual prototypes are shown in the middle row. The changes in zonal wind stress forecast from P5 to P6 and P6 to P7 are shown in the bottom row.

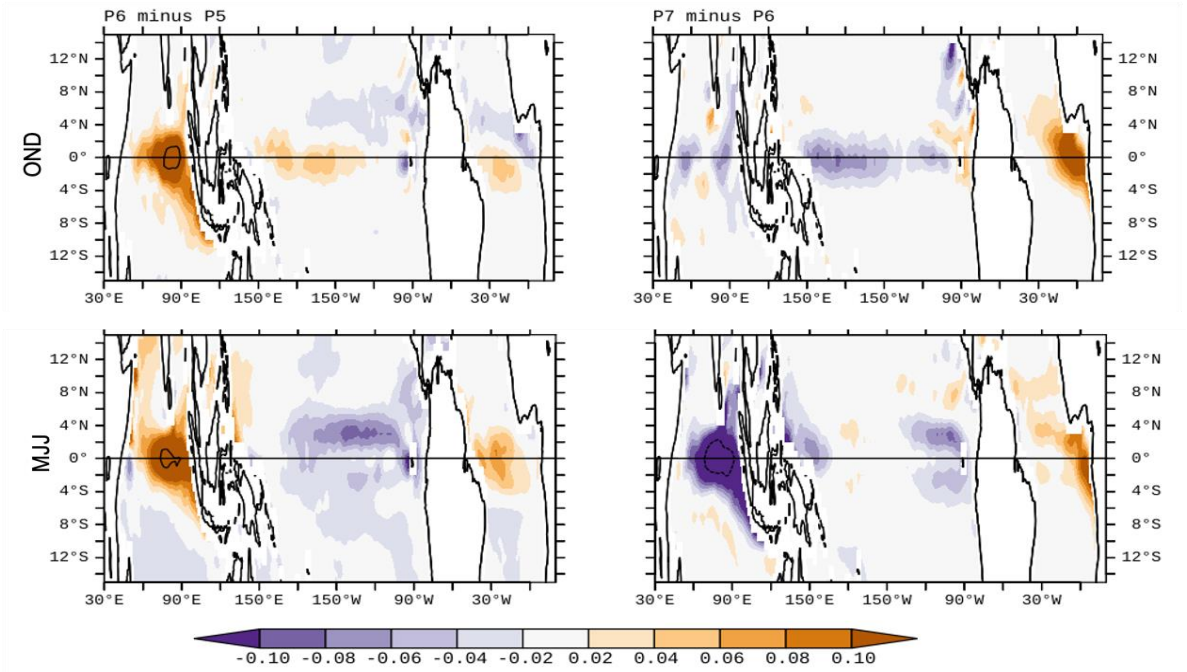

Figure 9.24: Changes in week 3&4 forecast of surface currents from P5 to P6 (left) and P6 to P7 (right) averaged for the OND (top) and MJJ (bottom).

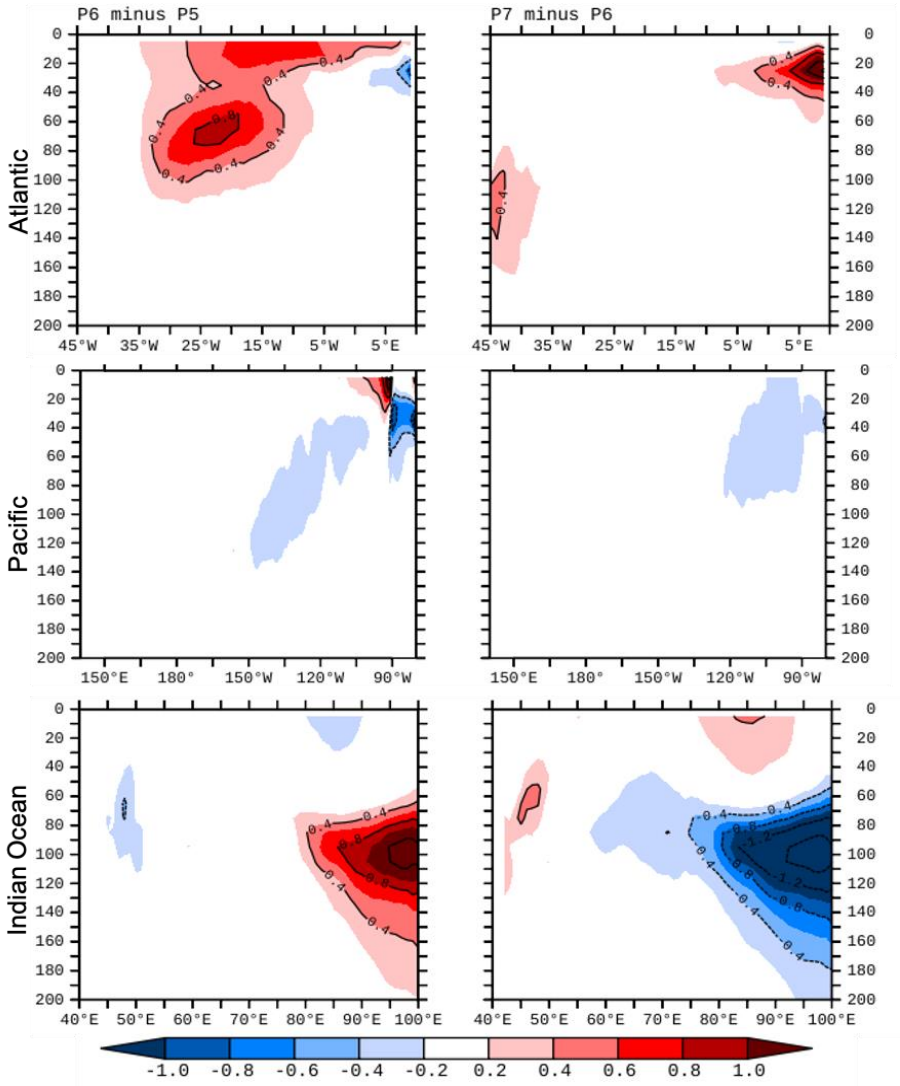

Figure 9.25: Changes in week 3&4 forecast of MJJ averaged equatorial (2S:2N) sections of temperature in the Atlantic (top), the Pacific (middle) and Indian Ocean (bottom) from P5 to P6 (left) and P6 to P7 (right).

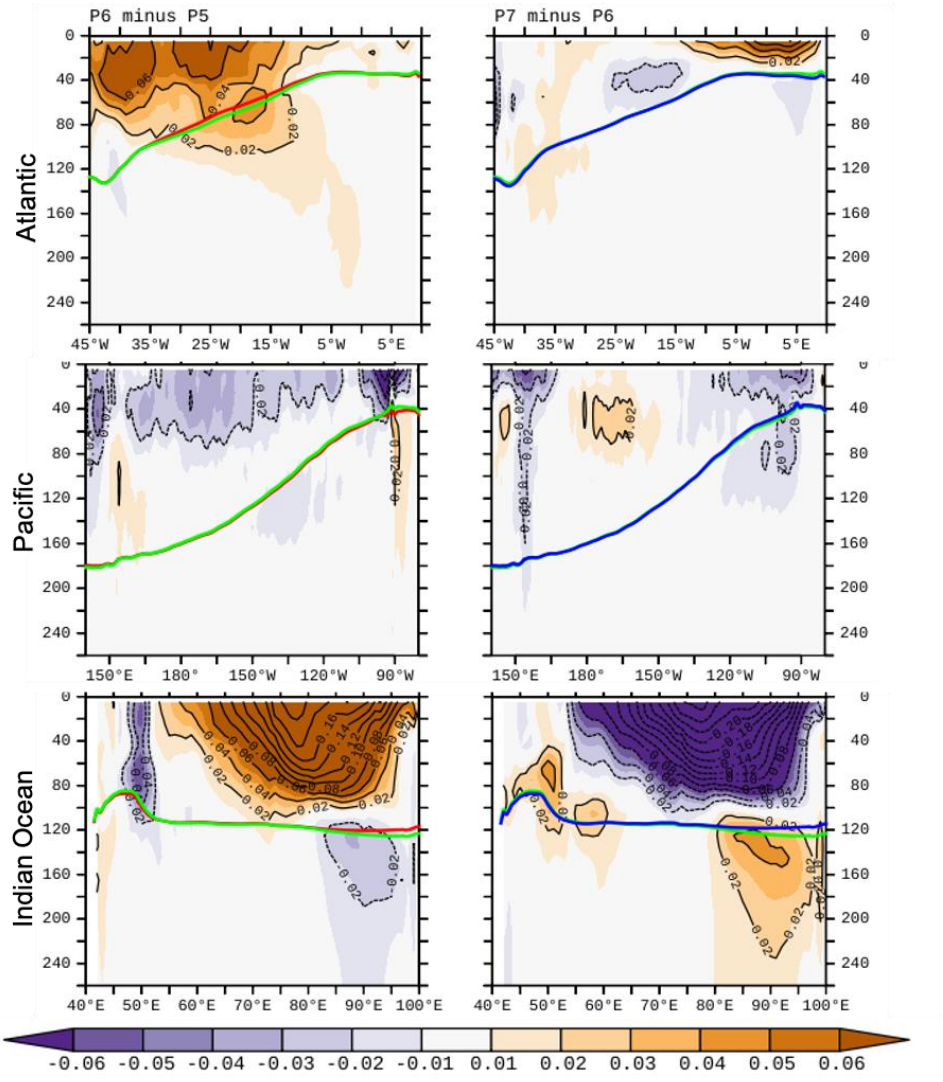

Figure 9.26: Changes in week 3&4 forecast of MJJ averaged equatorial (2°S:2°N) sections of zonal current in the Atlantic (top), the Pacific (middle) and Indian Ocean (bottom) from P5 to P6 (left) and P6 to P7 (right). Overlaid by the week 3&4 forecast of MJJ averaged D20 from P5 (red), P6 (green), and P7 (blue).

P6 Figures

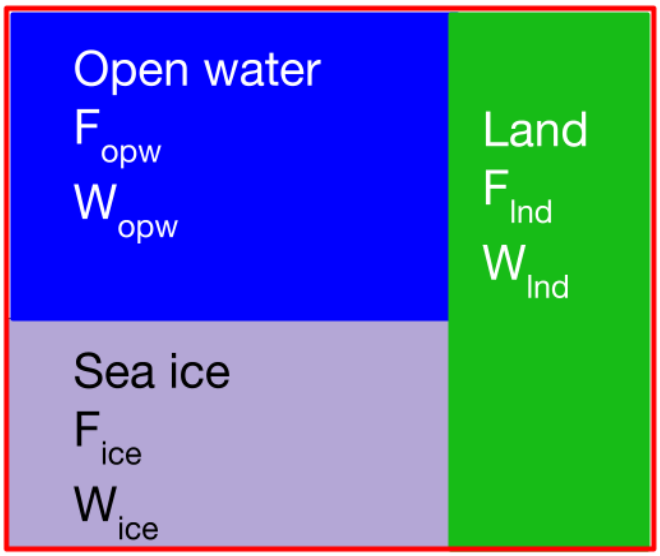

Figure 10.1: a schematic of an atmospheric grid cell (outlined in red) consisting of land, open water and sea-ice.

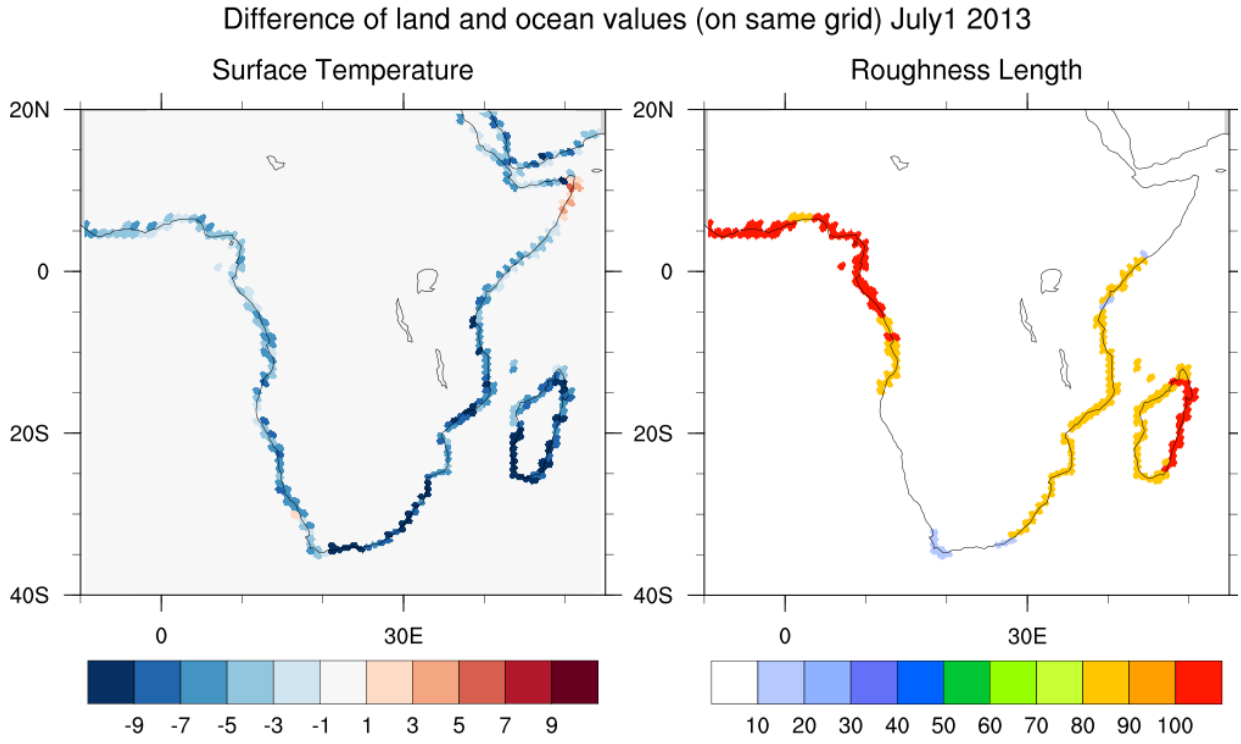

Figure 10.2: Difference between land and ocean values (on the same grid) of temperature (left, in K) and roughness length (right, in cm).

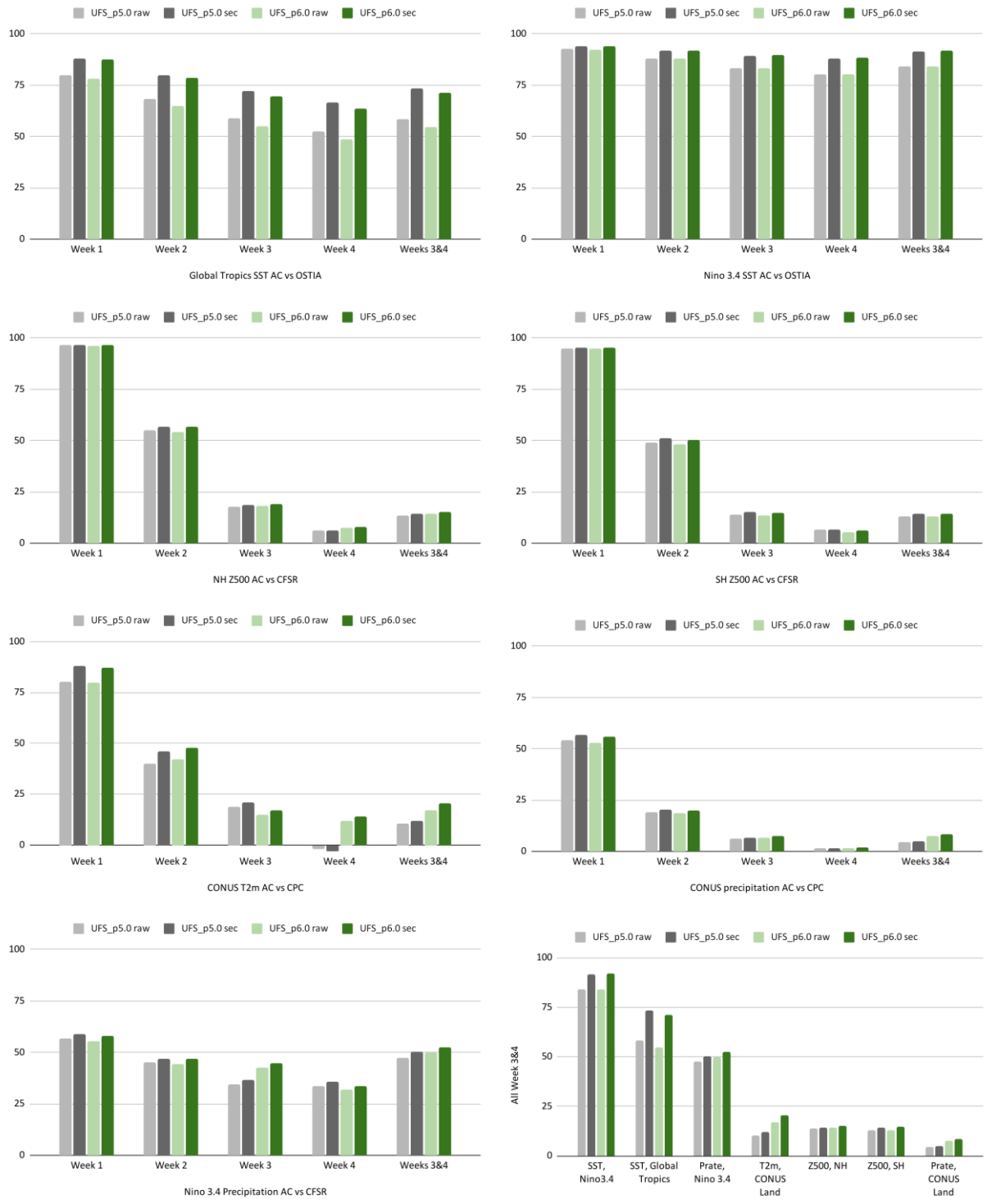

Figure 10.3: Anomaly correlation scores by forecast week of P5 (RAW-pale gray, SEC-dark gray) and P6 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of the anomaly correlation scores for Weeks 3&4.

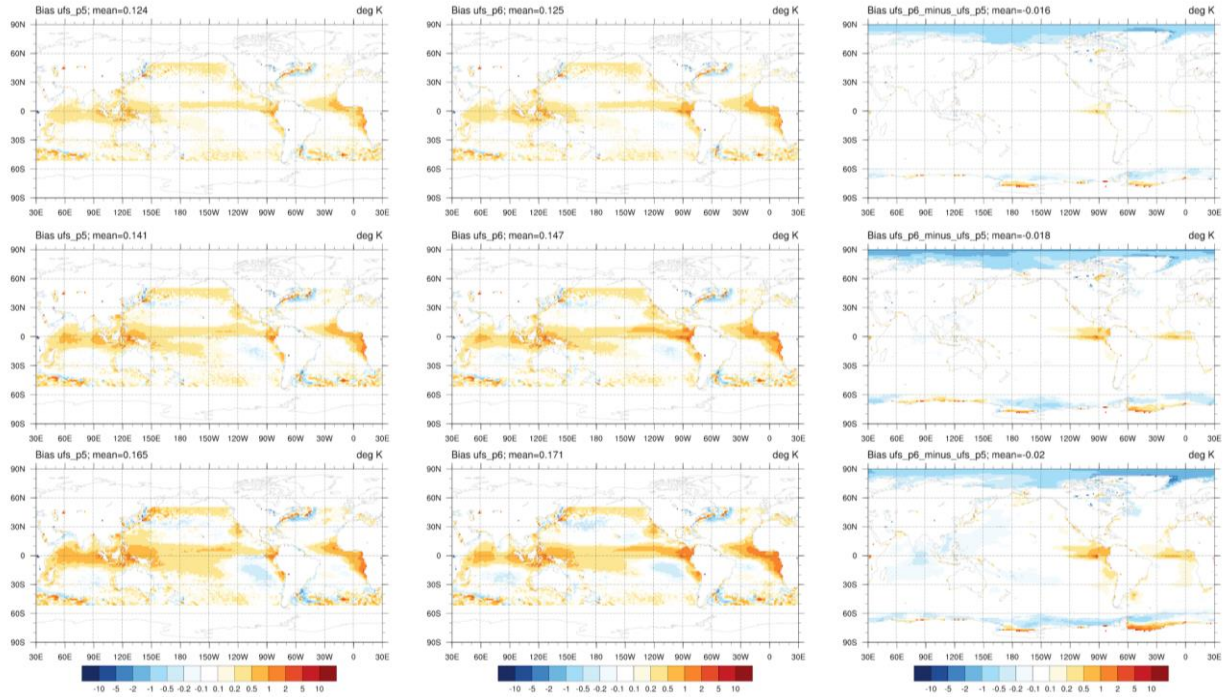

Figure 10.4: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean SST bias wrt OSTIA from 50°S to 50°N in P5 (left column) and P6 (middle column); Difference (P6 minus P5) in surface temperatures over ocean grid points, including sea ice (right column).

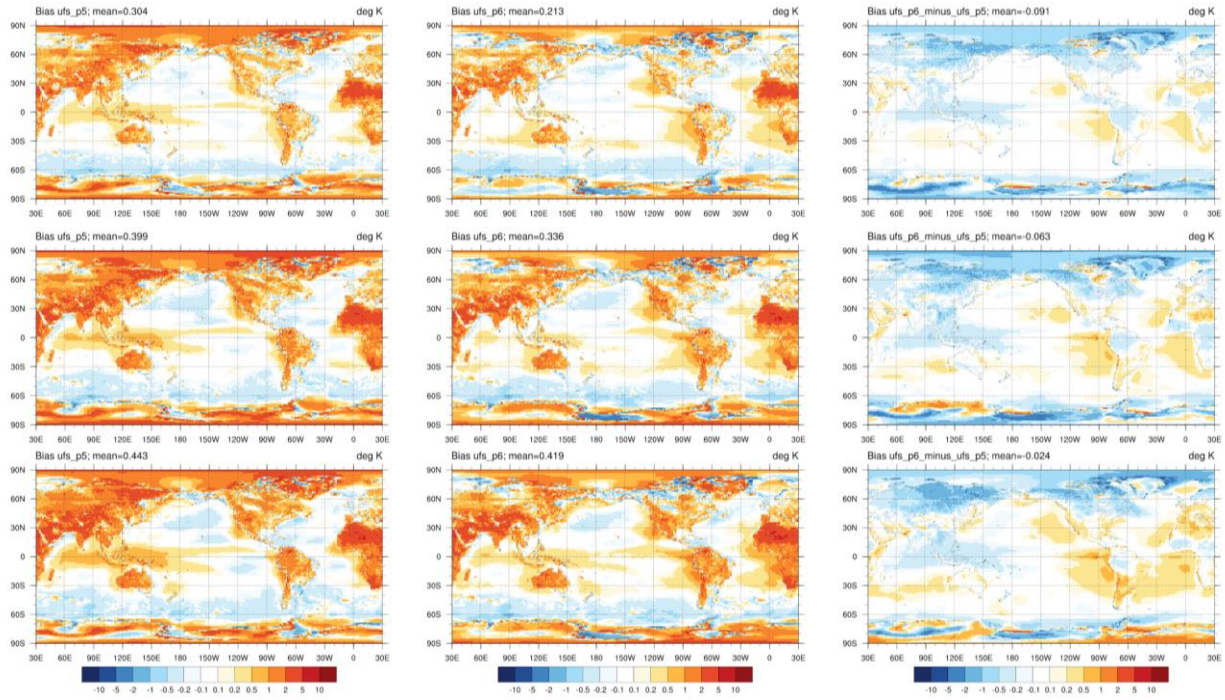

Figure 10.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in P5 (left column) and P6 (middle column); Difference (P6minus P5) in 2-m air temperature (right column).

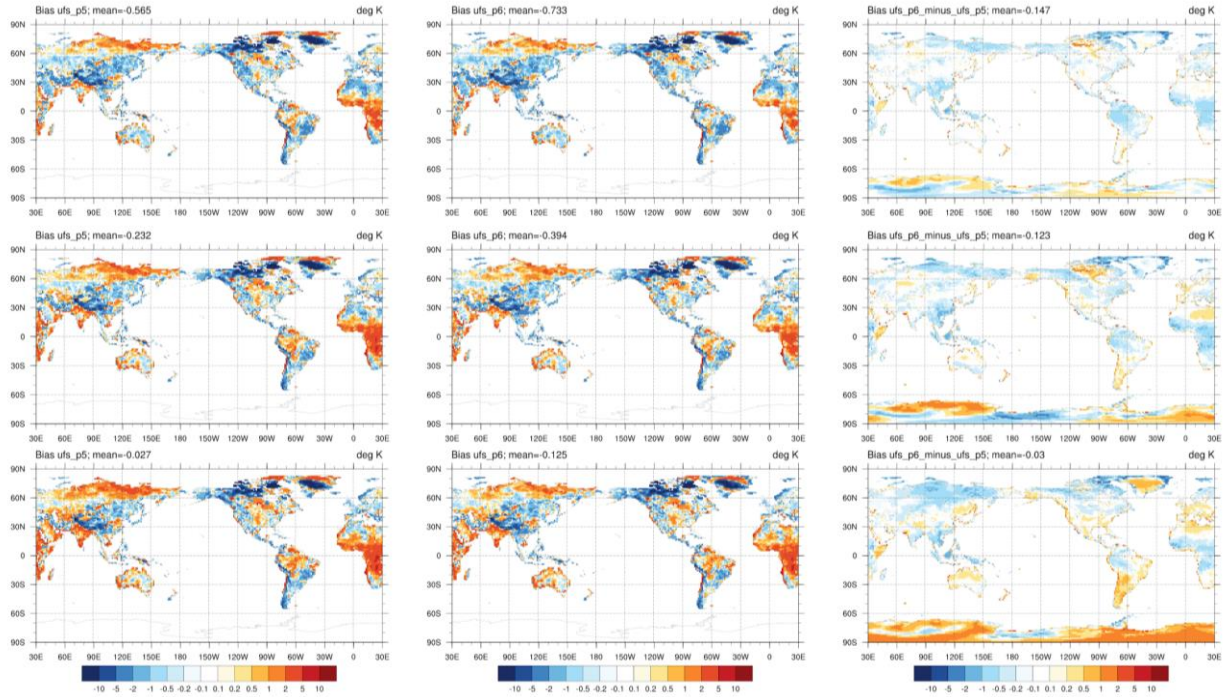

Figure 10.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in P5(left column) and P6 (middle column); Difference (P6 minus P5) in 2-m daily maximum air temperature (right column).

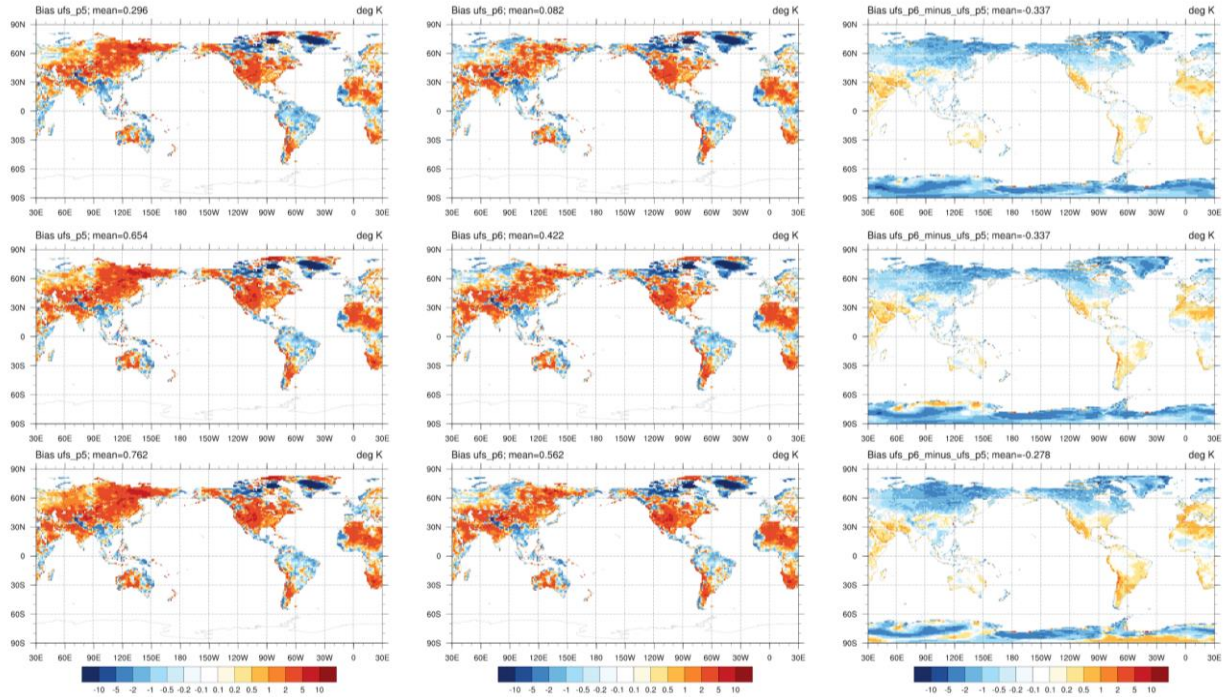

Figure 10.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in P5 (left column) and P6 (middle column); Difference (P6 minus P5) in 2-m daily minimum air temperature (right column).

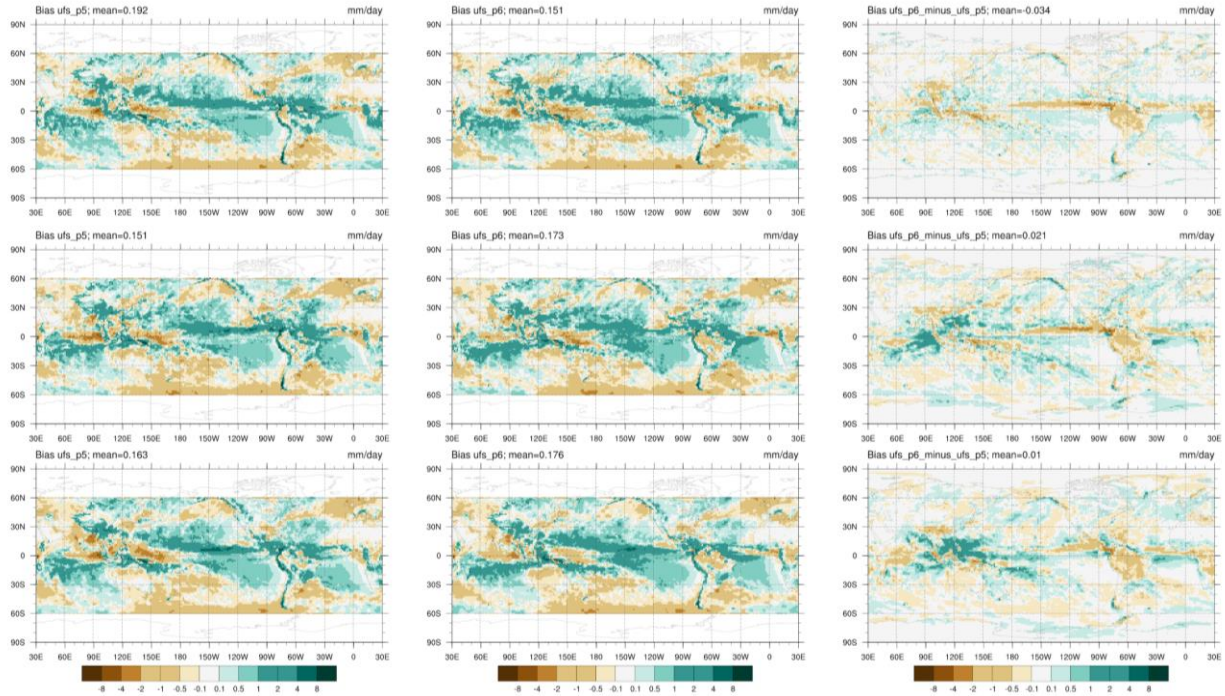

Figure 10.8: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in P5 (left column) and P6 (middle column); Difference (P6 minus P5) in precipitation rate (right column

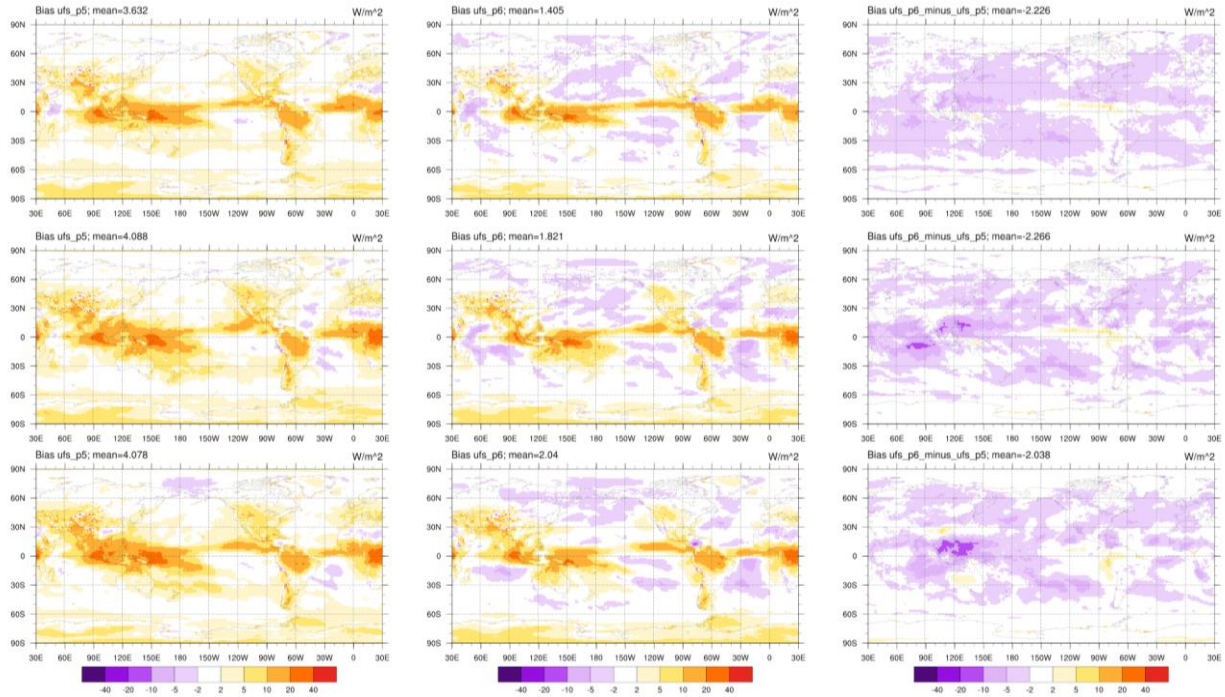

Figure 10.9: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in P5 (left column) and P6 (middle column); Difference (P6 minus P5) in OLR (right column).

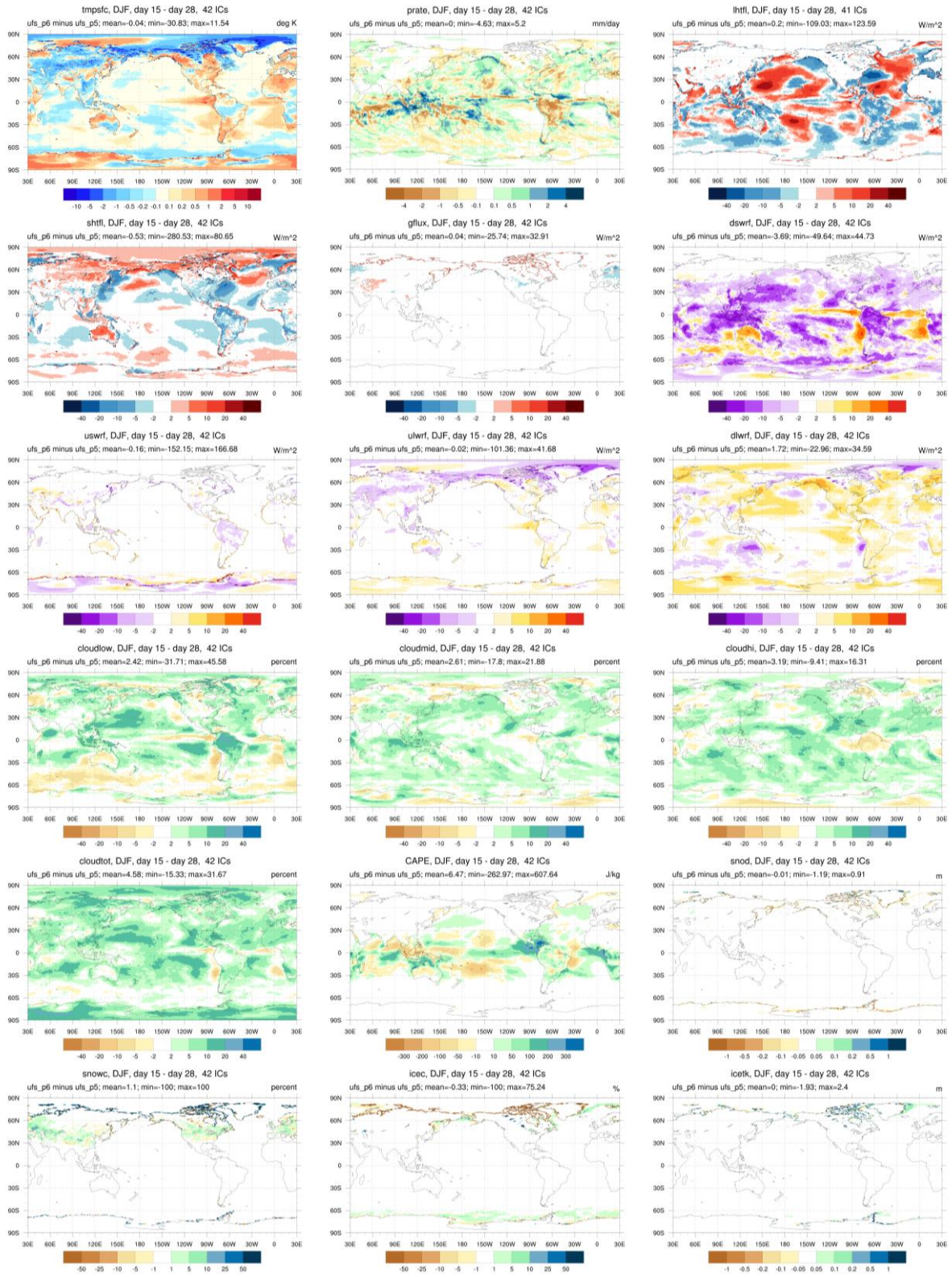

Figure 10.10: Week 3&4 average difference (P6 minus P5) for select forecast fields as indicated on subplot titles from forecasts initialized in DJF.

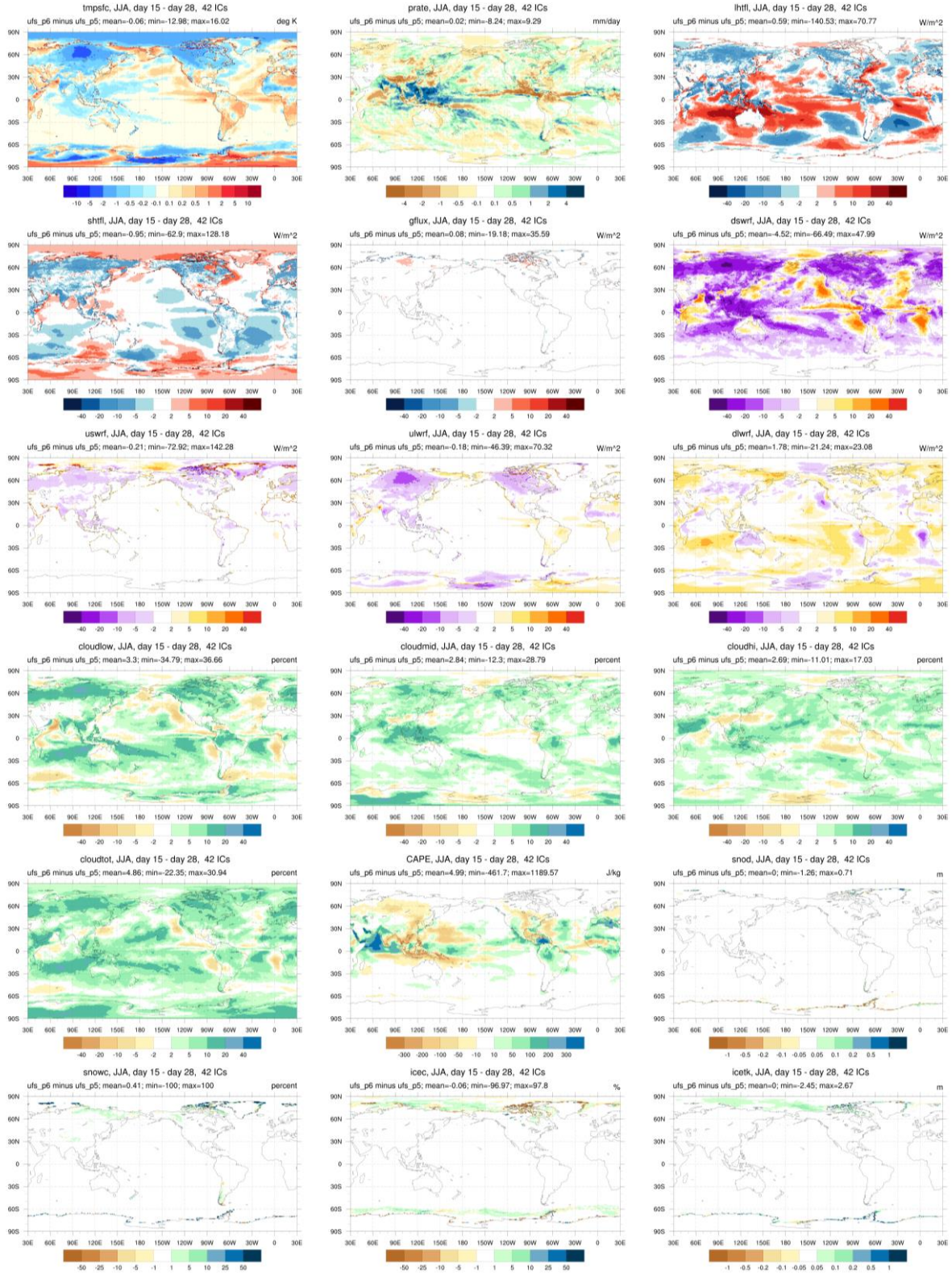

Figure 10.11: Week 3&4 average difference (P6 minus P5) for select forecast fields as indicated on subplot titles from forecasts initialized in JJA.

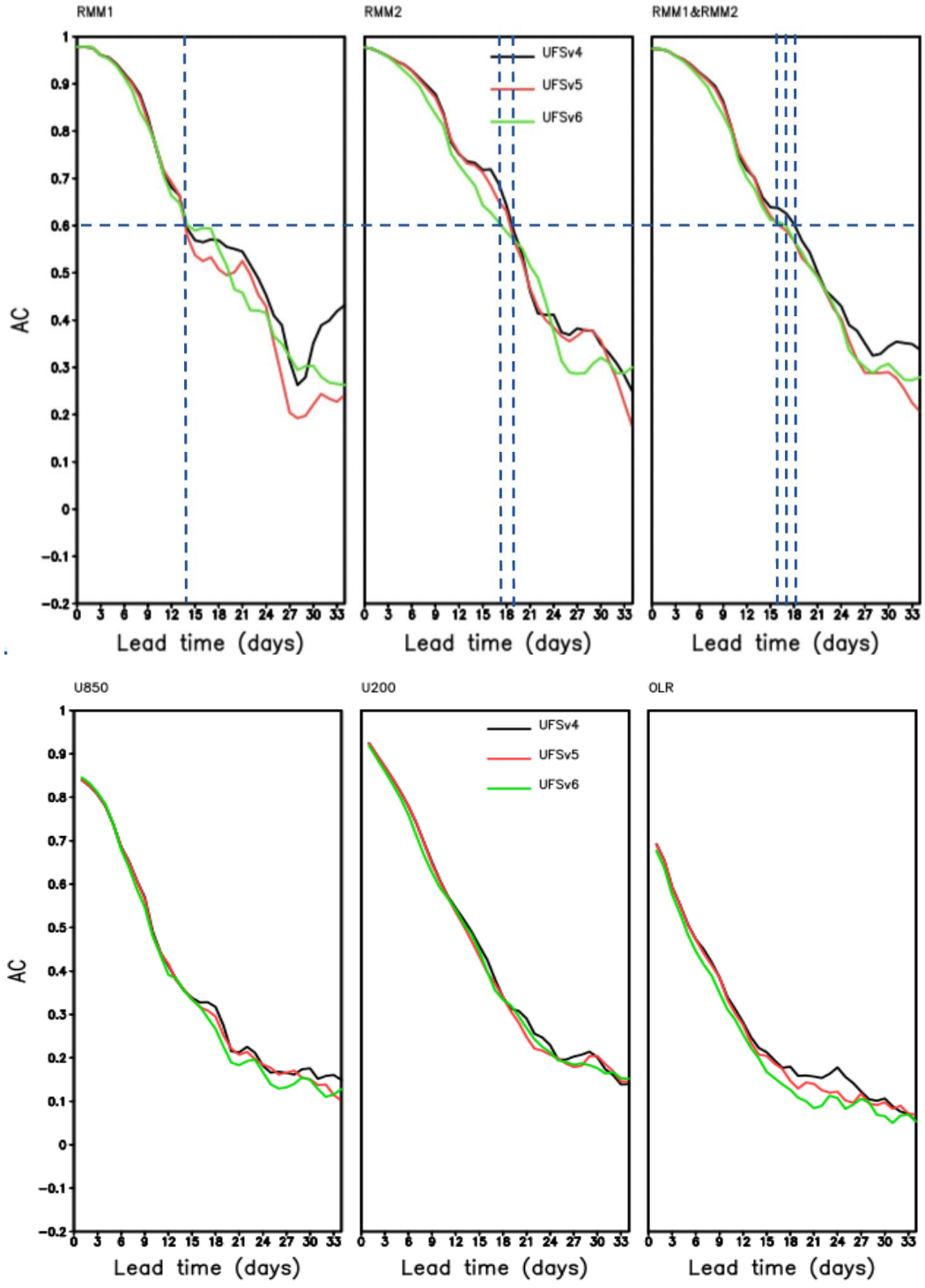

Figure 10.12: Top: AC of MJO RMM1 (left), RMM2 (middle) and combined RMM1&RMM2 (right). Bottom: AC of equatorial filtered zonal wind at 850mb (left) and 200mb (middle), and equatorial filtered OLR (right) in P4 (black line), P5 (red line), and P6 (green line).

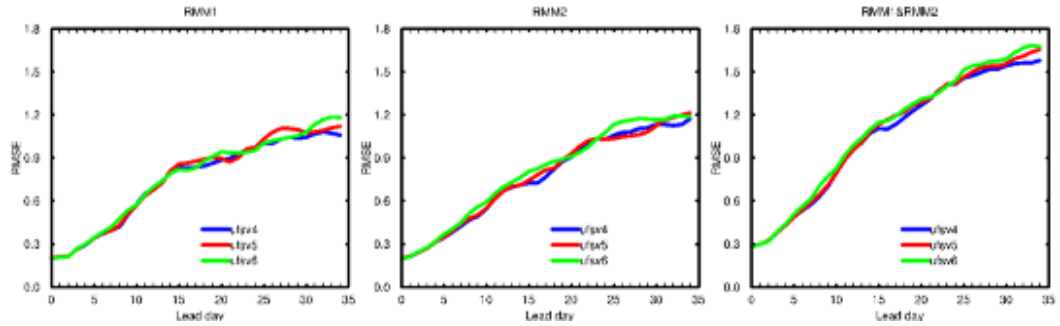

Figure 10.13: RMSE of MJO RMM1, RMM2 and RMM1&2 for P4 (blue), P5 (red) and P6 (green).

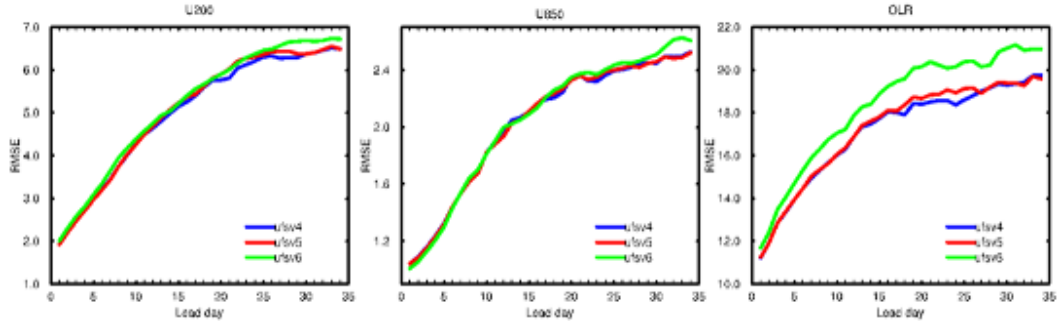

Figure 10.14: RMSE of MJO components (U200, U850 and OLR) for P4 (blue), P5 (red), and P6 (green).

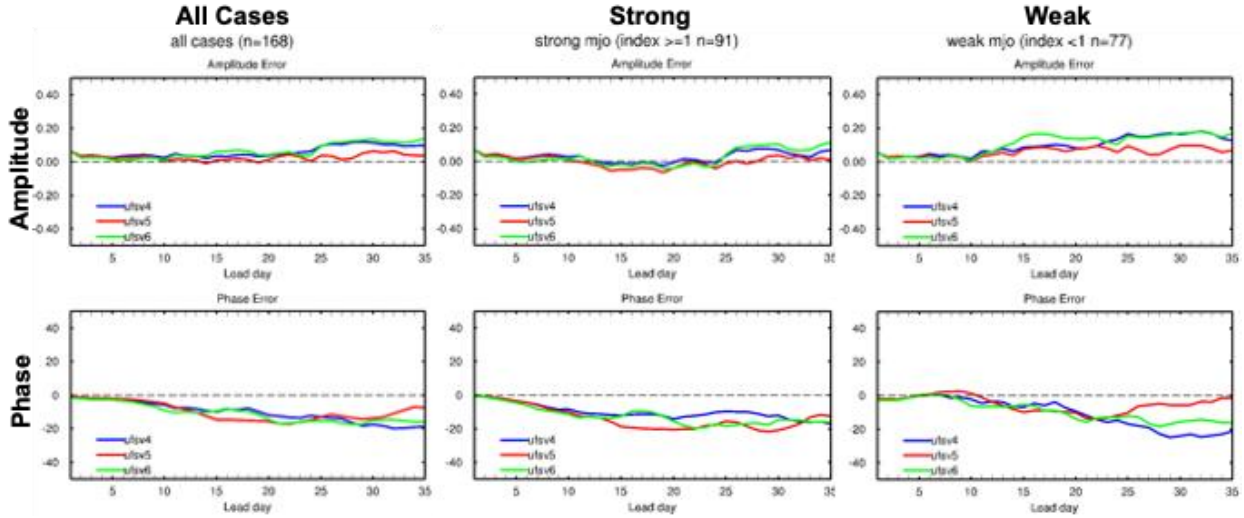

Figure 10.15: MJO amplitude and phase error for P4 (blue), P5 (red), and P6 (green) separated into strong and weak MJOs.

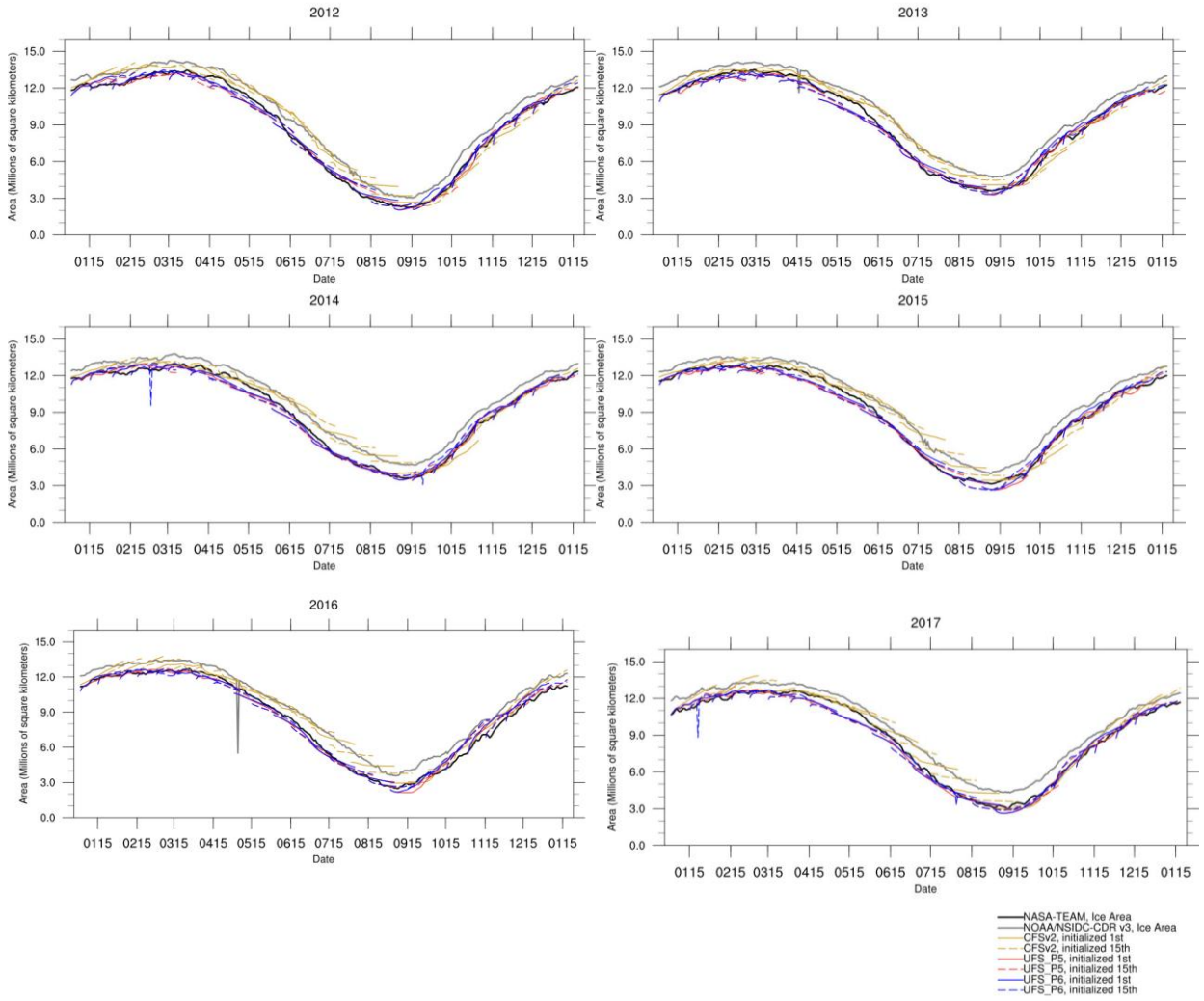

Figure 10.16: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow), P5 (red), and P6 (blue). Thick lines correspond to the observational estimate derived from ice concentrations from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date.

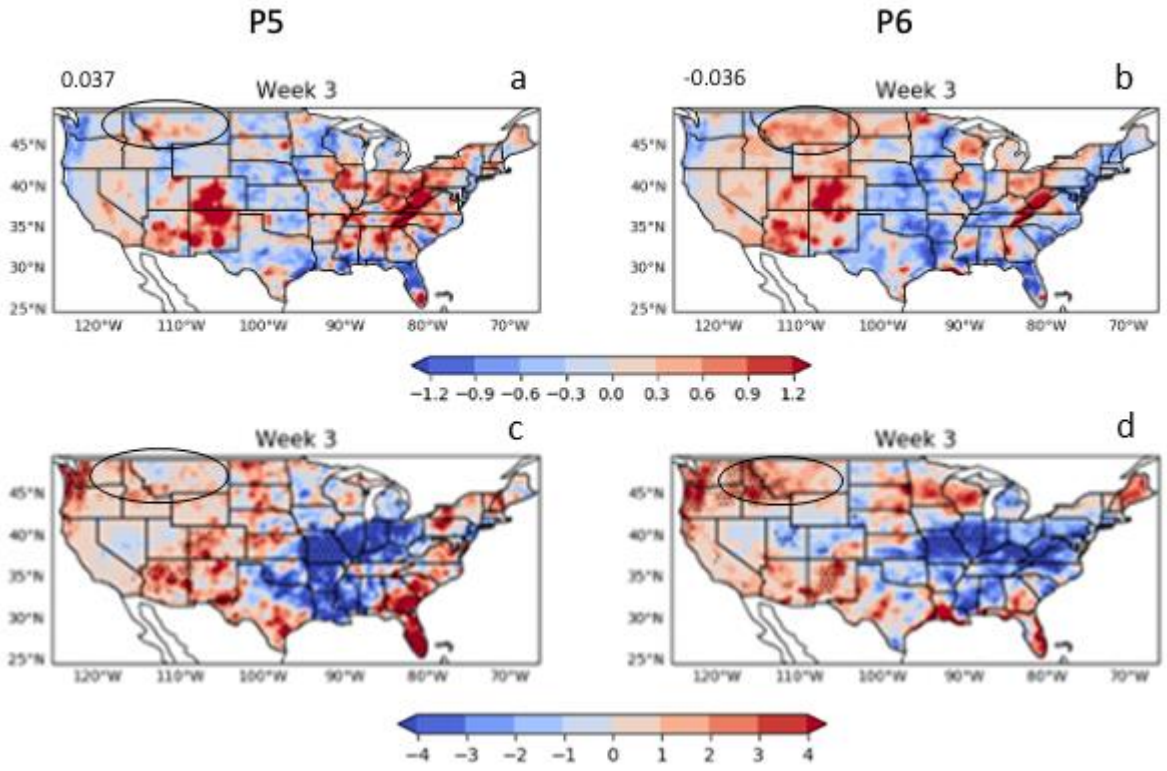

Figure 10.17: Boreal summer week 3 precipitation bias (a),(b) and regression of tropical Pacific SST bias onto precipitation bias (c),(d). The bias is defined as the difference between all forecasts initialized on 1st and 15th of each month during June-September and observations for the same dates (CPCU for precipitation and OSST for SST). [Units: mm/day and mm/day per unit standard deviation of SST time series]. Numbers in (a) and (b) represent the mean bias over CONUS. The regions where the regression coefficient is statistically significant above 5% are stippled. Highlighted regions are discussed in the text.

P7 Figures

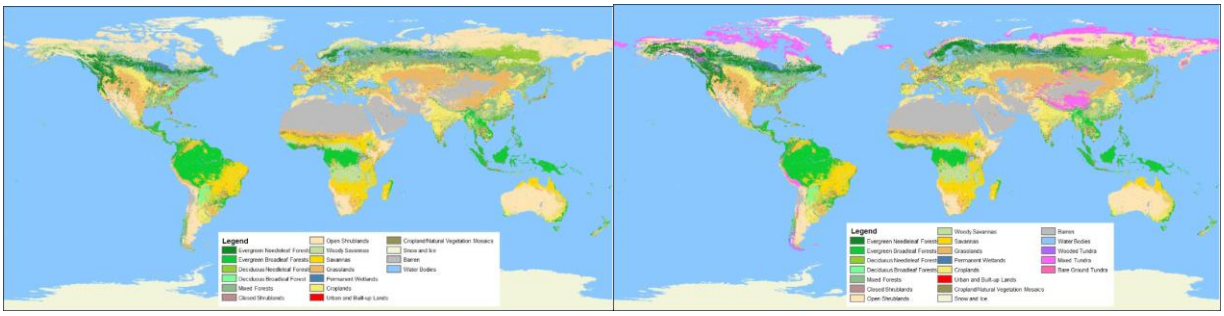

Figure 11.1: New climatological surface type product based on 2012-2019 NESDIS VIIRS surface type product (left) and final surface type product with additional tundra classes used in Prototype 7 (right).

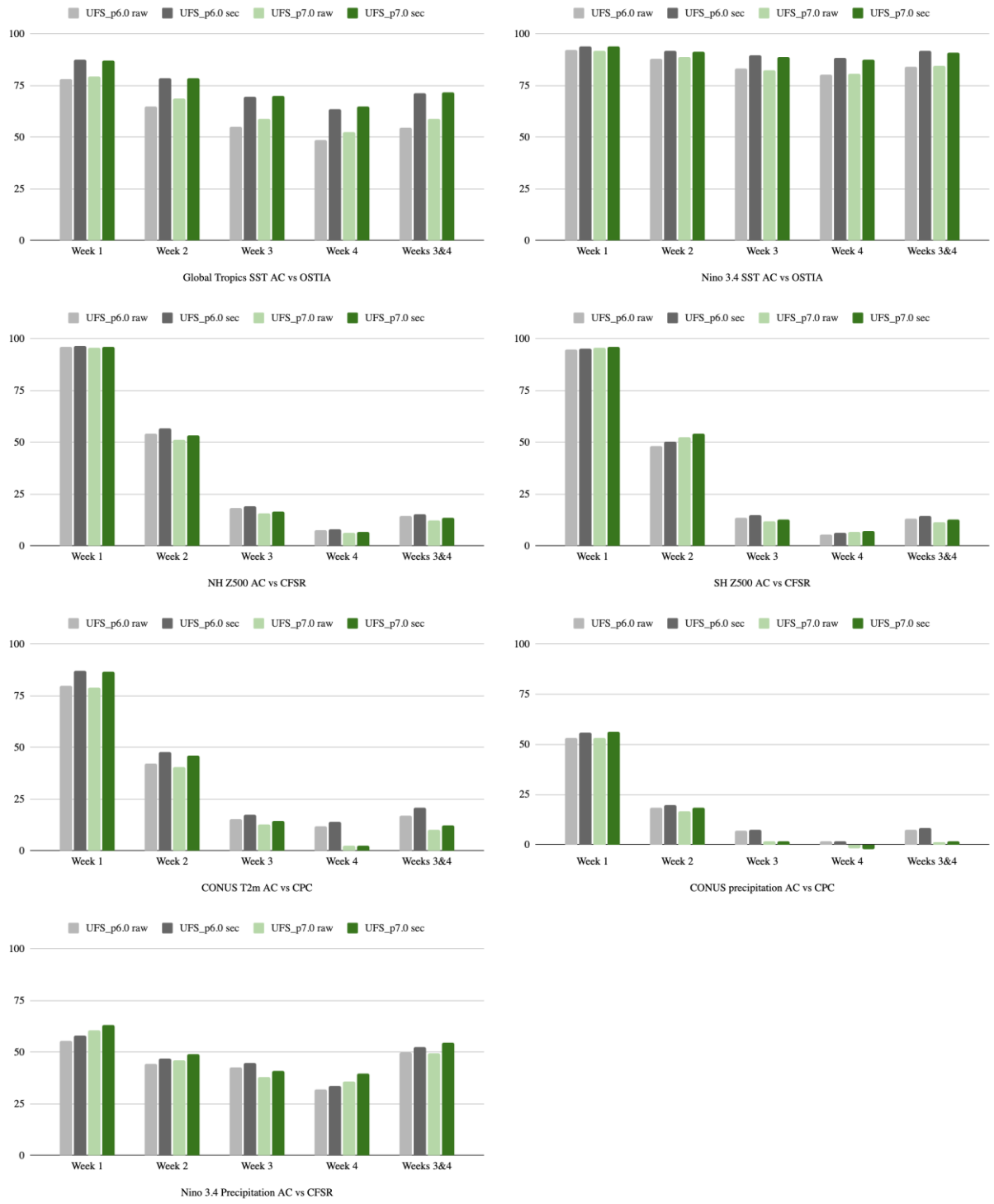

Figure 11.2: Anomaly correlation scores by forecast week of P6 (RAW-pale gray, SEC-dark gray) and P7 (RAW-pale green, SEC-dark green) for a) Global Tropics SST wrt OSTIA, b) Nino 3.4 SST wrt OSTIA, c) NH Z500 wrt CFSR, d) SH Z500 wrt CFSR, e) Nino 3.4 precipitation wrt CFSR; and f) Summary of the anomaly correlation scores for Weeks 3&4. **Note that the SST AC calculations shown here for P7 here were made using the model's NSST instead of SST which is not strictly correct. For an AC comparison of true SST from the ocean model, please see Fig 9.13 (top panel).*

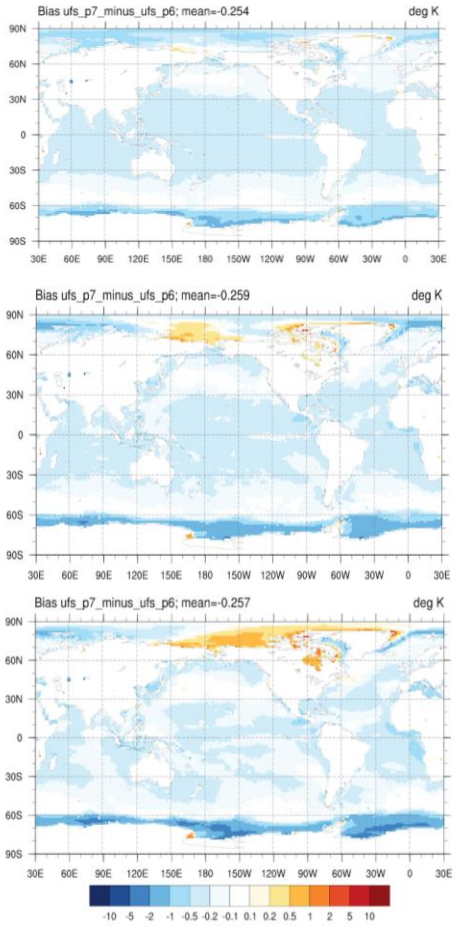

Figure 11.3: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Difference (P7 minus P6) in surface temperatures seen by the atmosphere over ocean grid points, including sea ice (left column).

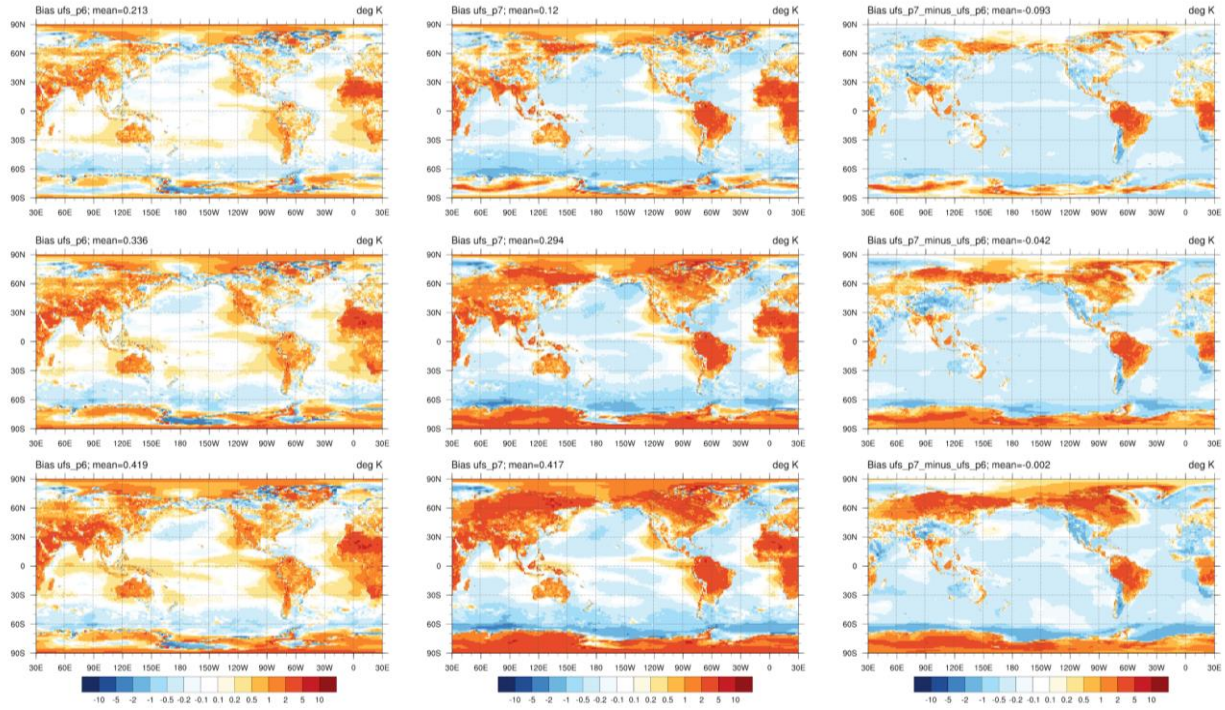

Figure 11.4a: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m air temperature bias wrt CFSR in P6 (left column) and P7 (middle column); Difference (P7 minus P6) in 2-m air temperature (right column).

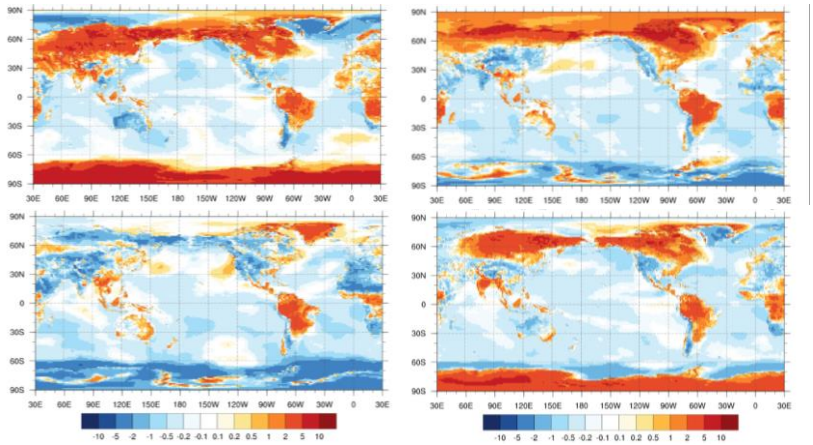

Figure 11.4b: Difference (P7 minus P6) of 2-m air temperature in Weeks 3&4 for DJF (top left), MAM (top right), JJA (bottom left), SON (bottom right).

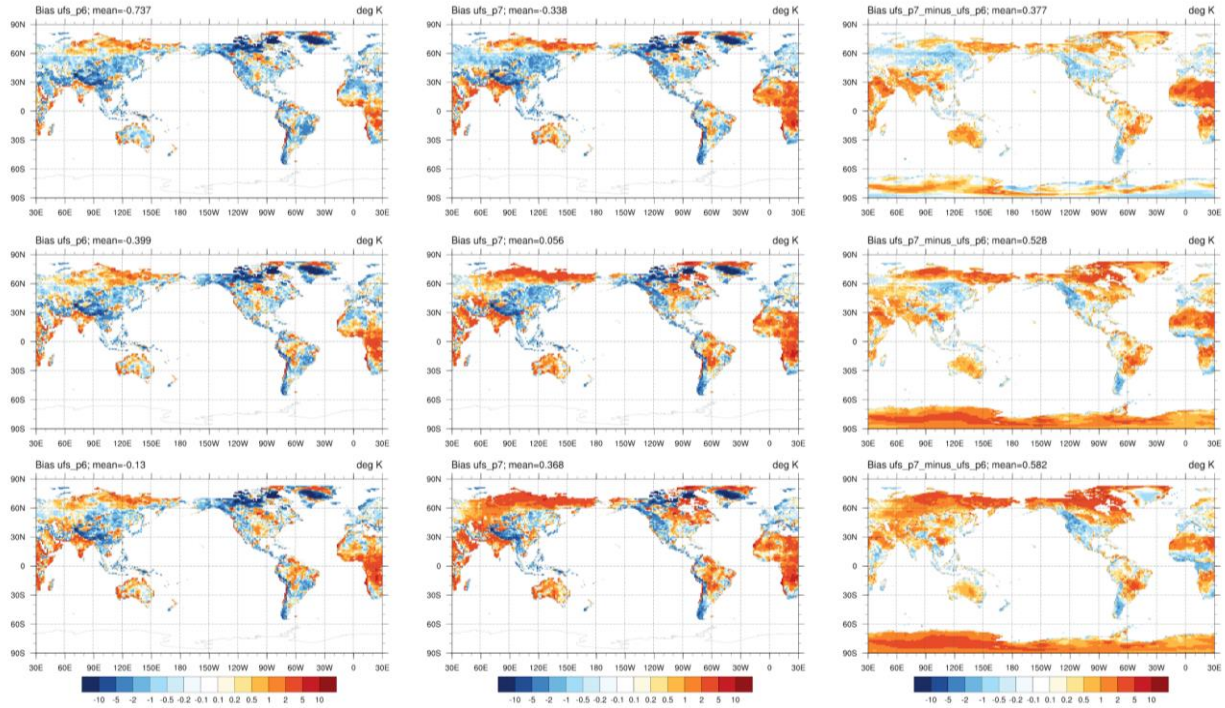

Figure 11.5: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily maximum air temperature bias wrt CPC in P6 (left column) and P7 (middle column); Difference (P7 minus P6) in 2-m daily maximum air temperature (right column).

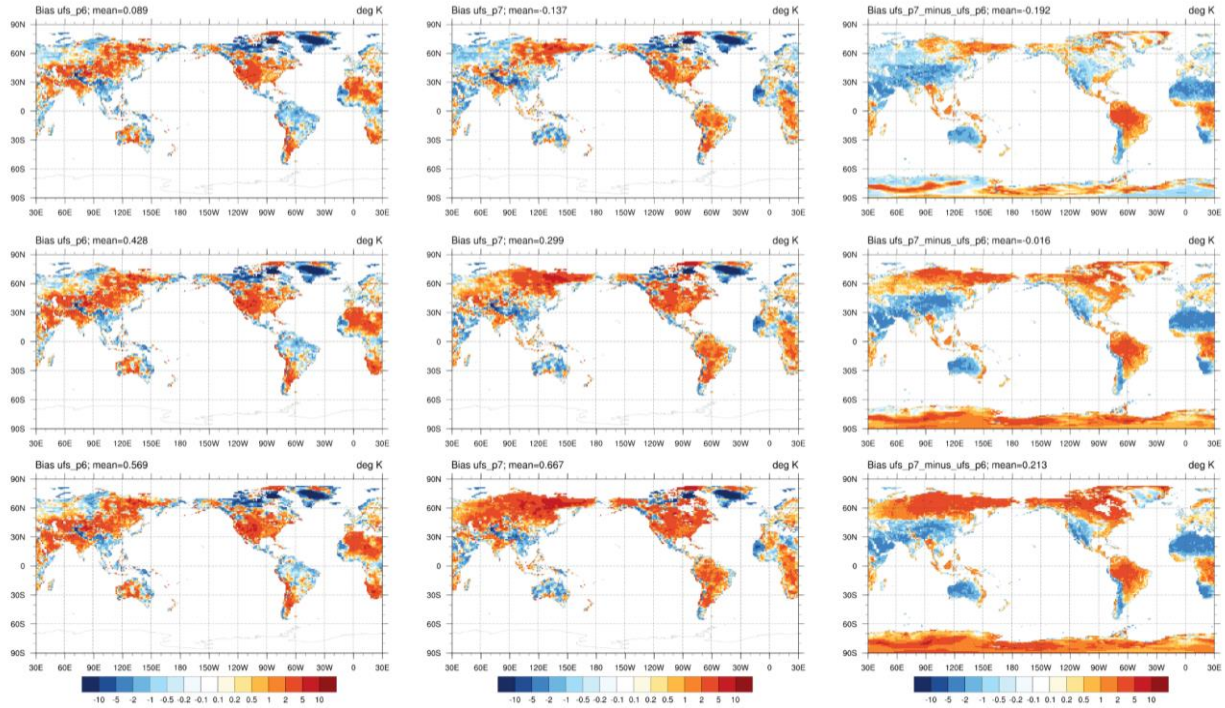

Figure 11.6: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean 2-m daily minimum air temperature bias wrt CPC in P6 (left column) and P7 (middle column); Difference (P7 minus P6) in 2-m daily minimum air temperature (right column).

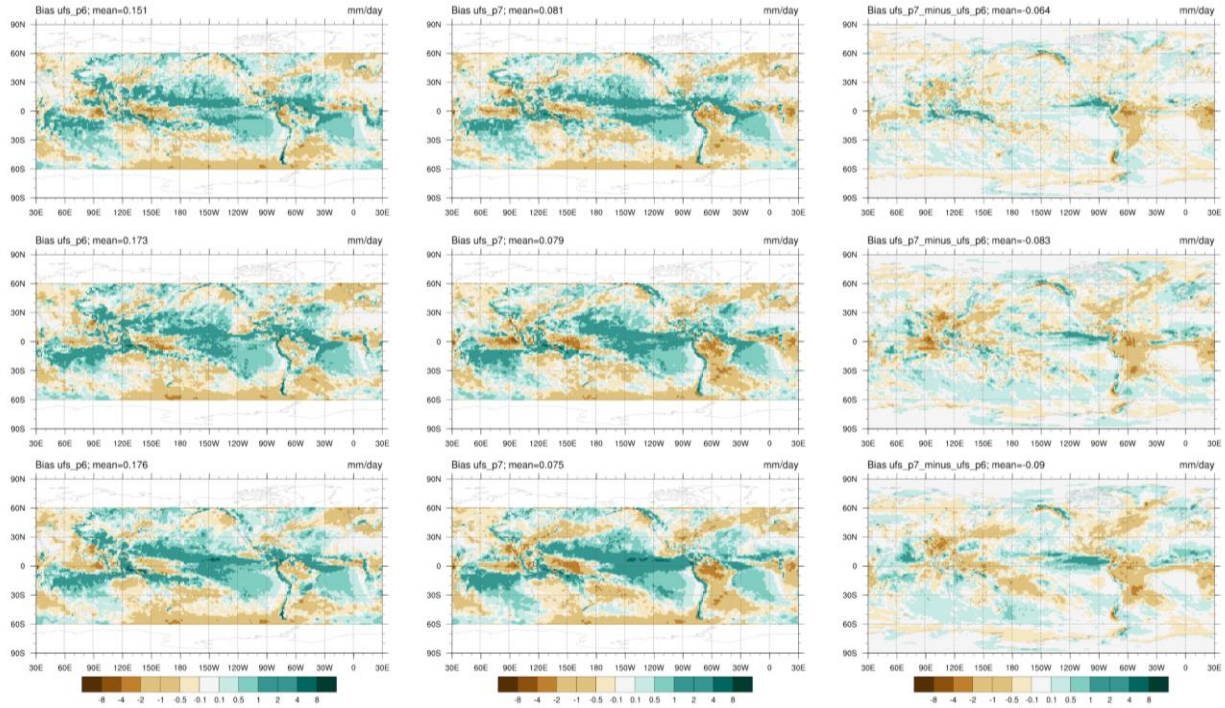

Figure 11.7: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean precipitation rate bias wrt IMERG in P6 (left column) and P7 (middle column); Difference (P7 minus P6) in precipitation rate (right column).

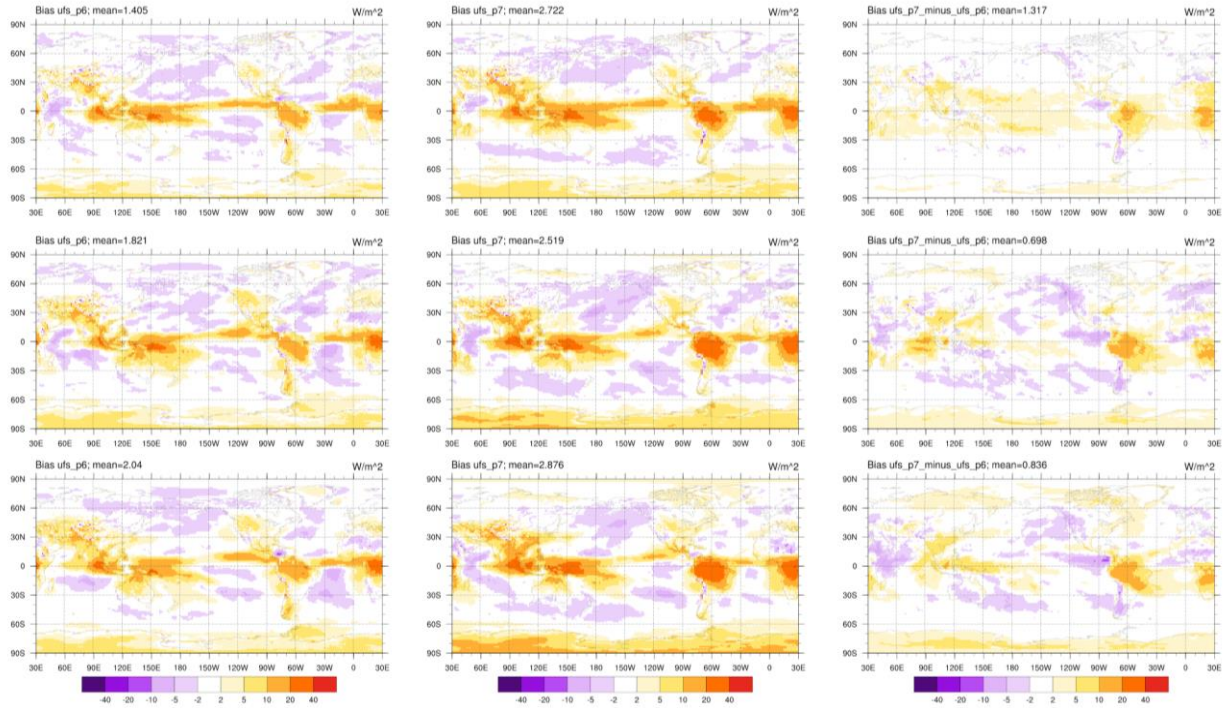

Figure 11.8: Week 1 (top row), Week 2 (middle row), and Weeks 3&4 (bottom row): Annual mean OLR bias wrt NOAA-CDR in P6 (left column) and P7 (right column); Difference (P6 minus P7) in OLR (left column).

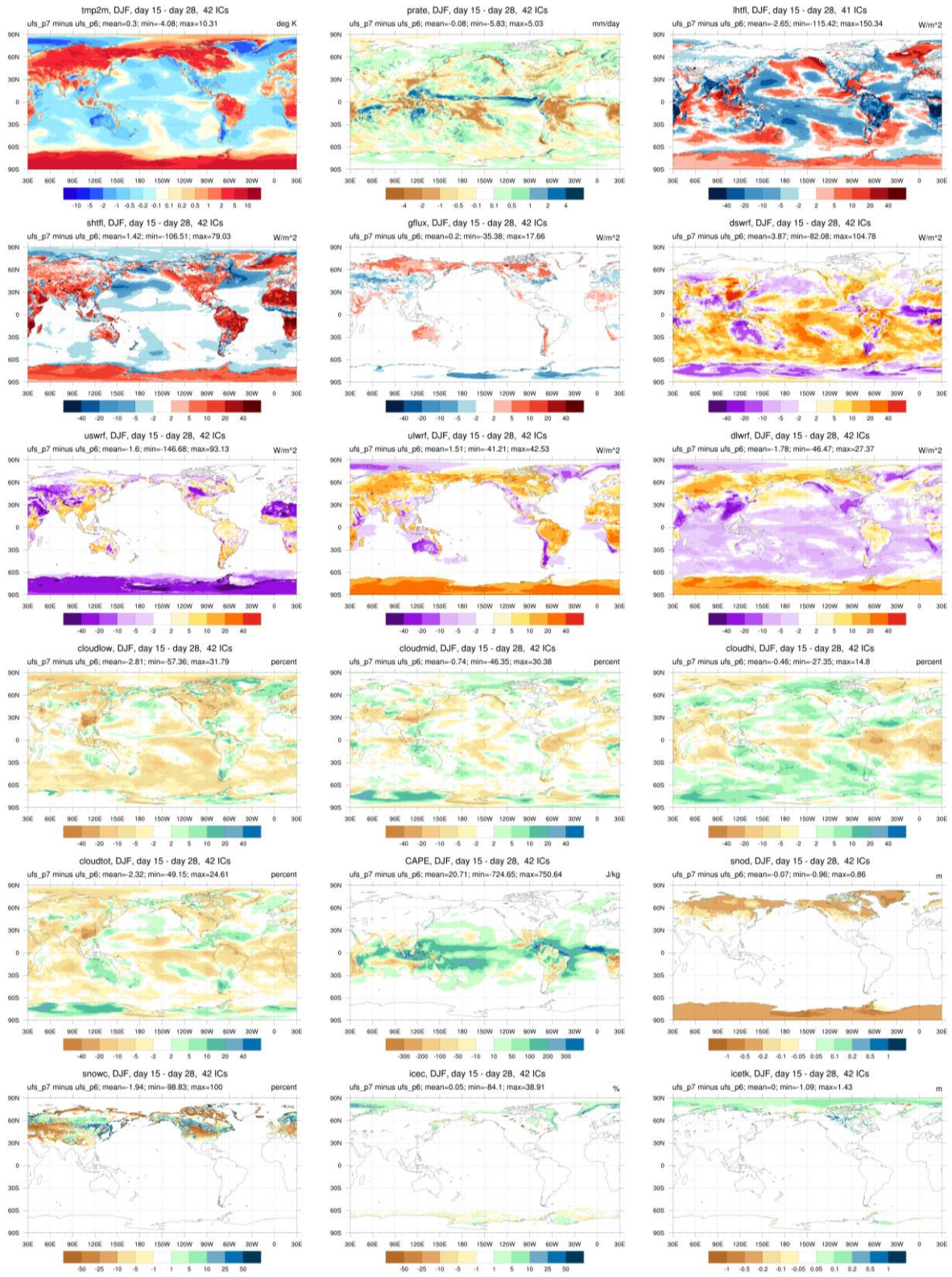

Figure 11.9: Week 3&4 average difference (P7 minus P6) for select forecast fields as indicated on subplot titles from forecasts initialized in DJF.

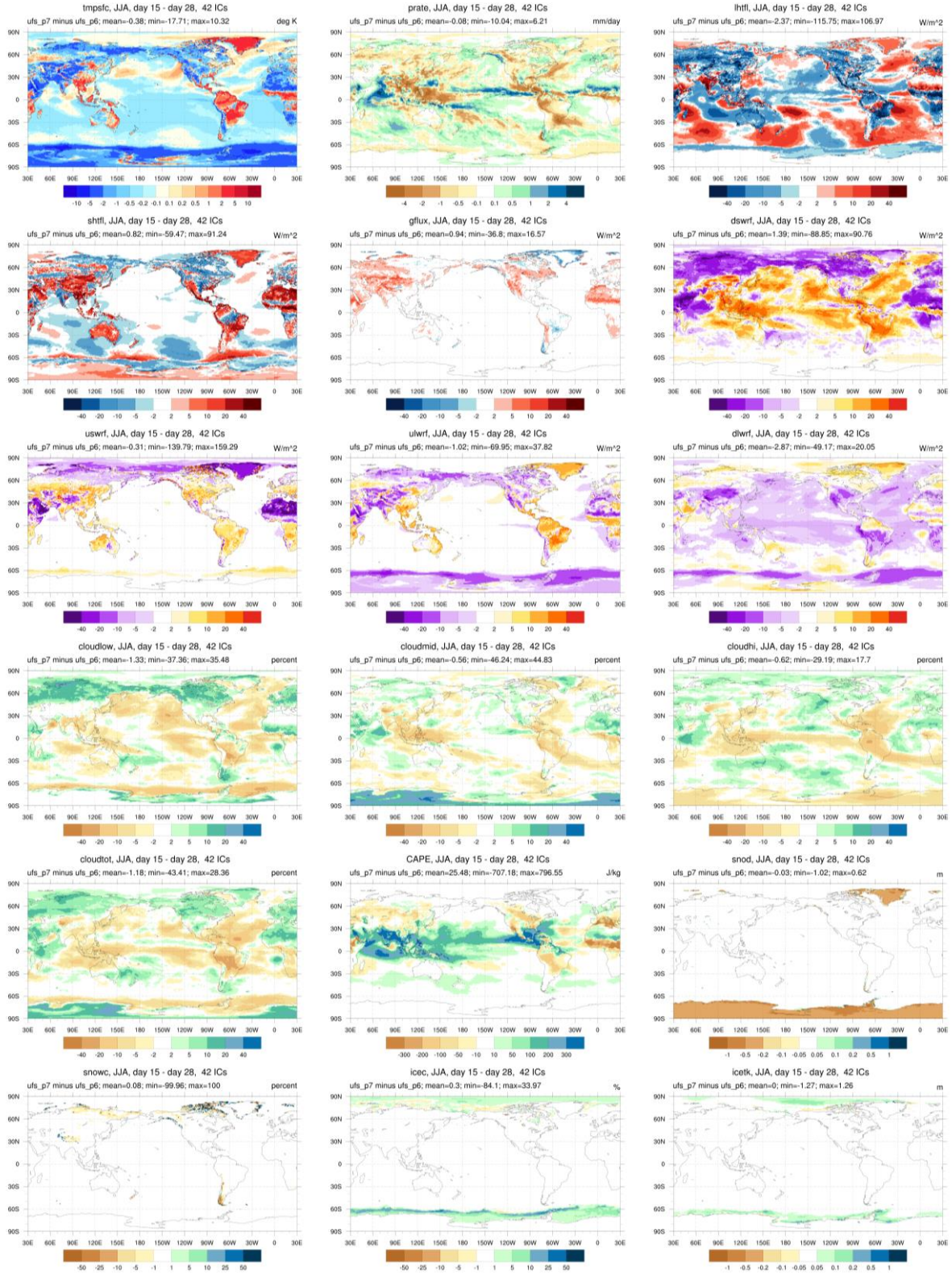

Figure 11.10: Week 3&4 average difference (P7 minus P6) for select forecast fields as indicated on subplot titles from forecasts initialized in JJA.

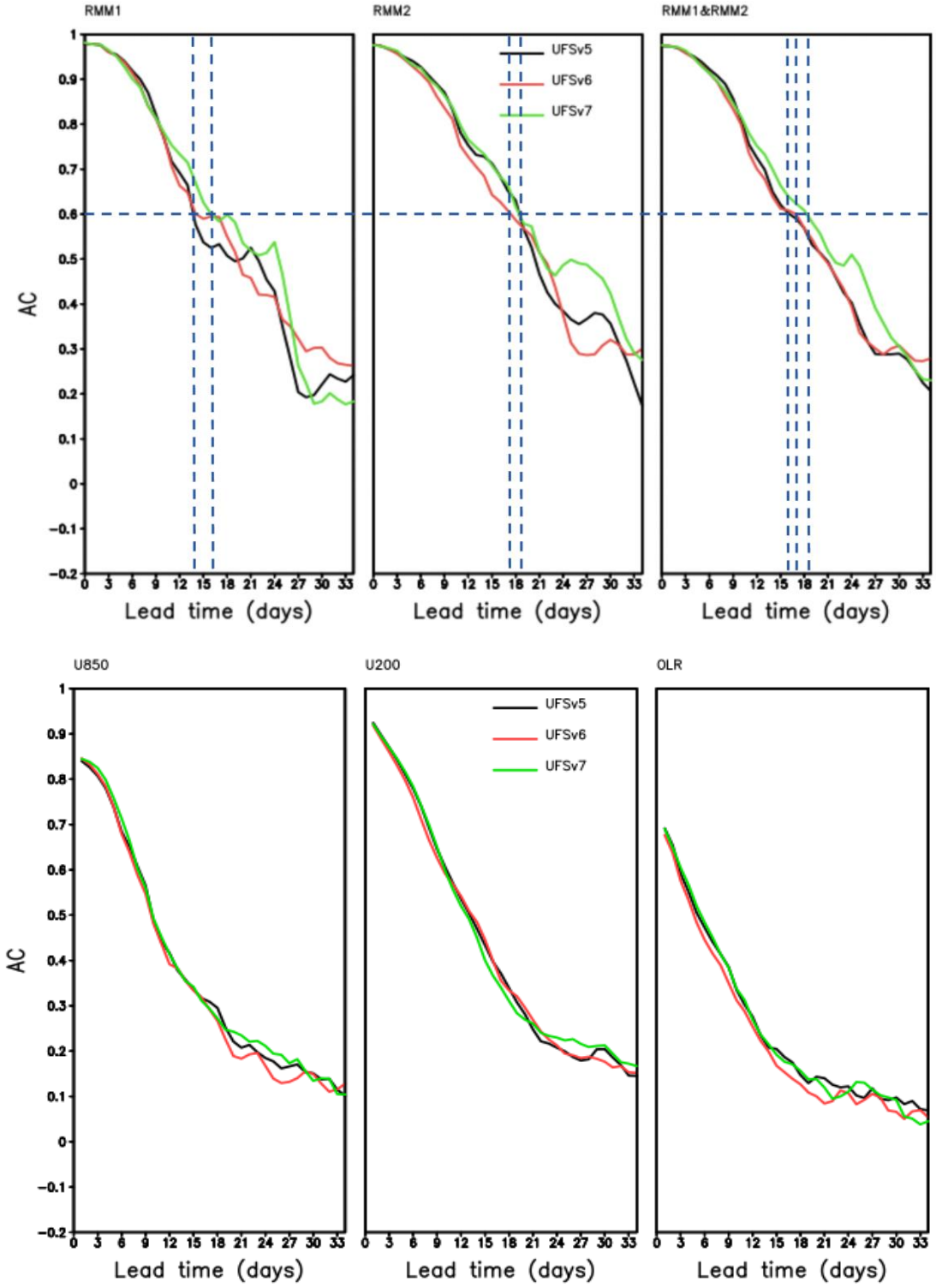

Figure 11.11: Top: AC of MJO RMM1 (left), RMM2 (middle) and combined RMM1&RMM2 (right). Bottom: AC of equatorial filtered zonal wind at 850mb (left) and 200mb (middle), and equatorial filtered OLR (right) in P5 (black line), P6 (red line), and P7 (green line).

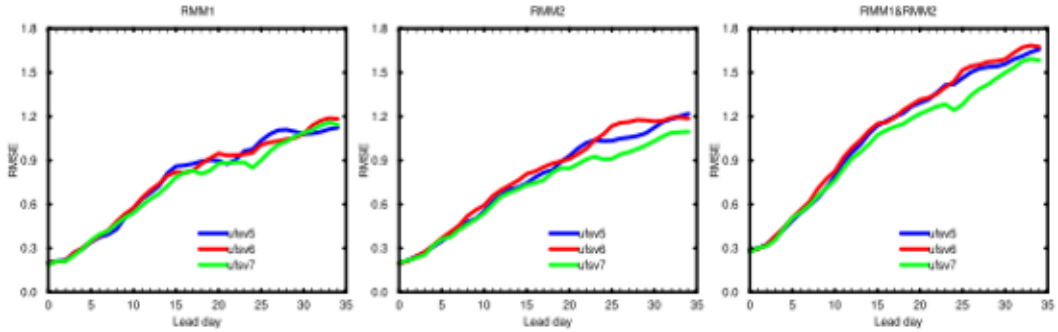

Figure 11.12: RMSE of MJO RMM1, RMM2 and RMM1&2 for P5 (blue), P6 (red), and P7 (green).

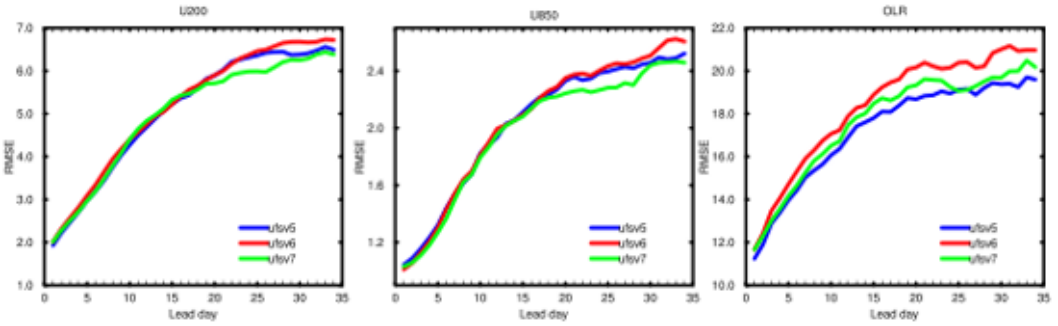

Figure 11.13: RMSE of MJO components (U200, U850 and OLR) for P5 (blue), P6 (red), and P7 (green).

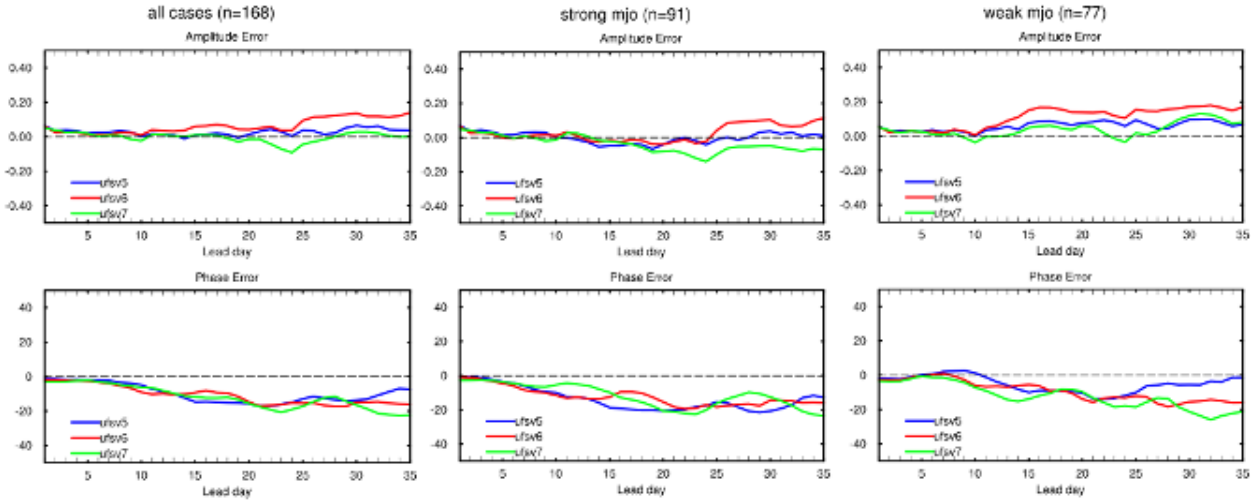

Figure 11.14: MJO amplitude and phase error for P5 (blue), P6 (red), and P7 (green) separated into strong and weak MJOs.

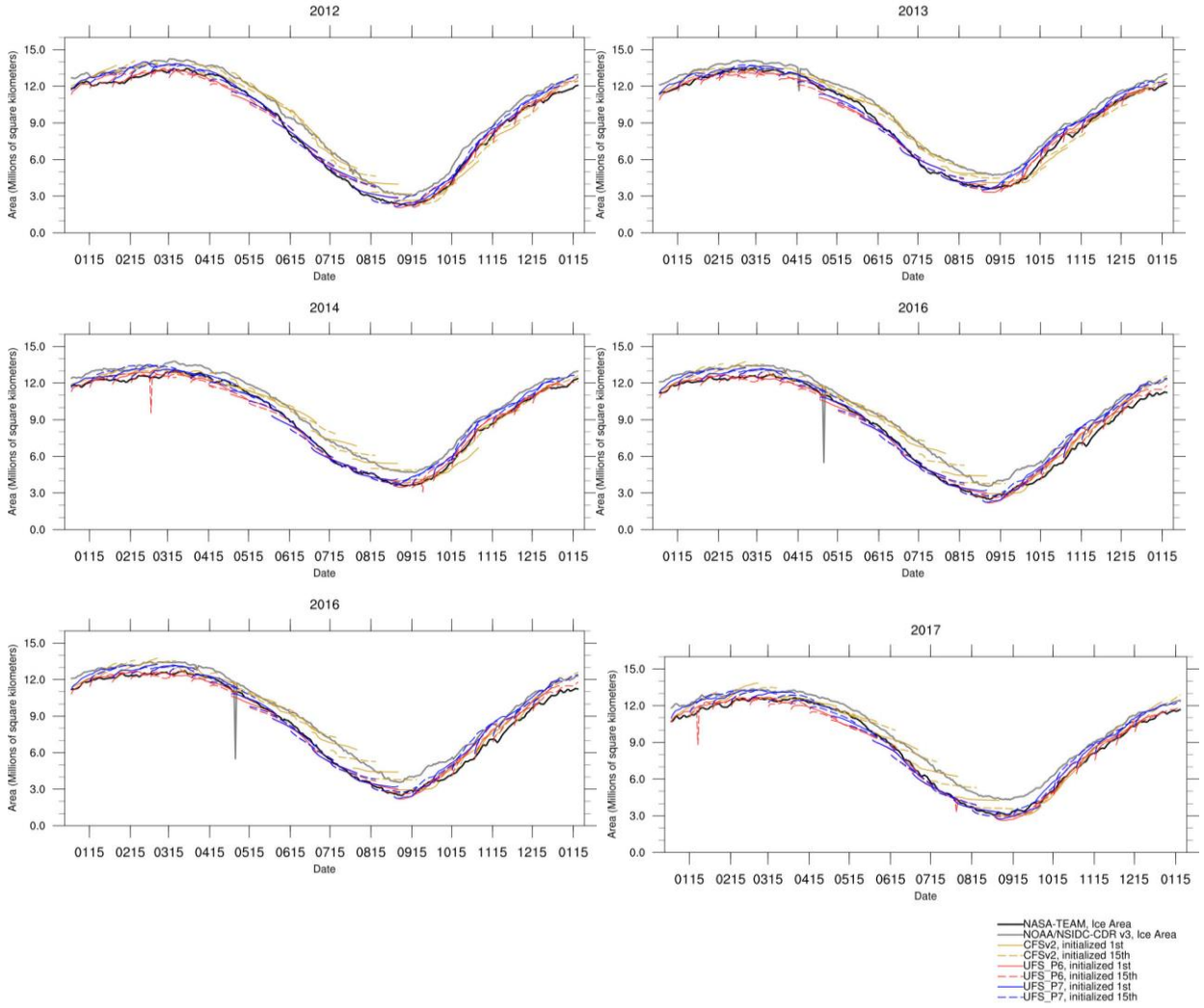

Figure 11.15: Arctic sea ice area for all forecasts between 2012-2017. Solid/dashed lines correspond to forecasts initialized on the 1st/15th of the month for CFSv2 (yellow), P6(red), and P7 (blue). Thick lines correspond to the observational estimate derived from ice concentrations from NASA-Team (black) and NOAA/NSIDC CDR v3 (gray). Any spikes are artifacts due to occasional data point corruption during post-processing, and do not reflect the actual model forecast for that date

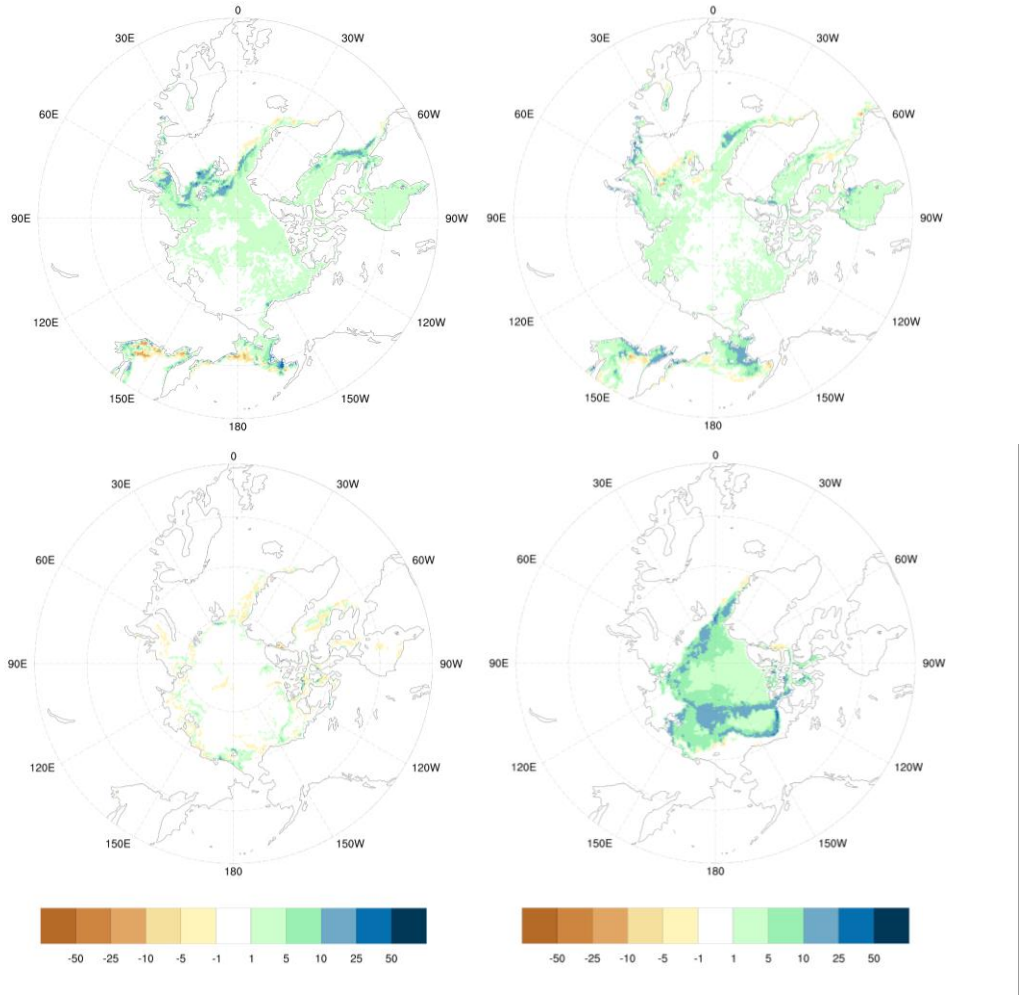

Figure 11.16: Average difference (P7b minus P7a) of ice concentration [%] over week 1 in forecasts initialized on the 1st of January (top left), April (top right), July (bottom left) and October (bottom right), 2012 and 2013.

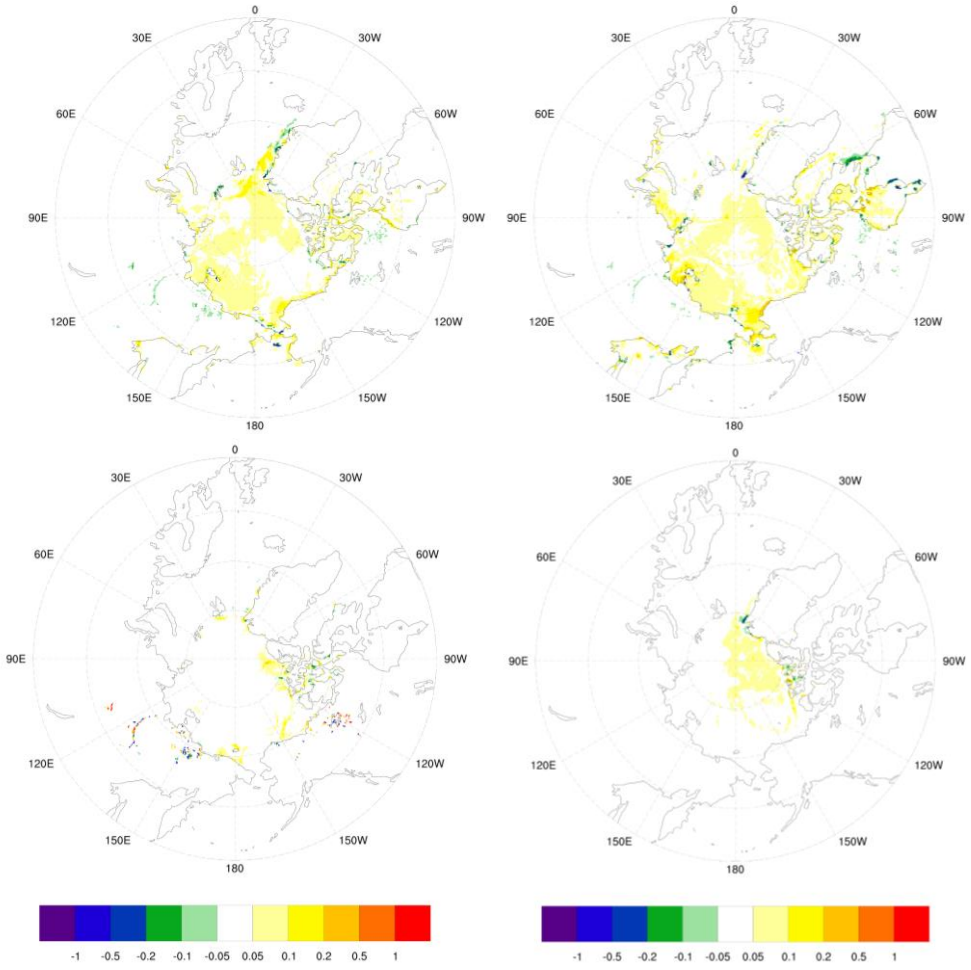

Figure 11.17: Average difference (P7b minus P7a) of ice thickness [m] over week 1 in forecasts initialized on the 1st of January (top left), April (top right), July (bottom left) and October (bottom right), 2012 and 2013.

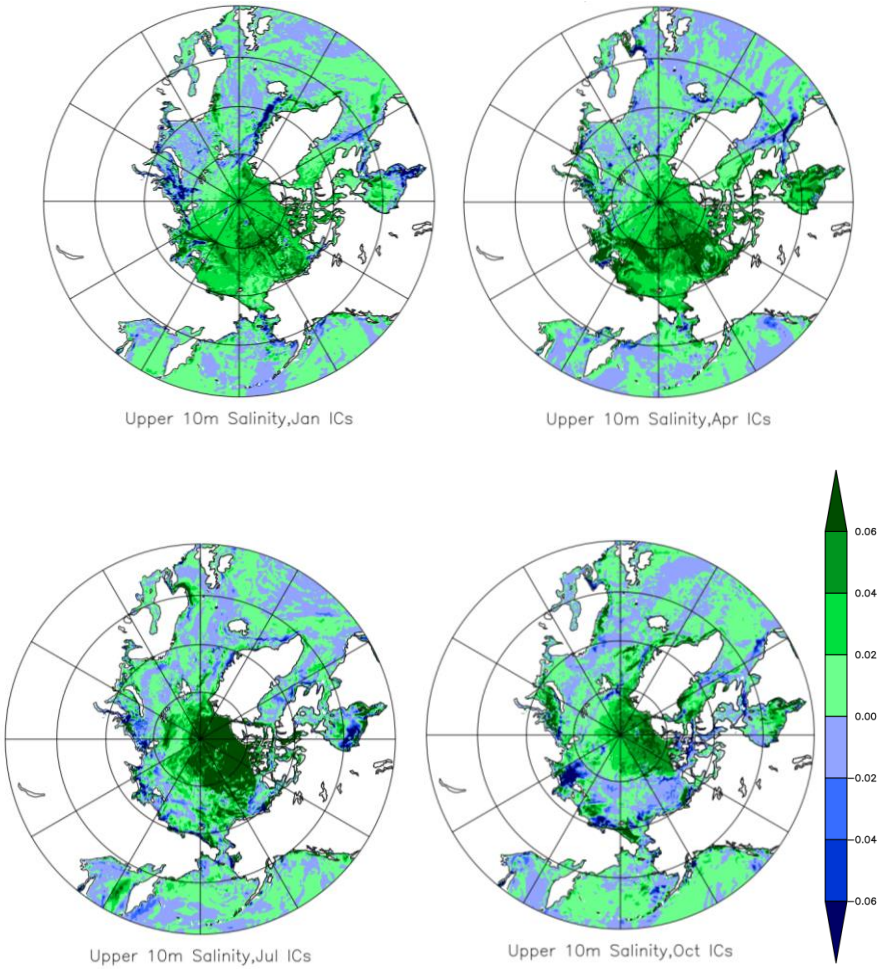

Figure 11.18: Average difference (P7b minus P7a) of upper 10m salinity [psu] over week 1 in forecasts initialized on the 1st of January (top left), April (top right), July (bottom left) and October (bottom right), 2012 and 2013.

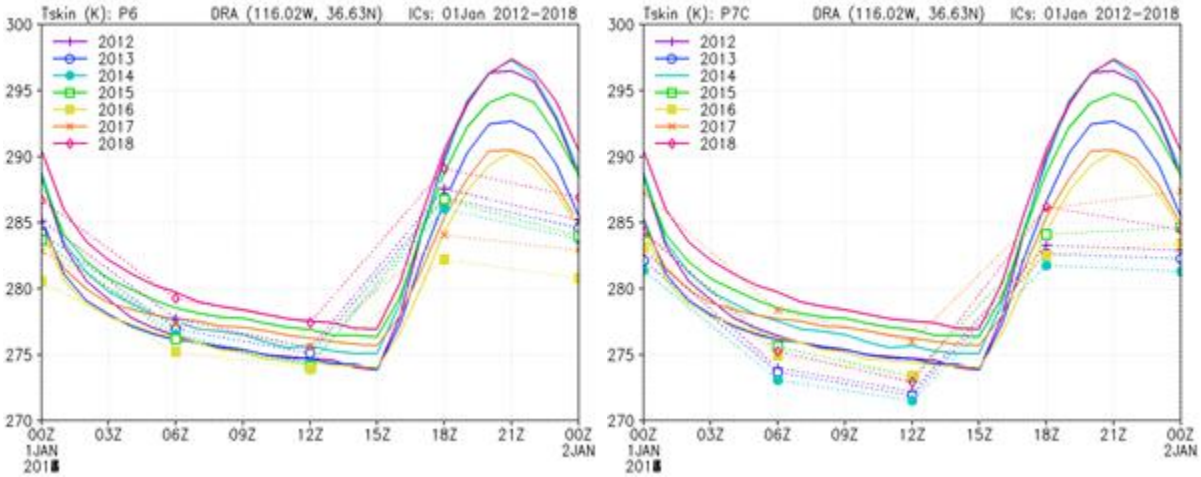

Figure 11.19: Average diurnal cycle of land skin temperatures in January 2012–2018 from SURFRAD (solid lines) and UFS (dashed lines) at Desert Rock, Nevada. (a) is from P6 and (b) is from P7.

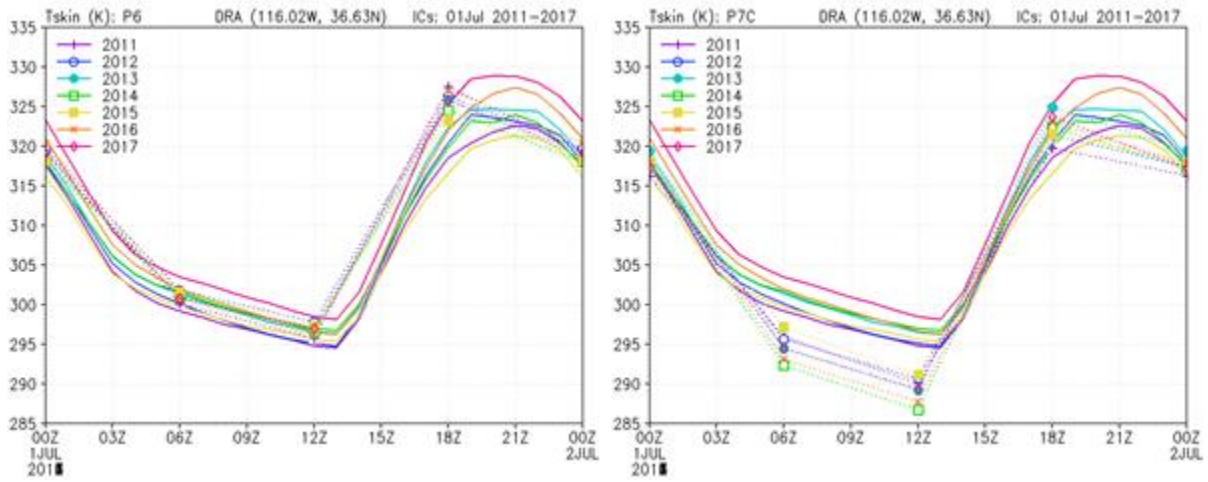

Figure 11.20: Average diurnal cycle of land skin temperatures in July 2011–2017 from SURFRAD (solid lines) and UFS (dashed lines) at Desert Rock, Nevada. (a) is from P6 and (b) is from P7.

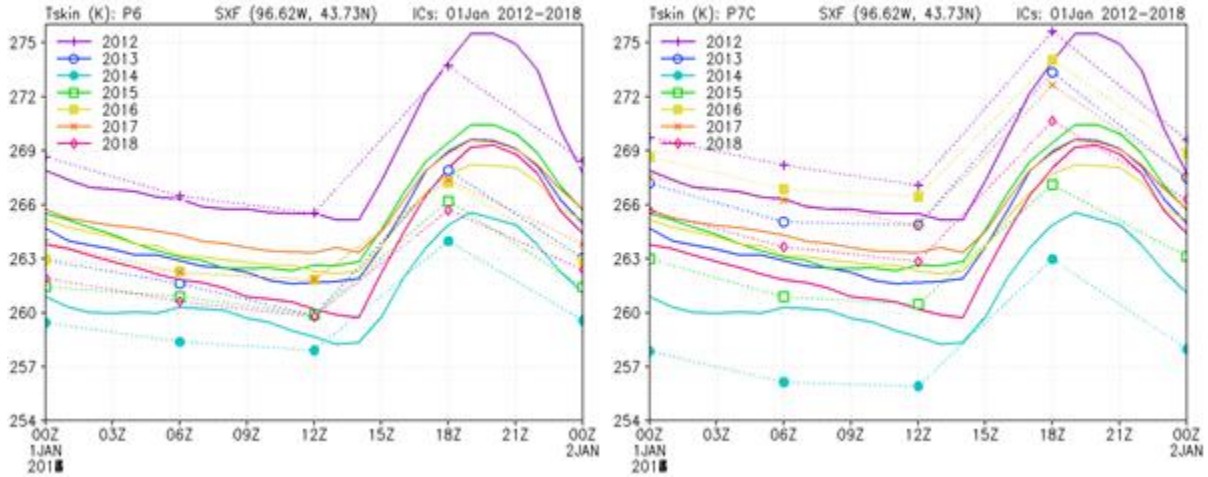

Figure 11.21a: Average diurnal cycle of land skin temperatures in January 2012–2018 from SURFRAD (solid lines) and UFS (dashed lines) at Sioux Falls, South Dakota. Results from P6 (left) and P7 (right).

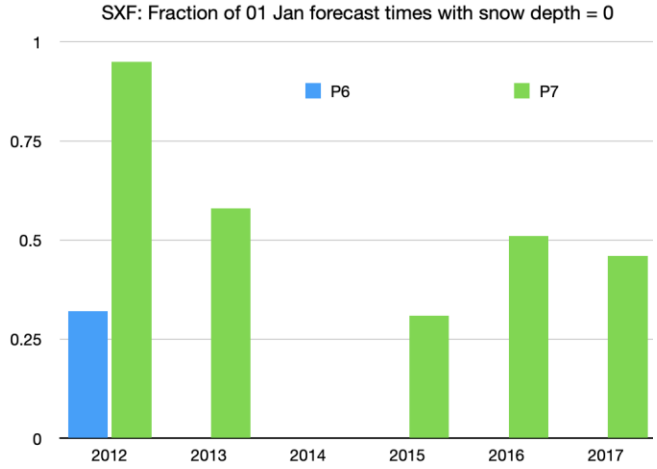

Figure 11.21b: Fraction of forecast times in Jan 2012–2018 P6 (blue) and P7 (green) with snow-free ground at Sioux Falls, South Dakota.

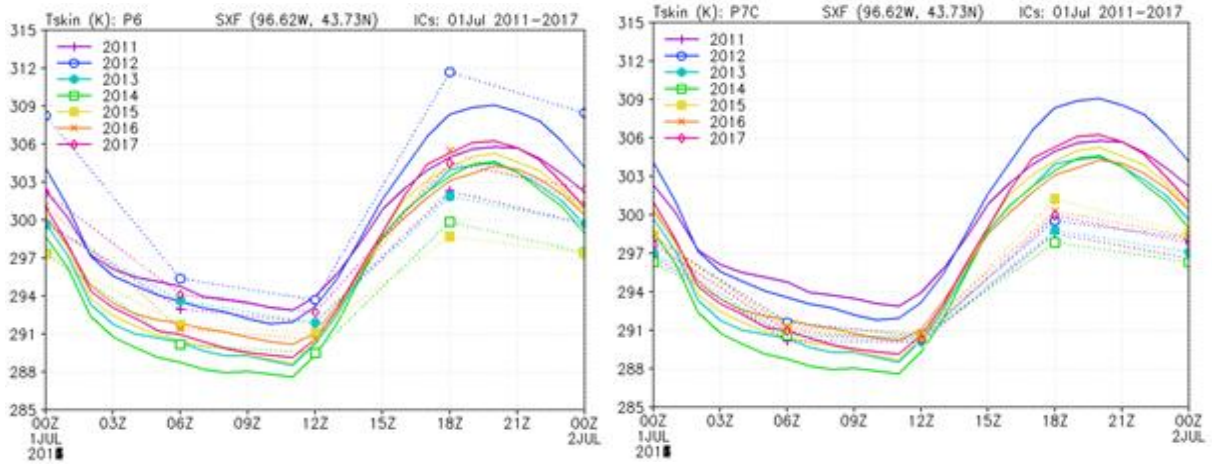

Figure 11.22: Average diurnal cycle of land skin temperatures in July 2011–2017 from SURFRAD (solid lines) and UFS (dashed lines) at Sioux Falls, South Dakota. Results from P6 (left) and P7 (right).

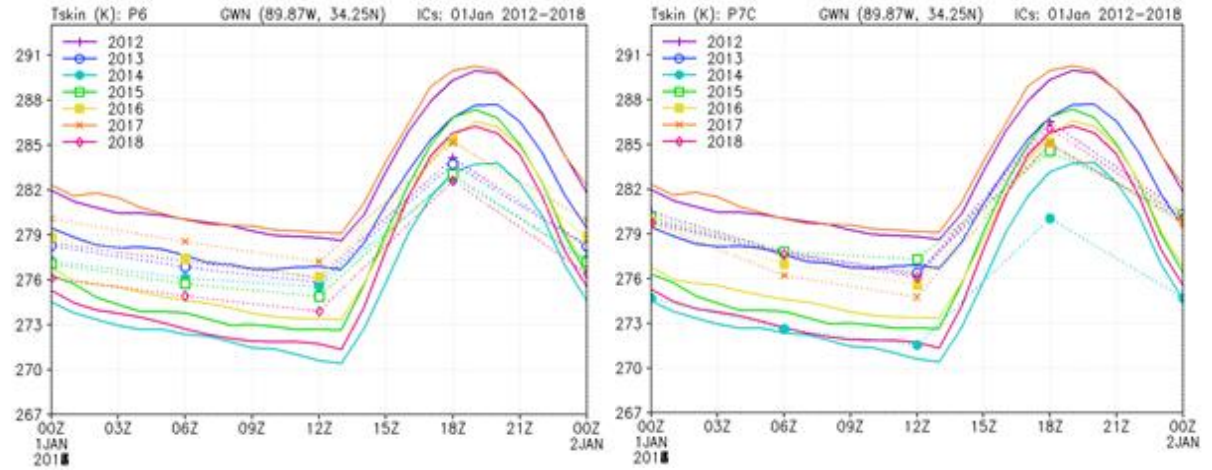

Figure 11.23: Average diurnal cycle of land skin temperatures in January 2012–2018 from SURFRAD (solid lines) and UFS (dashed lines) at Goodwin Creek, Mississippi. Results from P6 (left) and P7 (right).

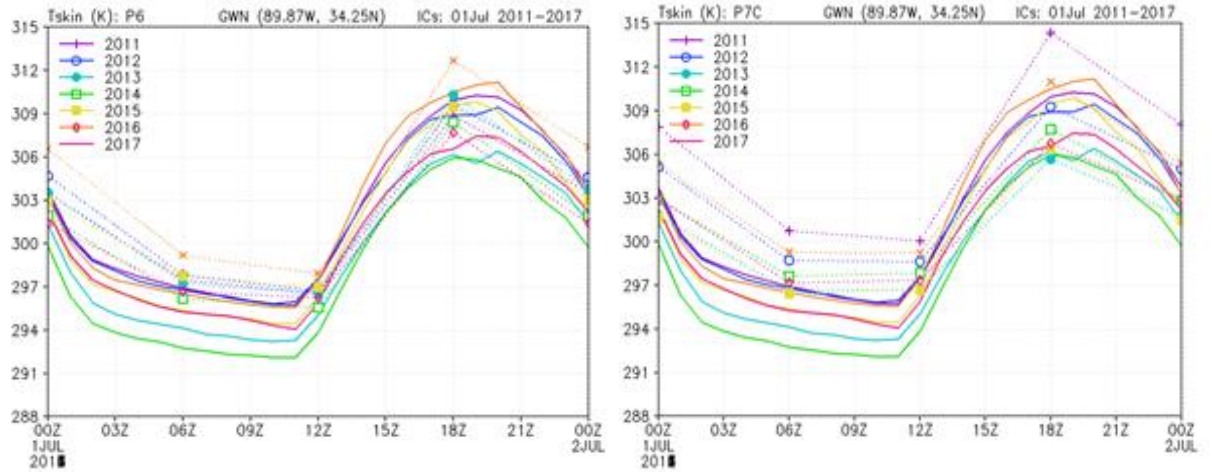

Figure 11.24: Average diurnal cycle of land skin temperatures in July 2011–2017 from SURFRAD (solid lines) and UFS (dashed lines) at Goodwin Creek, Mississippi. Results from P6 (left) and P7 (right).

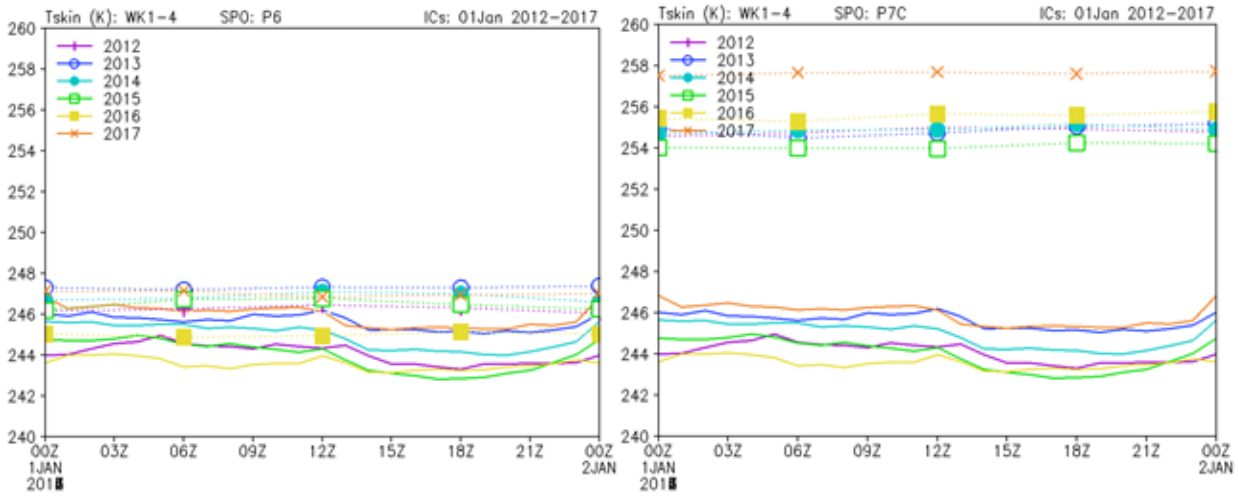

Figure 11.25: Average diurnal cycle of land skin temperatures at South Pole in January 2012-2017 from the observation (solid lines) and UFS (dashed lines). Results from P6 (left) and P7 (right).

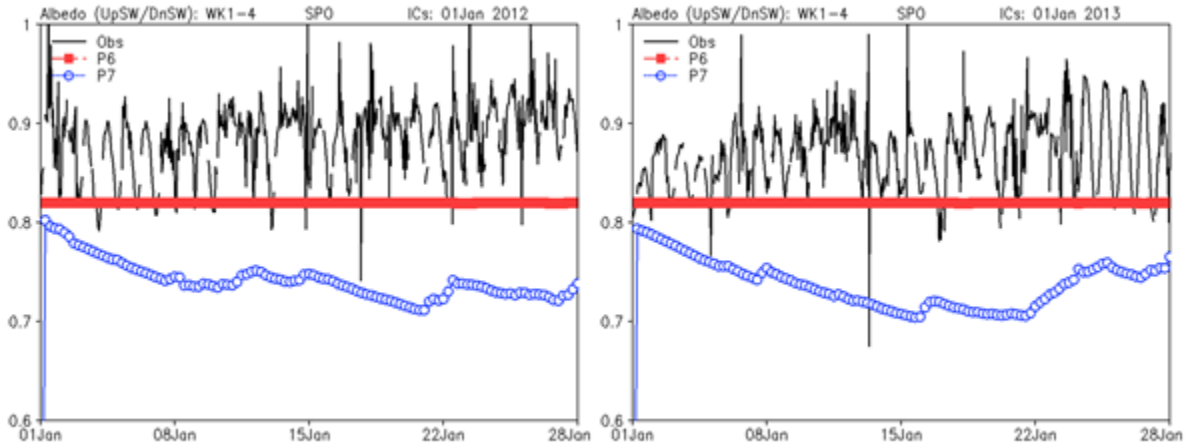

Figure 11.26: Time series of ratio of upward SW flux to downward SW flux in 2012 (a) and 2013 for P6 (red lines) and P7 (blue lines) against the observation (black lines).

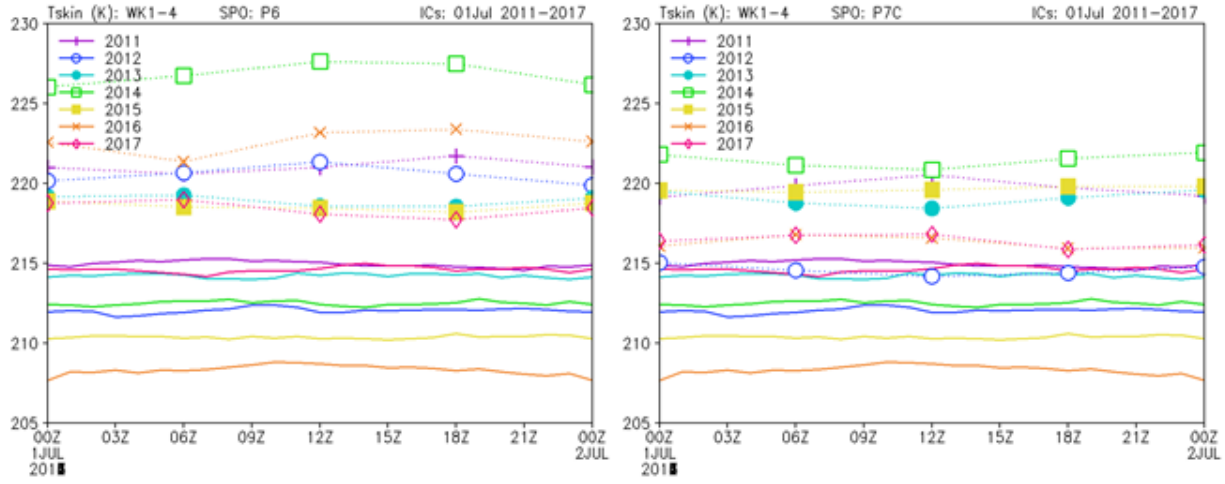

Figure 11.27: Average diurnal cycle of land skin temperatures at South Pole in July 2011-2017 from the observation (solid lines) and UFS (dashed lines). Results from P6 (left) and P7 (right).

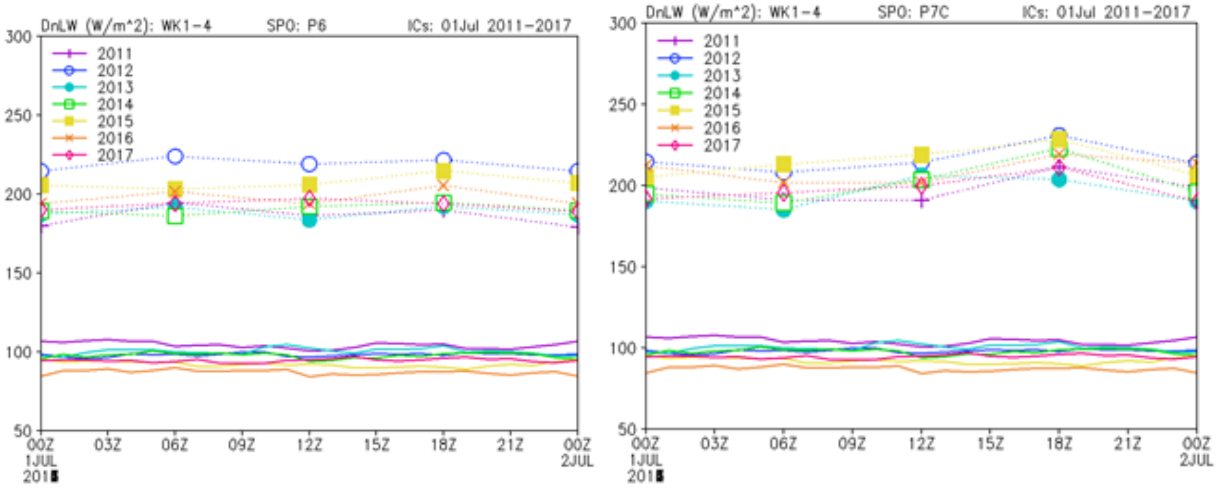

Figure 11.28: Average diurnal cycle of downward long-wave radiation at the surface at South Pole in July 2011-2017 from the observation (solid lines) and UFS (dashed lines). Results from P6 (left) and P7 (right).

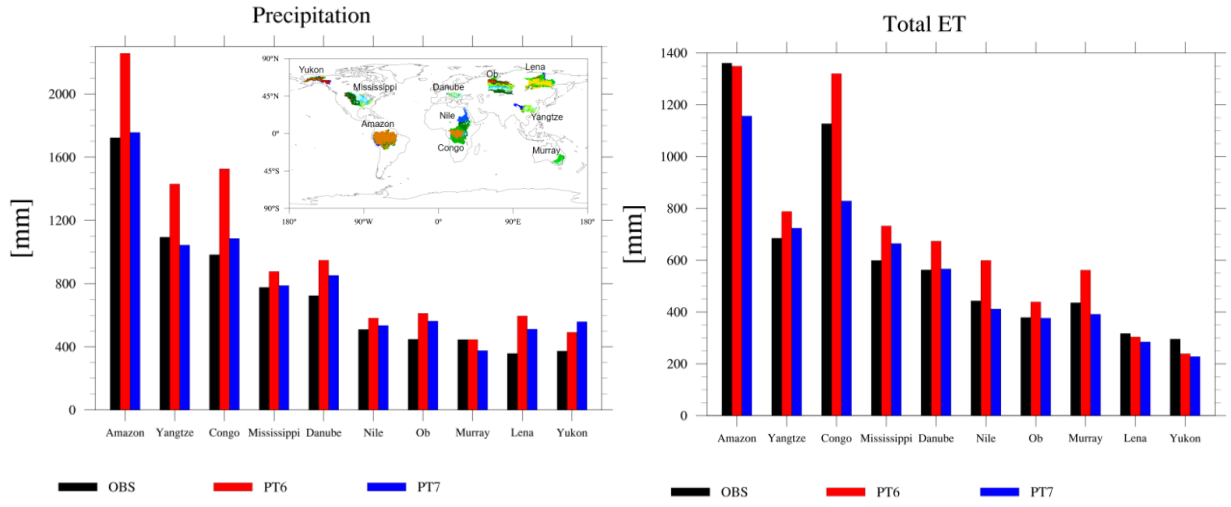

Figure 11.29: Mean annual precipitation [mm] and evapotranspiration [mm] over 10 large river basins. The basins are organized from left to right based on decreasing observed precipitation. Observations from CPC global 0.5° daily (precipitation) and GLEAM global 0.25° monthly (ET) products.

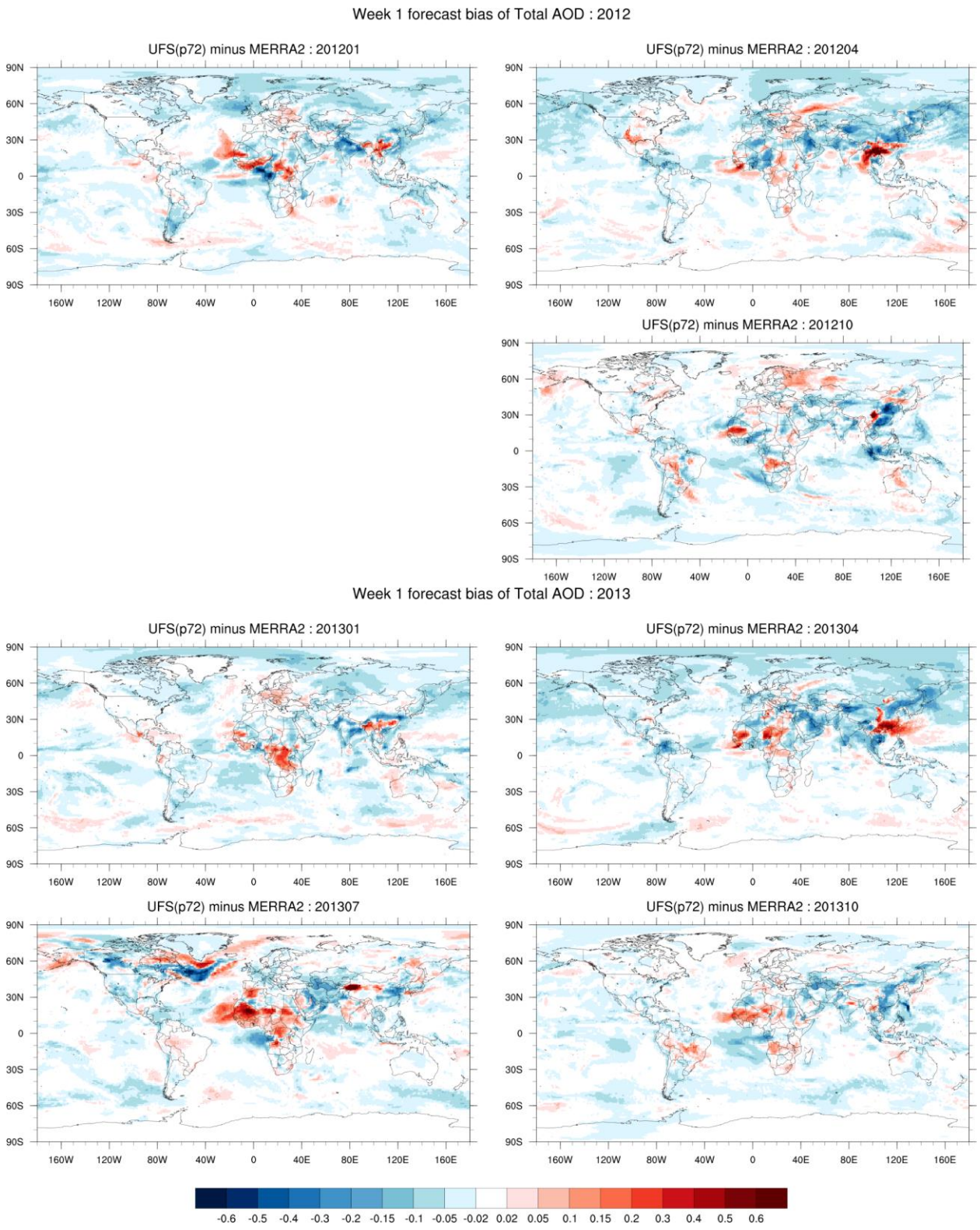

Figure 11.30: Total AOD forecast error for Week 1 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

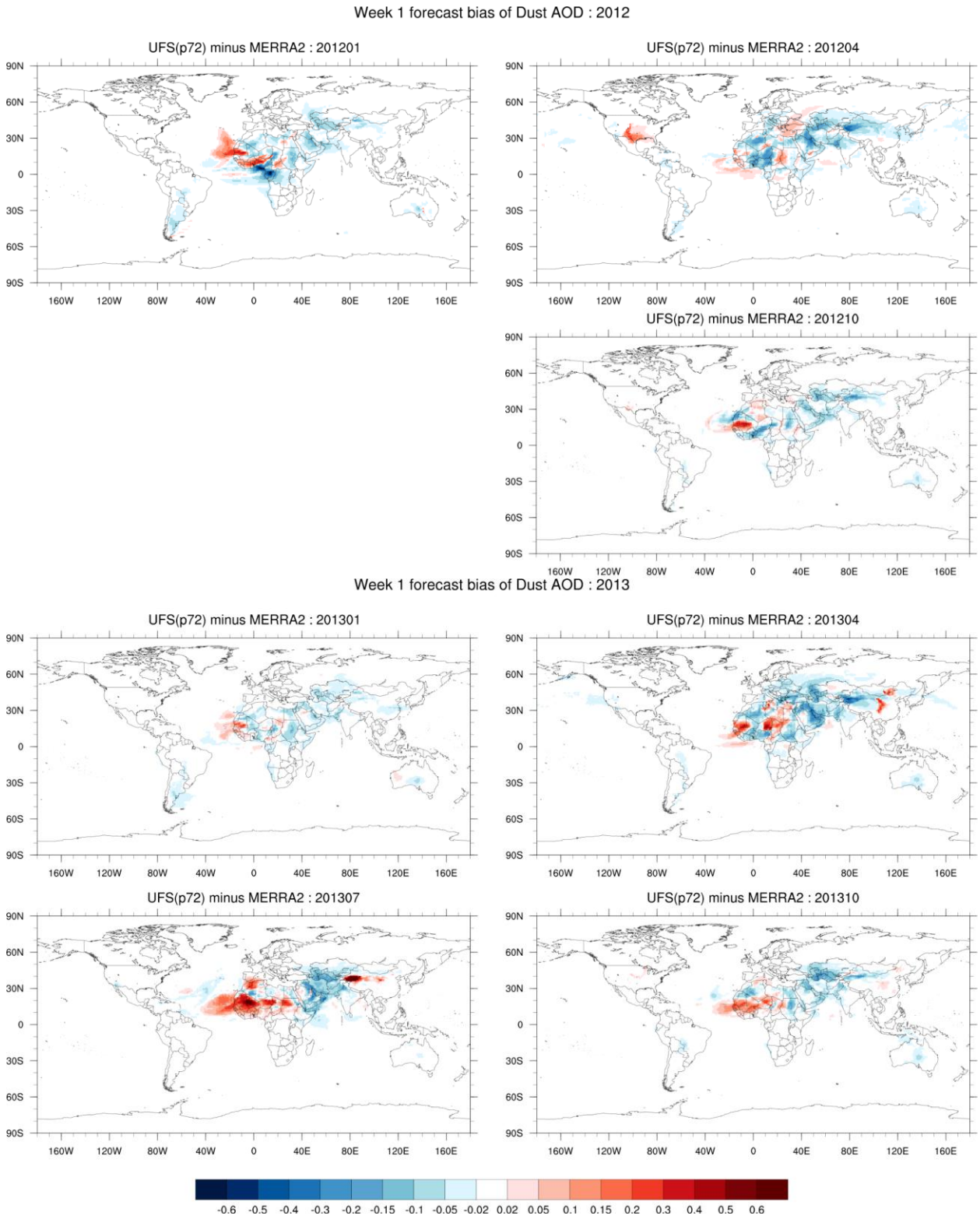

Figure 11.31: Dust AOD forecast error for Week 1 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels)

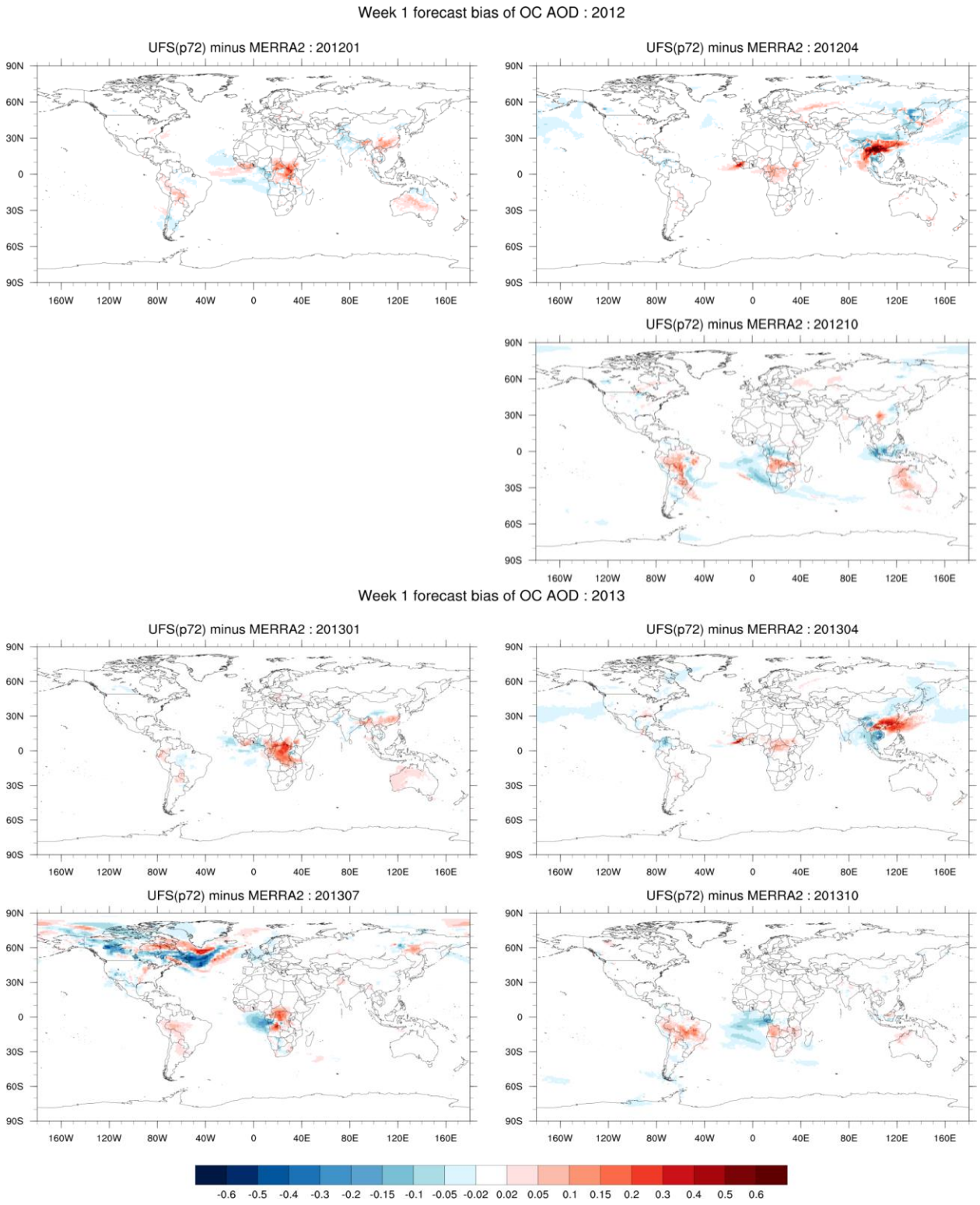

Figure 11.32: Organic carbon AOD forecast error for Week 1 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

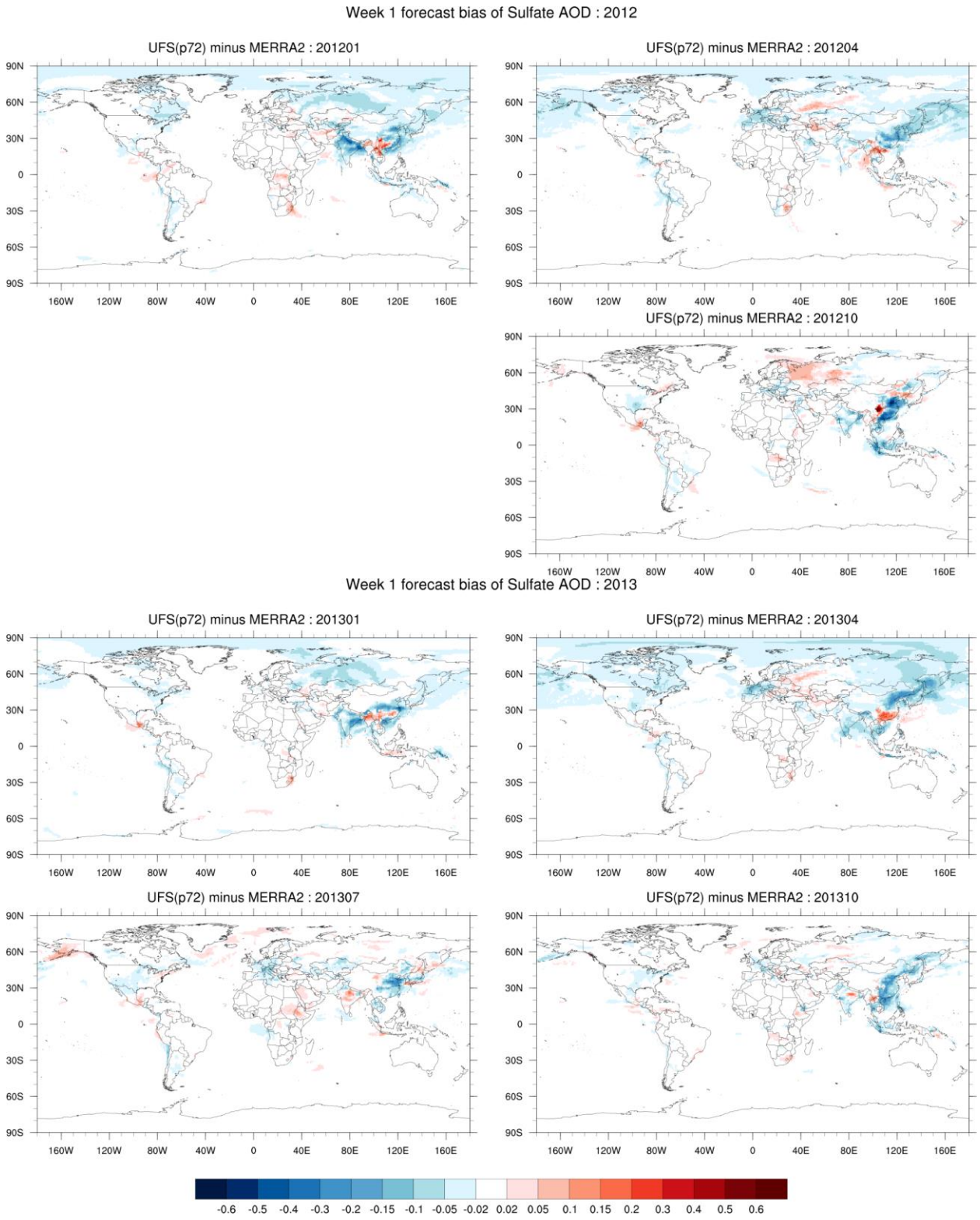

Figure 11.33: Sulfate AOD forecast error for Week 1 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

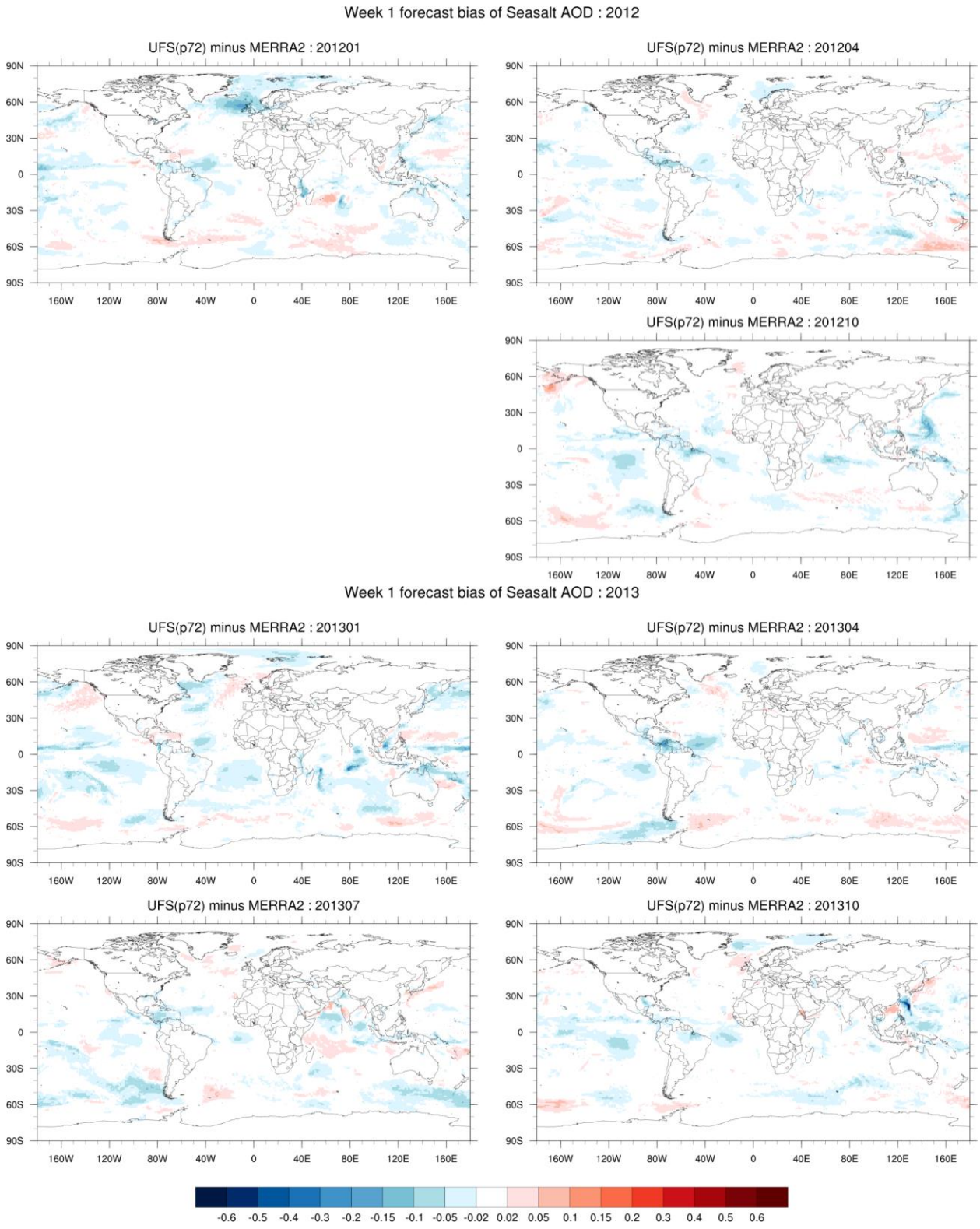

Figure 11.34: Sea salt AOD forecast error for Week 1 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

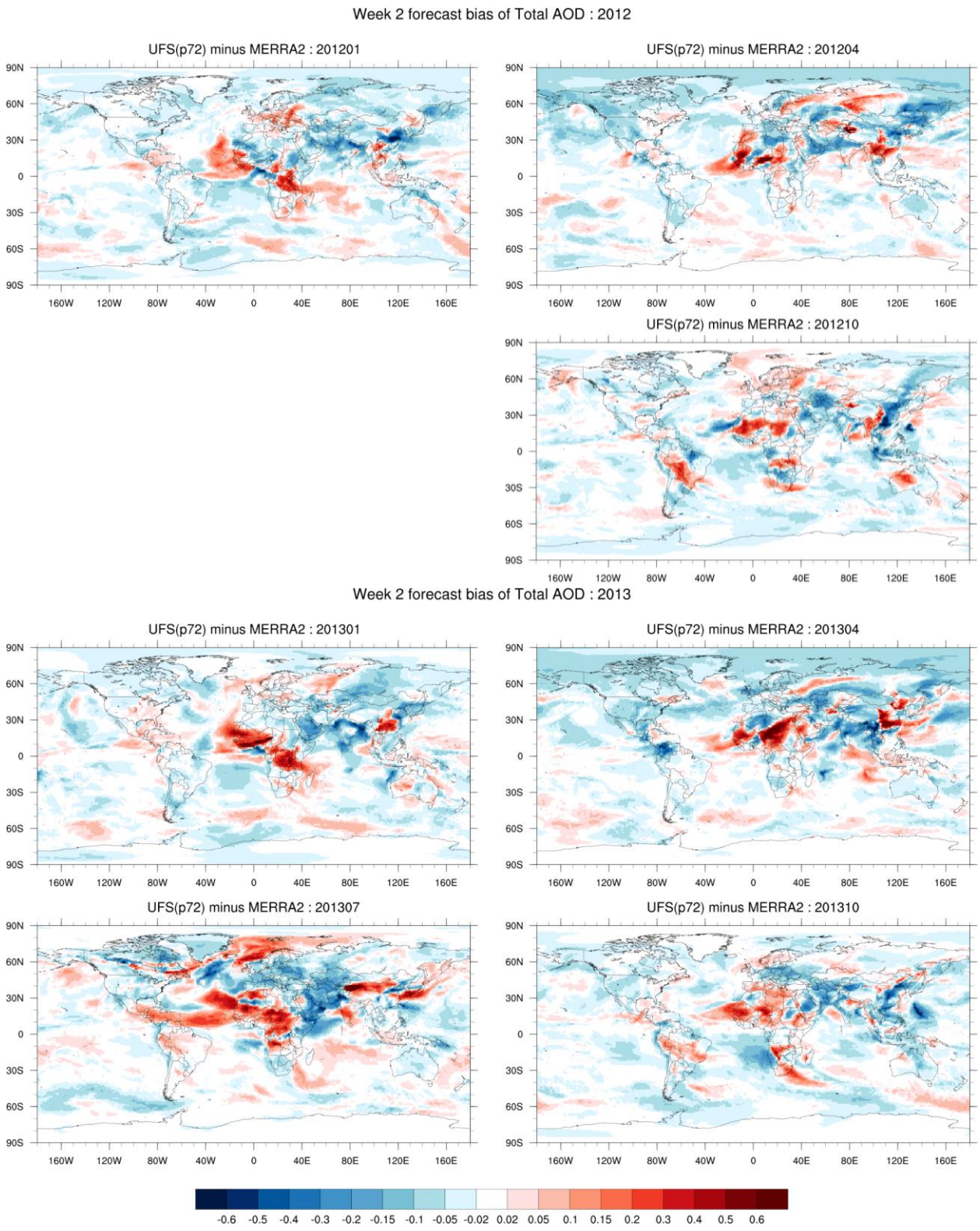

Figure 11.35: Total AOD forecast error for Week 2 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

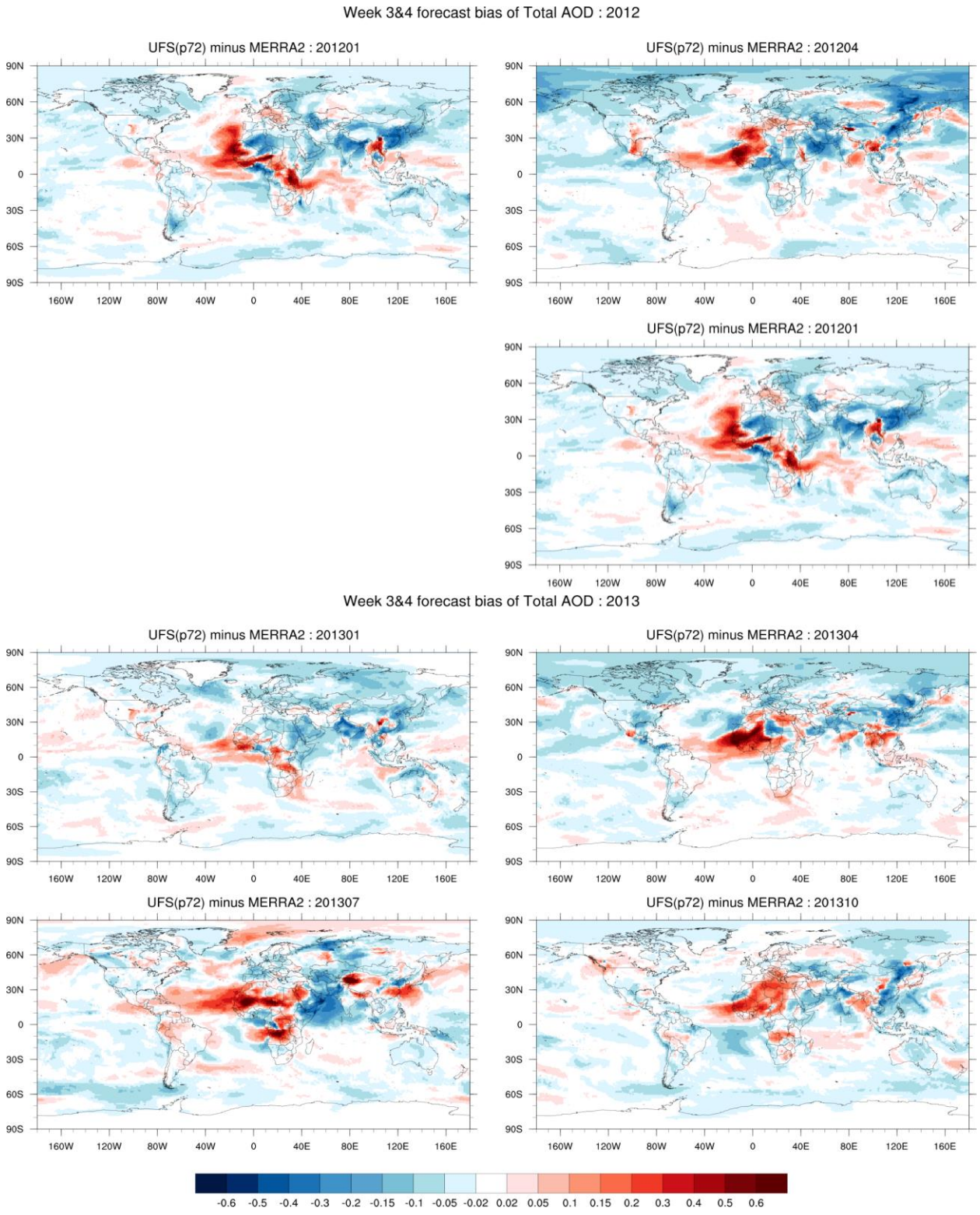

Figure 11.36: Total AOD forecast error for Weeks 3&4 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

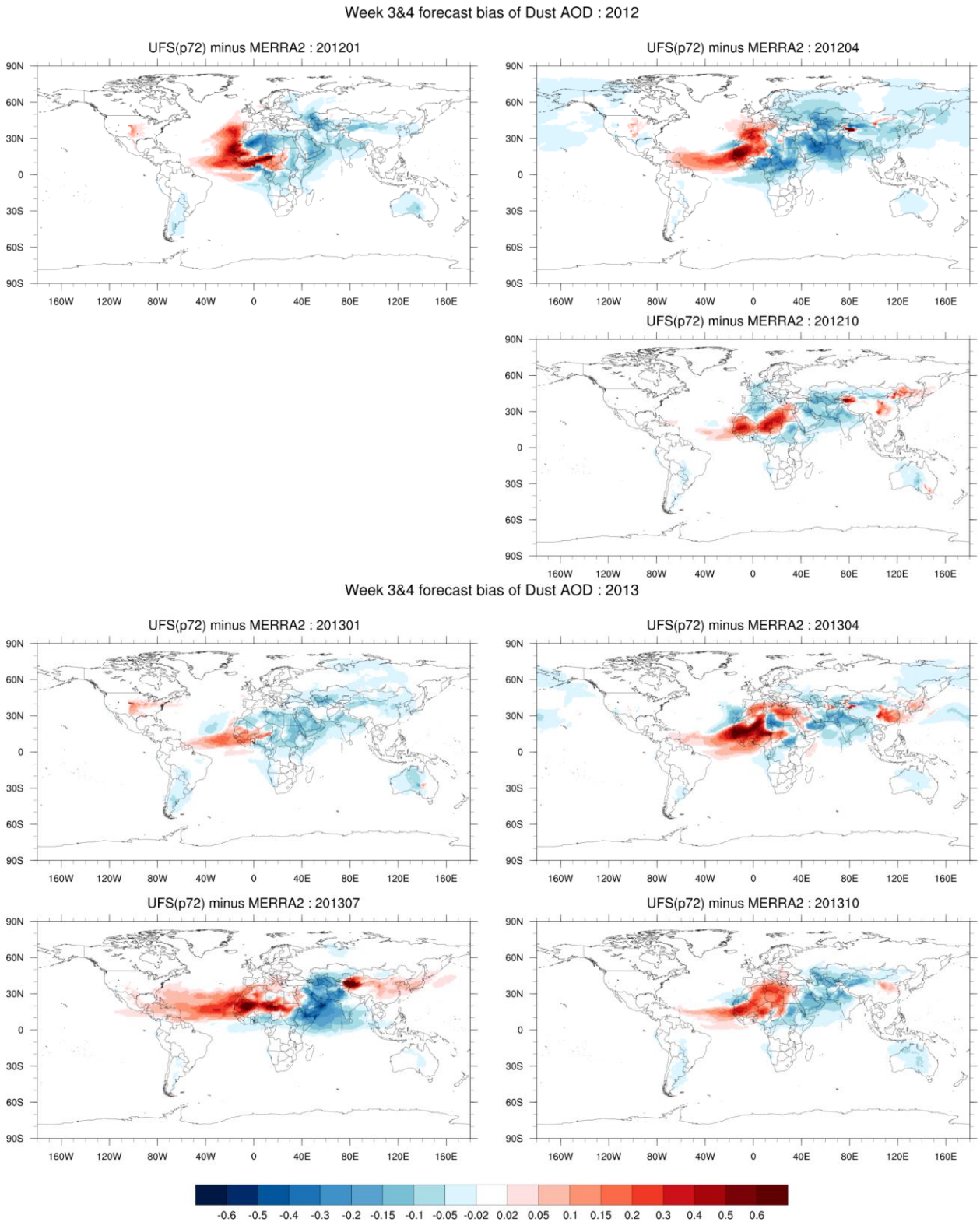

Figure 11.37: Dust AOD forecast error for Weeks 3&4 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

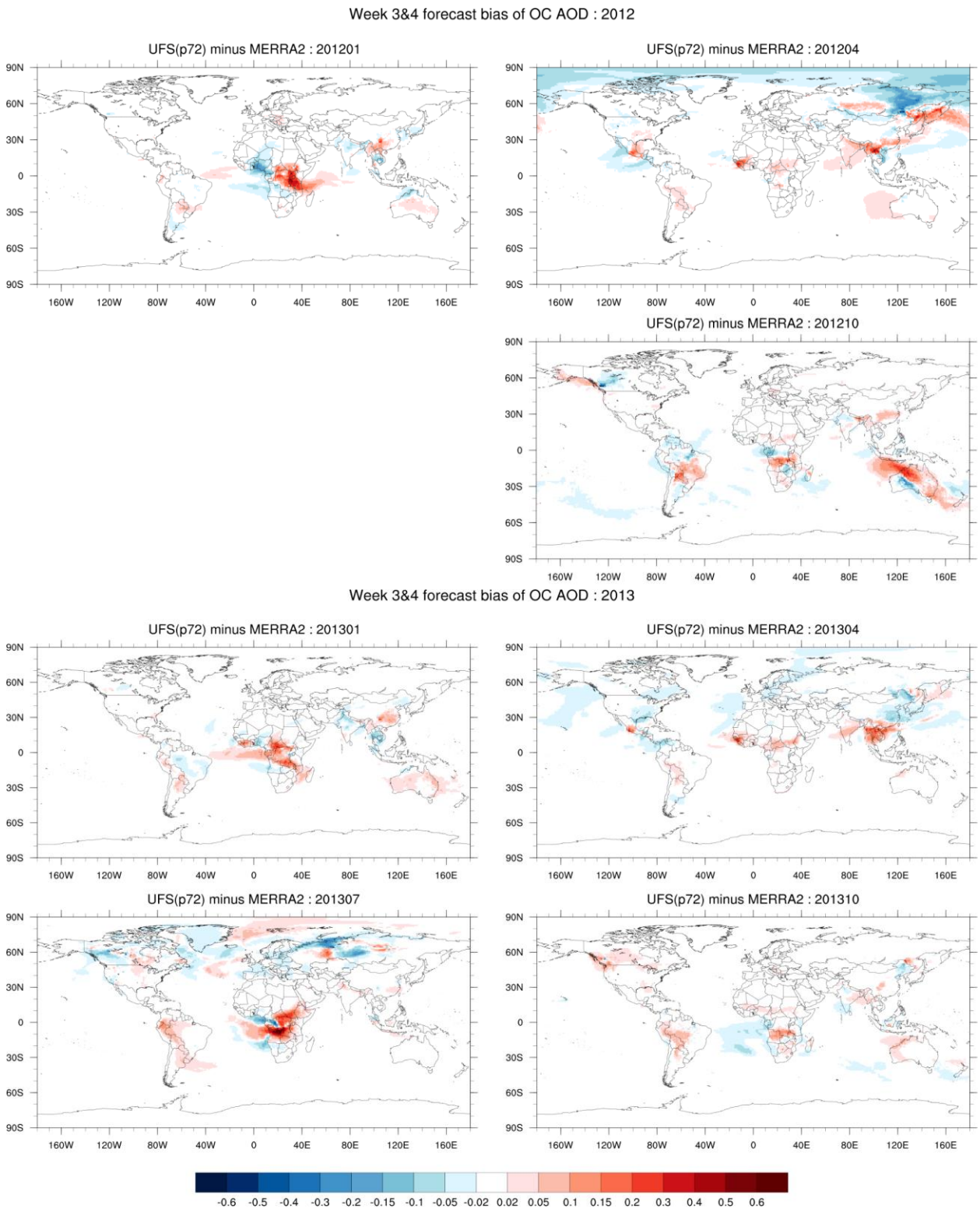

Figure 11.38: Organic carbon AOD forecast error for Weeks 3&4 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

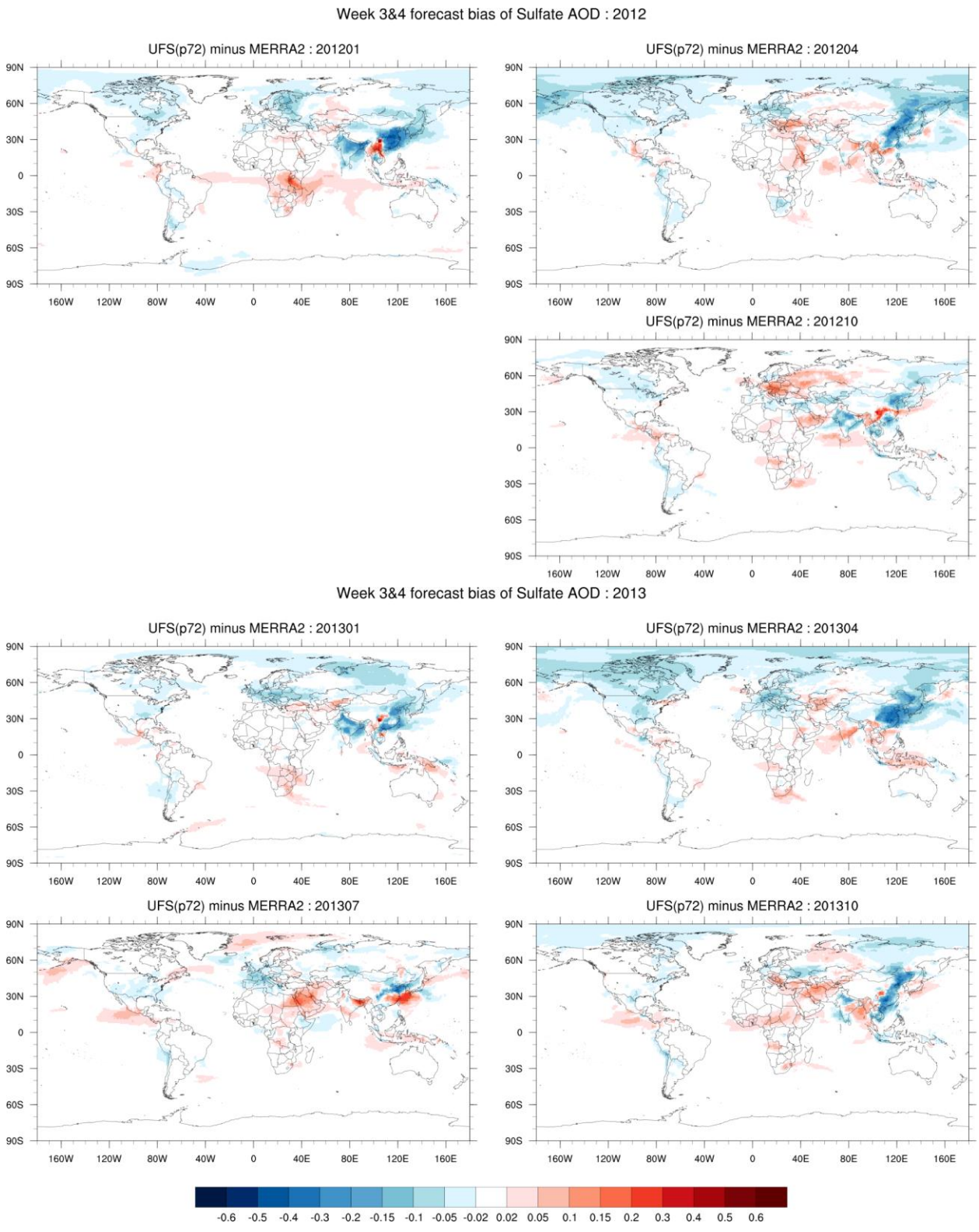

Figure 11.39: Sulfate AOD forecast error for Weeks 3.4 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

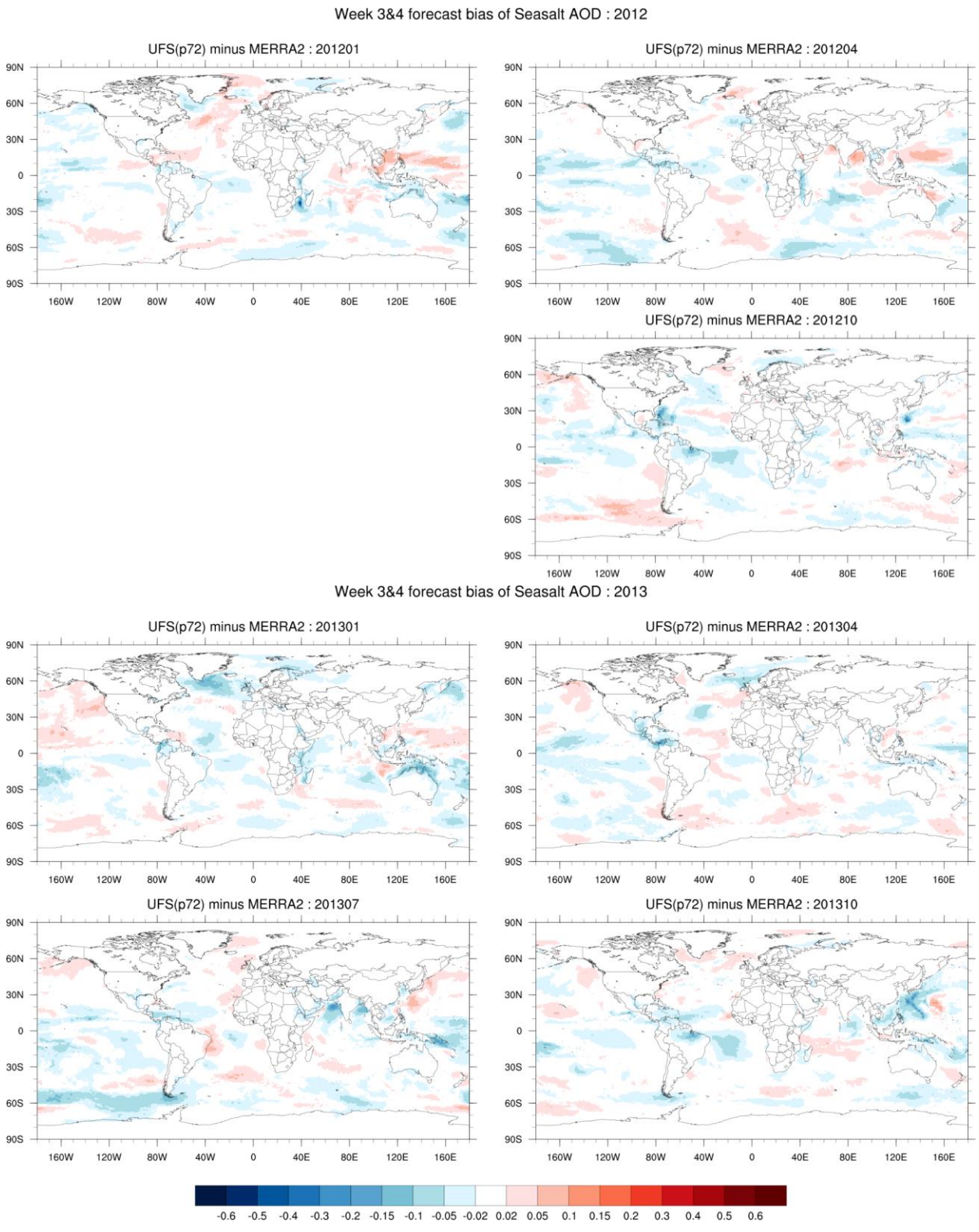

Figure 11.40: Sea salt AOD forecast error for Weeks 3.4 (P7.2 minus MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

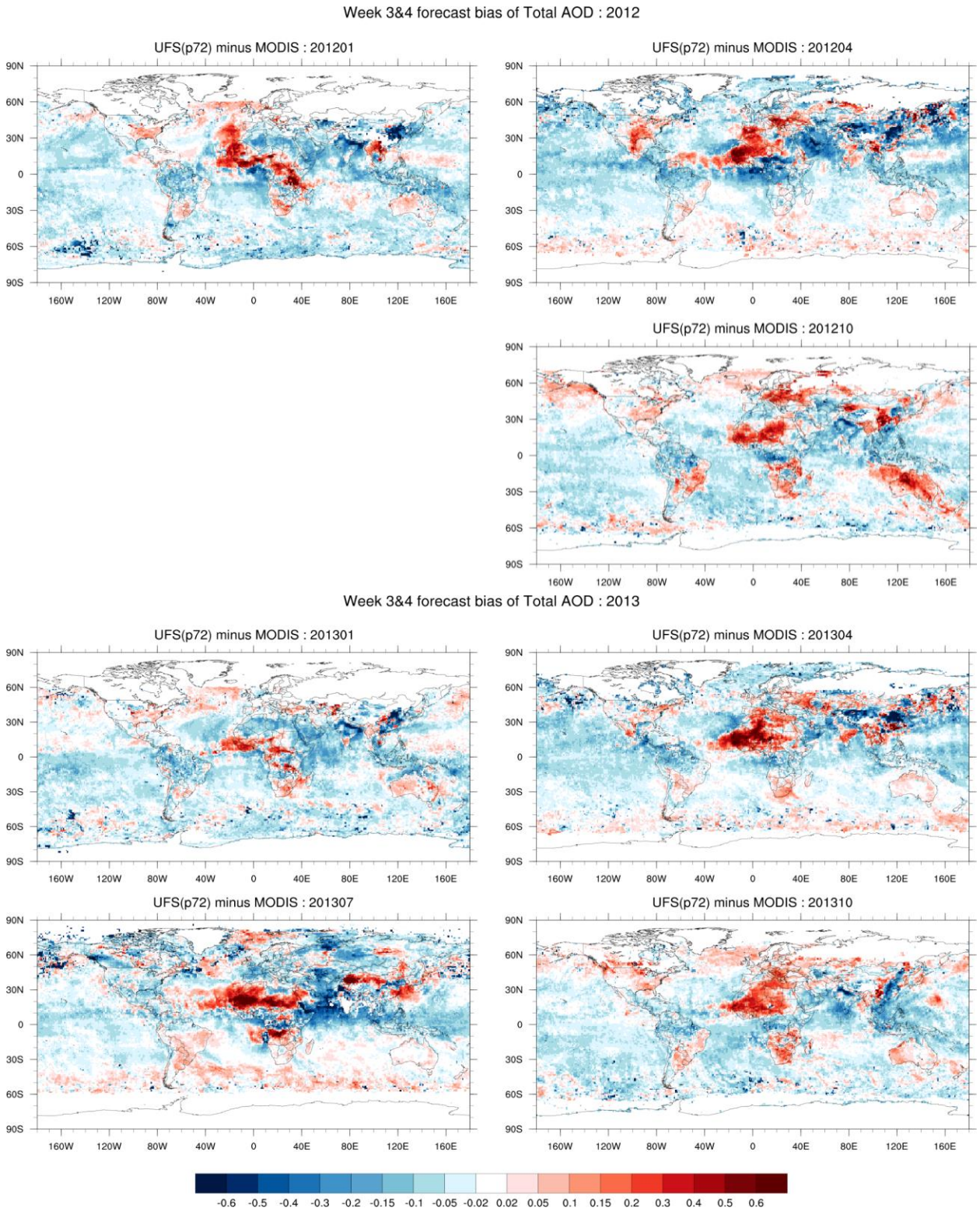

Figure 11.41: Total AOD forecast error for Weeks 3&4 (P7.2 minus MODIS). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

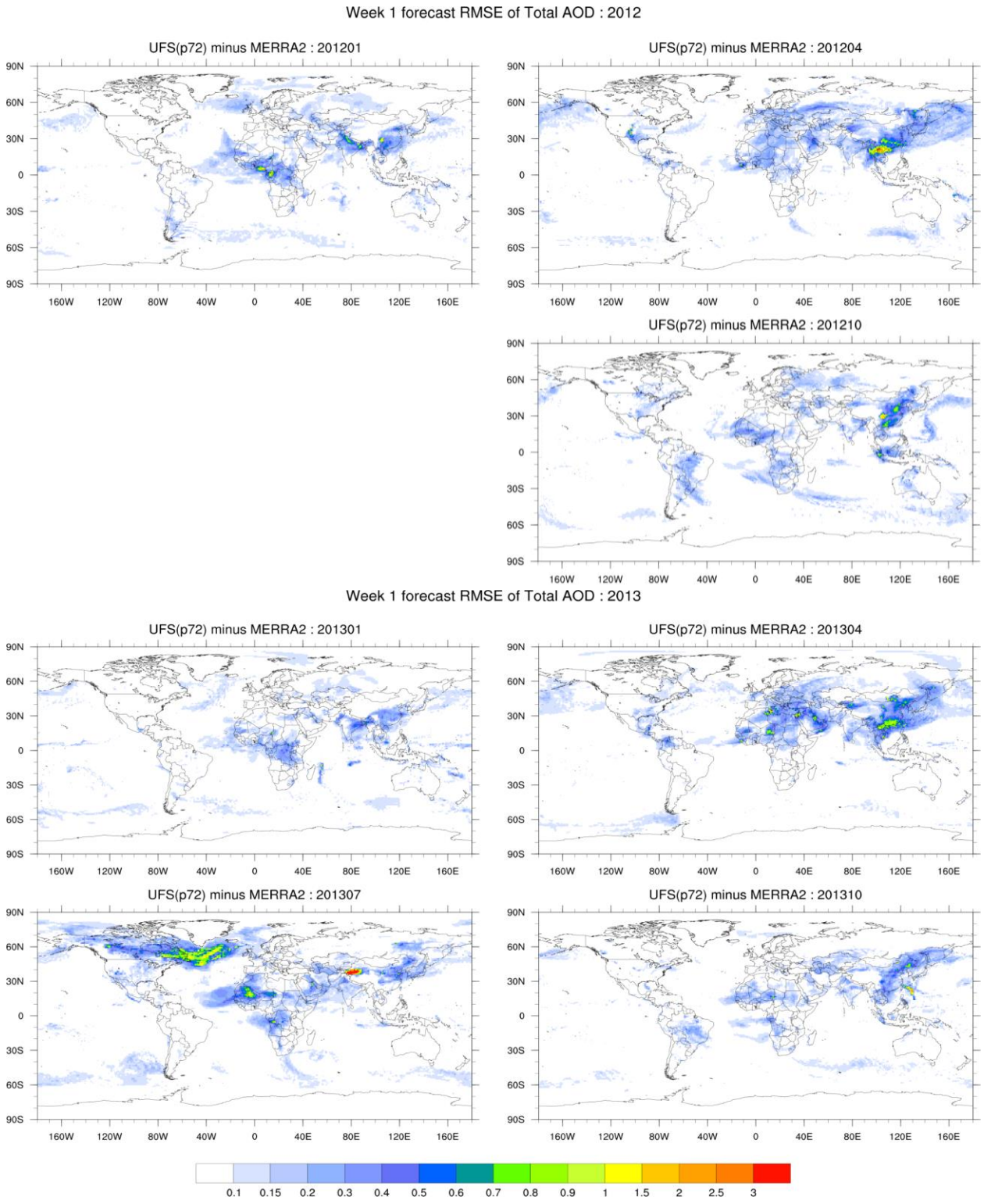

Figure 11.42: Total AOD forecast RMSE for Week 1 (P7.2 vs MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

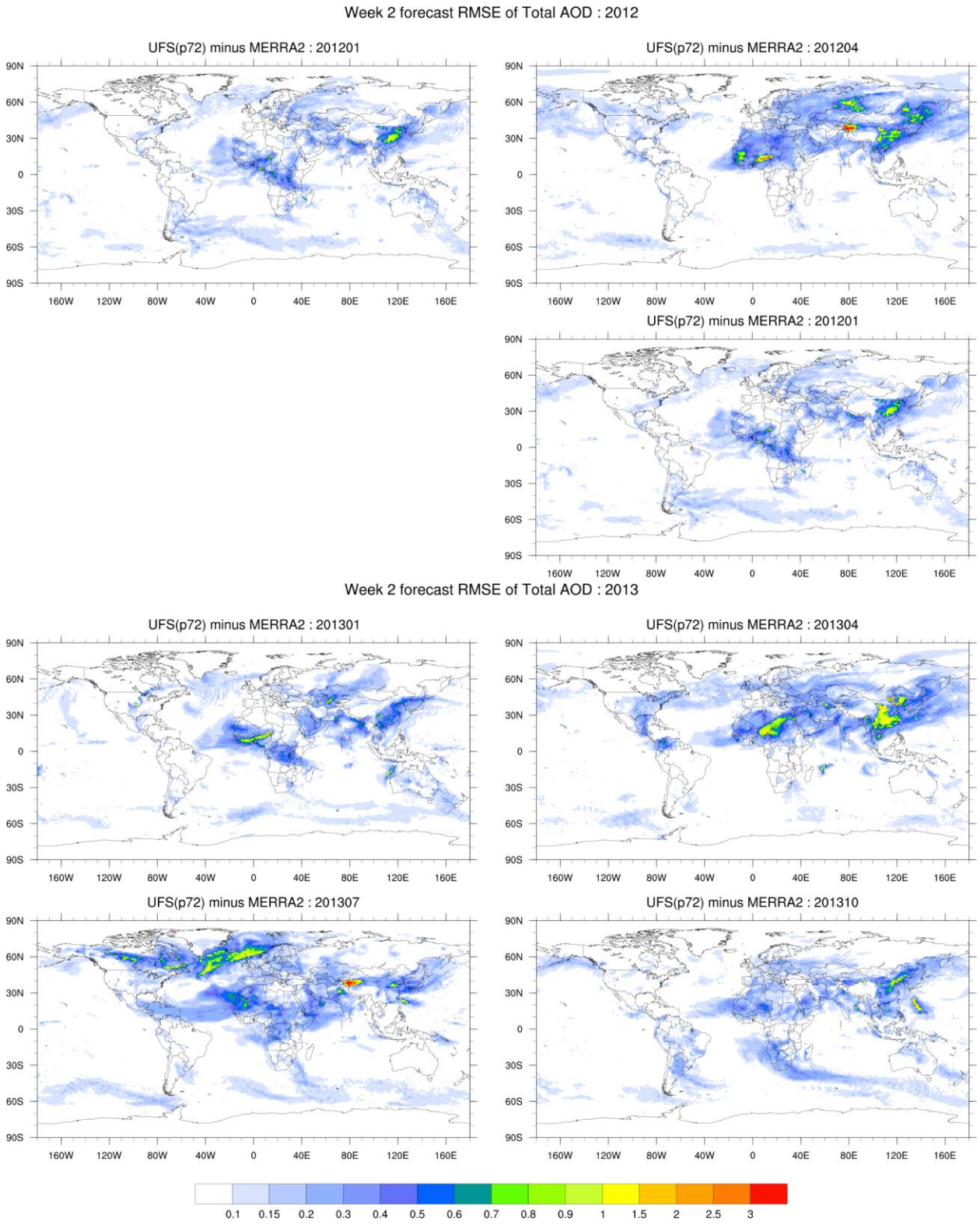

Figure 11.43: Total AOD forecast RMSE for Week 2 (P7.2 vs MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

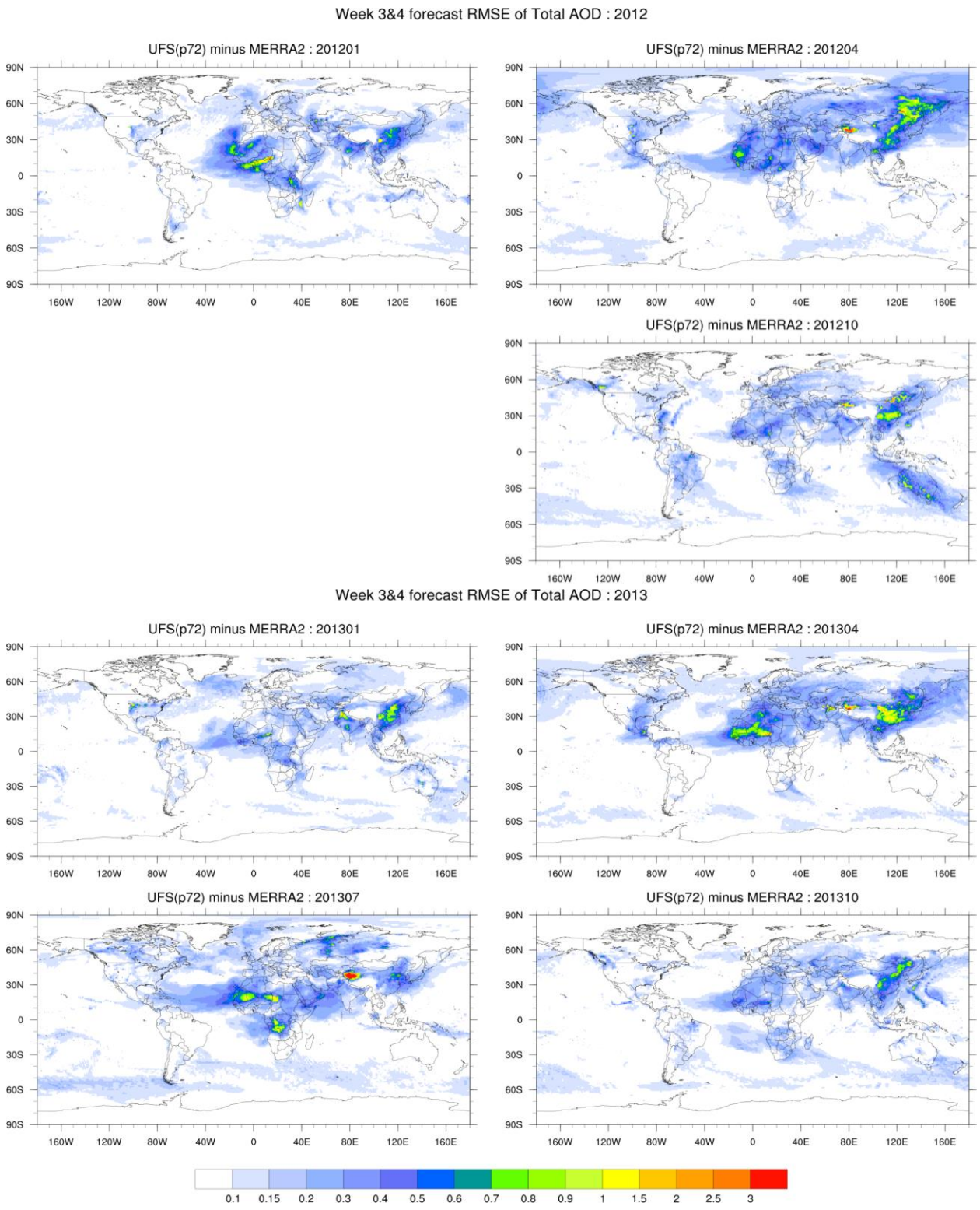

Figure 11.44: Total AOD forecast RMSE for Weeks 3&4 (P7.2 vs MERRA-2). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

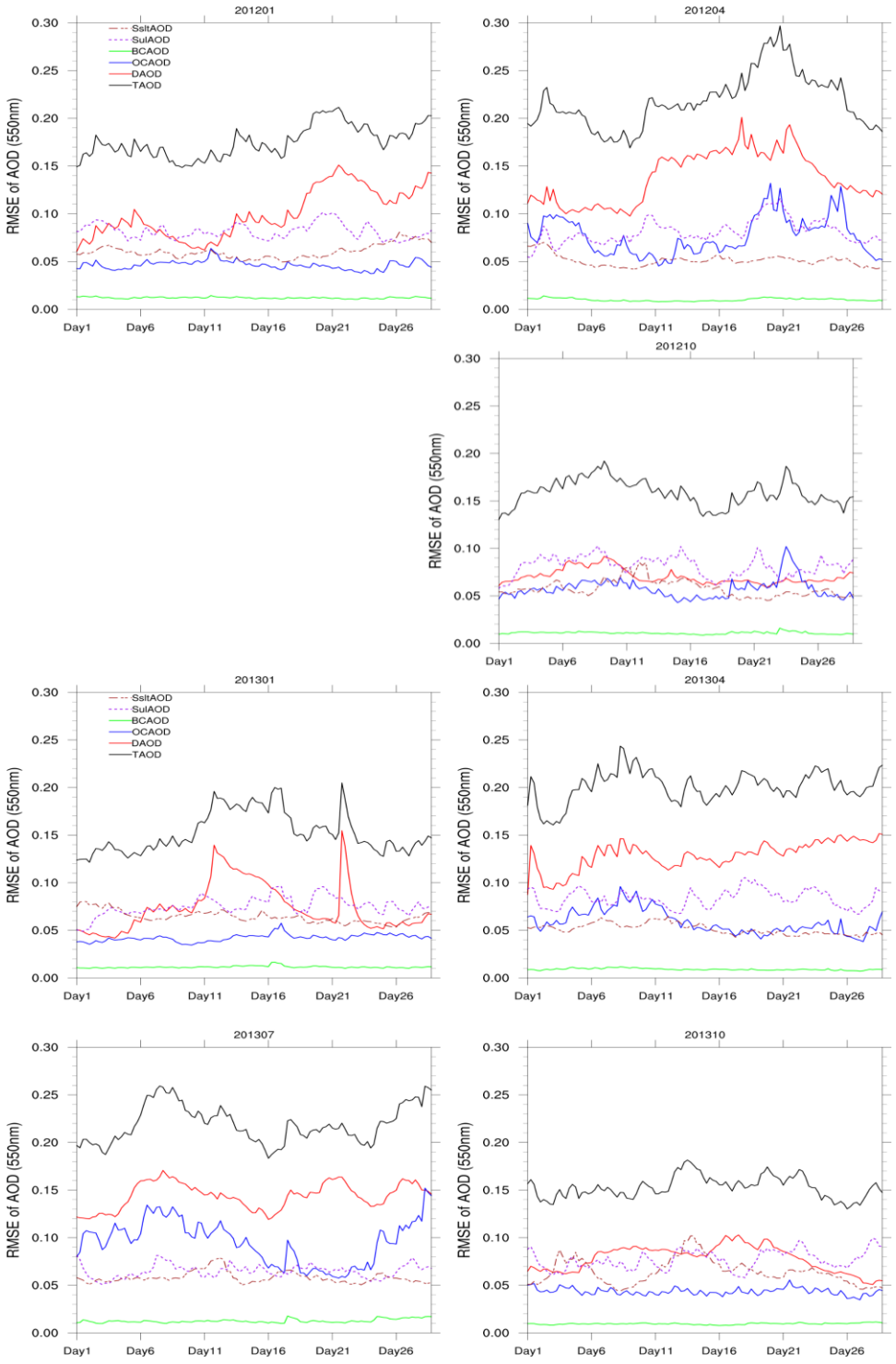

Figure 11.45: Global RMSE for forecast days 1–28 (P7.2 vs MERRA-2) for the AOD of: sea salt (brown), sulfates (purple), black carbon (green), organic carbon (blue), dust (red), and total (black). Forecasts initialized on 01 Jan, Apr, Oct of 2012 (top three panels) and 01 Jan, Apr, Jul, Oct of 2013 (bottom four panels).

Summary Tables and Figures

Table 12.1: Annually aggregated RAW Benchmark AC scores [%] for week 1, week 2, and weeks 3&4 forecasts from CFSv2 and P1–P7. The highest score(s) in each line are in bold red, and the lowest—in bold blue. Prototype scores exceeding the CFSv2 are marked with ↑ and those below CFSv2—with ↓. Ties with the CFSv2 score are marked with ↕. Note that the SST scores shown for P7 (in italic) are calculated using the NSST instead of the actual SST, and are thus somewhat inaccurate.

Week 1, RAW	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
SST, Nino3.4	91.0	**90.5** ↓	**92.9** ↑	92.8 ↑	92.6 ↑	92.6 ↑	92.4 ↑	*92.0* ↑
SST, Global Tropics	**76.0**	78.0 ↑	**80.6** ↑	80.4 ↑	79.8 ↑	79.8 ↑	78.3 ↑	*79.4* ↑
Prate, Nino 3.4	**54.4**	55.2 ↑	57.1 ↑	57.4 ↑	56.8 ↑	56.8 ↑	55.6 ↑	**60.6** ↑
T2m, CONUS Land	**79.1**	**81.3** ↑	**81.3** ↑	80.9 ↑	79.4 ↑	80.4 ↑	79.7 ↑	**79.1** ↕
Z500, NH	**95.0**	96.3 ↑	96.3 ↑	**96.4** ↑	**96.4** ↑	**96.4** ↑	96.2 ↑	95.6 ↑
Z500, SH	**93.6**	94.9 ↑	94.9 ↑	94.9 ↑	94.9 ↑	94.9 ↑	94.8 ↑	**95.5** ↑
Prate, CONUS Land	**48.2**	53.5 ↑	53.4 ↑	53.6 ↑	53.8 ↑	**54.0** ↑	53.1 ↑	53.4 ↑

Week 2, RAW	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
SST, Nino3.4	**86.0**	87.2 ↑	**89.1** ↑	88.5 ↑	88.0 ↑	88.1 ↑	87.8 ↑	*88.7* ↑
SST, Global Tropics	**66.6**	68.2 ↑	**70.3** ↑	69.7 ↑	68.4 ↑	68.2 ↑	65.0 ↓	*68.6* ↑
Prate, Nino 3.4	**42.0**	42.0 ↕	**47.4** ↑	46.7 ↑	44.9 ↑	45.3 ↑	44.6 ↑	46.2 ↑
T2m, CONUS Land	40.7	**42.6** ↑	41.3 ↑	39.0 ↓	**35.8** ↓	40.1 ↓	42.1 ↑	40.3 ↓
Z500, NH	**49.6**	54.8 ↑	54.7 ↑	54.1 ↑	**54.9** ↑	**54.9** ↑	54.3 ↑	51.2 ↑
Z500, SH	**46.6**	51.3 ↑	50.5 ↑	50.1 ↑	48.8 ↑	49.2 ↑	48.2 ↑	**52.4** ↑
Prate, CONUS Land	16.9	**20.4** ↑	19.2 ↑	19.1 ↑	**16.8** ↓	19.2 ↑	18.5 ↑	16.9 ↕

Weeks 3&4, RAW	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
SST, Nino3.4	84.8	85.5 ↑	**86.0** ↑	85.3 ↑	**84.0** ↓	**84.0** ↓	84.1 ↓	*84.4* ↓
SST, Global Tropics	57.3	59.4 ↑	**61.5** ↑	60.5 ↑	58.8 ↑	58.4 ↑	**54.6** ↓	*58.8* ↑
Prate, Nino 3.4	48.0	**46.9** ↓	**50.0** ↑	47.9 ↓	47.6 ↓	47.4 ↓	50.0 ↑	49.4 ↑
T2m, CONUS Land	10.7	17.5 ↑	17.5 ↑	17.6 ↑	**21.9** ↑	10.4 ↓	17.1 ↑	**10.2** ↓
Z500, NH	12.7	17.8 ↑	**18.5** ↑	15.0 ↑	16.6 ↑	13.7 ↑	14.3 ↑	**12.3** ↓
Z500, SH	**9.7**	10.9 ↑	**14.3** ↑	10.9 ↑	12.3 ↑	13.1 ↑	13.1 ↑	11.6 ↑
Prate, CONUS Land	3.2	2.2 ↓	4.4 ↑	5.1 ↑	6.0 ↑	4.6 ↑	**7.6** ↑	**1.4** ↓

Table 12.2: As table 12.1 but for SEC Benchmark AC scores.

Week 1, SEC	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
SST, Nino3.4	92.0	**91.8** ↓	**94.1** ↑	**94.1** ↑	94.0 ↑	94.0 ↑	94.0 ↑	*93.8* ↑
SST, Global Tropics	**80.8**	83.2 ↑	**88.2** ↑	**88.2** ↑	88.0 ↑	88.0 ↑	87.6 ↑	*87.2* ↑
Prate, Nino 3.4	**56.7**	57.2 ↑	59.0 ↑	59.2 ↑	58.7 ↑	58.7 ↑	58.2 ↑	**63.0** ↑
T2m, CONUS Land	87.3	87.6 ↑	88.0 ↑	87.8 ↑	87.8 ↑	87.8 ↑	87.3 ↕	**86.8** ↓
Z500, NH	**95.9**	**96.7** ↑	**96.7** ↑	**96.7** ↑	**96.7** ↑	**96.7** ↑	**96.7** ↑	96.2 ↑
Z500, SH	**94.8**	95.5 ↑	95.4 ↑	95.3 ↑	95.4 ↑	95.3 ↑	95.3 ↑	**96.0** ↑
Prate, CONUS Land	**52.4**	56.3 ↑	56.3 ↑	56.6 ↑	56.8 ↑	56.9 ↑	55.8 ↑	56.2 ↑

Week 2, SEC	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
SST, Nino3.4	**88.4**	89.4 ↑	**92.1** ↑	91.8 ↑	91.8 ↑	91.7 ↑	91.7 ↑	*91.3* ↑
SST, Global Tropics	**73.7**	76.7 ↑	**80.6** ↑	80.4 ↑	79.9 ↑	79.9 ↑	78.6 ↑	*78.5* ↑
Prate, Nino 3.4	44.6	**44.2** ↓	**49.3** ↑	48.6 ↑	46.7 ↑	47.1 ↑	46.8 ↑	49.1 ↑
T2m, CONUS Land	46.1	46.9 ↑	46.5 ↑	43.9 ↓	**41.3** ↓	46.0 ↓	**47.9** ↑	46.1 ↕
Z500, NH	**52.4**	**56.9** ↑	56.8 ↑	56.0 ↑	56.8 ↑	56.7 ↑	56.6 ↑	53.3 ↑
Z500, SH	**49.4**	53.5 ↑	52.7 ↑	52.2 ↑	50.7 ↑	51.3 ↑	50.2 ↑	**54.3** ↑
Prate, CONUS Land	18.6	**21.9** ↑	20.6 ↑	20.4 ↑	**18.1** ↓	20.6 ↑	19.8 ↑	18.3 ↓

Weeks 3&4, SEC	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
SST, Nino3.4	**89.1**	89.9 ↑	**92.2** ↑	92.0 ↑	91.6 ↑	91.5 ↑	91.8 ↑	*90.9* ↑
SST, Global Tropics	**67.9**	70.8 ↑	**74.7** ↑	74.5 ↑	73.9 ↑	73.5 ↑	71.2 ↑	*71.9* ↑
Prate, Nino 3.4	51.6	50.3 ↓	53.0 ↑	50.7 ↓	**50.2** ↓	50.4 ↓	52.4 ↑	**54.4** ↑
T2m, CONUS Land	13.5	20.0 ↑	20.7 ↑	20.7 ↑	**27.0** ↑	**12.0** ↓	20.7 ↑	12.1 ↓
Z500, NH	14.3	19.2 ↑	**20.0** ↑	15.9 ↑	17.8 ↑	14.5 ↑	15.4 ↑	**13.4** ↓
Z500, SH	**11.1**	11.7 ↑	**15.5** ↑	11.8 ↑	13.5 ↑	14.2 ↑	14.6 ↑	12.5 ↑
Prate, CONUS Land	3.6	2.4 ↓	4.8 ↑	5.6 ↑	6.5 ↑	5.0 ↑	**8.4** ↑	**1.5** ↓

Table 12.3: Approximate number of forecast days to RMM1/RMM2/RMM AC dropping below 0.6 in CFSv2 and coupled UFS prototypes P1–P7. The highest number(s) of each line are in bold red, and the lowest—in bold blue. Durations exceeding that of the skillful CFSv2 forecast are marked with ↑.

Number of days to AC<0.6	CFSv2	P1	P2	P3.1	P4	P5	P6	P7
RMM1	**10.5**	13.5 ↑	13.5 ↑	15.5 ↑	14 ↑	14 ↑	14 ↑	**16** ↑
RMM2	**15**	18 ↑	**20** ↑	19 ↑	19 ↑	19 ↑	17.5 ↑	18.5 ↑
RMM (RMM1&2)	**12**	17 ↑	**19** ↑	18 ↑	18 ↑	16 ↑	17 ↑	18.5 ↑

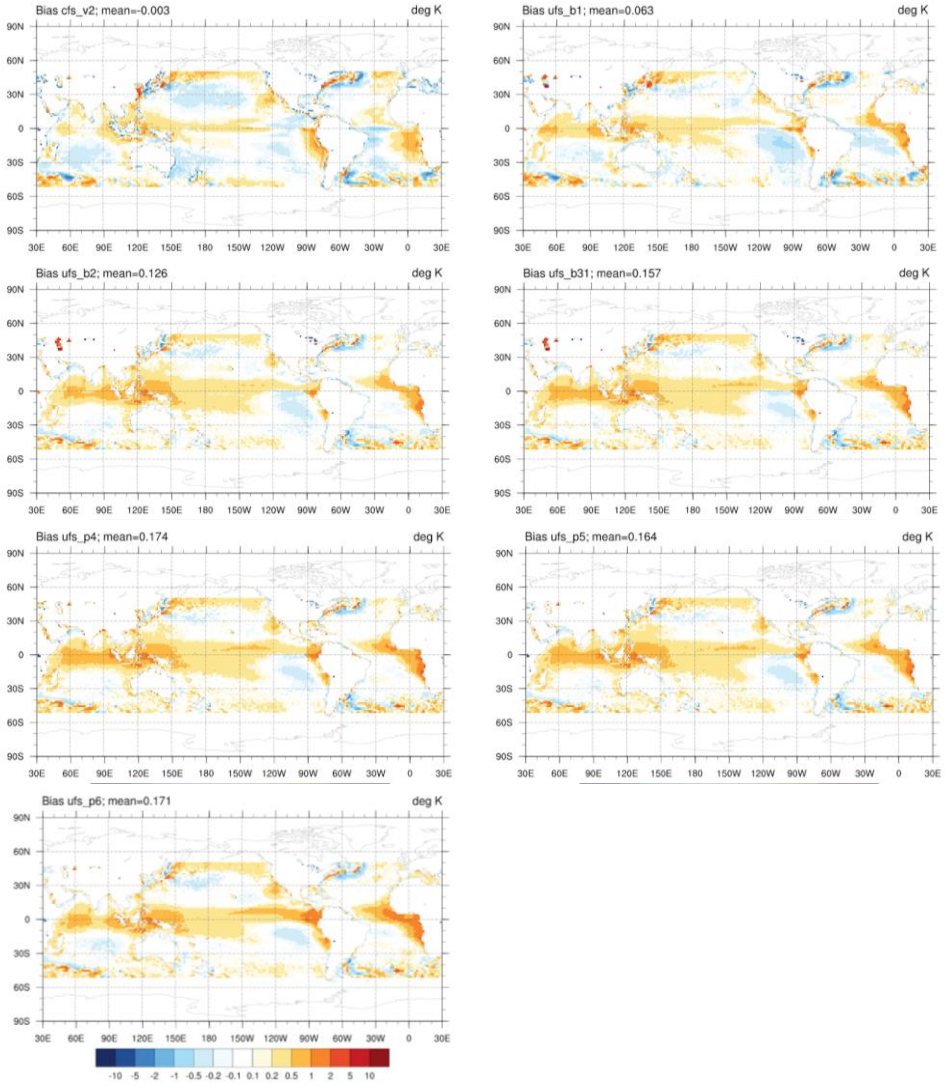

Figure 12.1: Annual average week 3&4 SST (50°S–50°N) bias wrt OSTIA in: CFSv2 (top left), P1 (top right), P2 (second row, left), P3.1 (second row, right), P4 (third row, left), P5 (third row, right), P6 (last row, left). Results for P7 are not shown here because the above plots are created using atmospheric-side output, which, in the case of P7, contains NSST and not SST. For P7 SST bias in comparison with P6 and P5, see Figs 9.15 and 9.16.

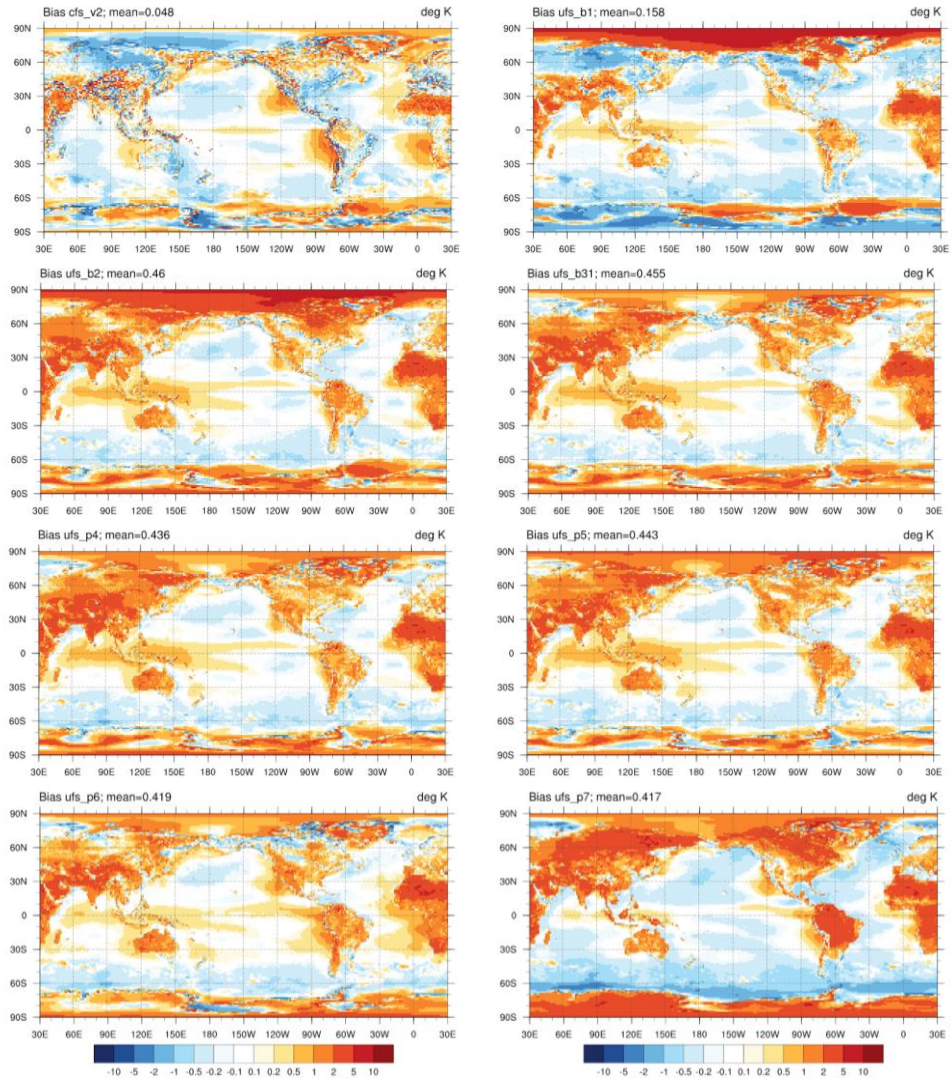

Figure 12.2: Annual average week 3&4 2-m air temperature bias wrt CFSR in: CFSv2 (top left), P1 (top right), P2 (second row, left), P3.1 (second row, right), P4 (third row, left), P5 (third row, right), P6 (last row, left), P7 (last row, right)

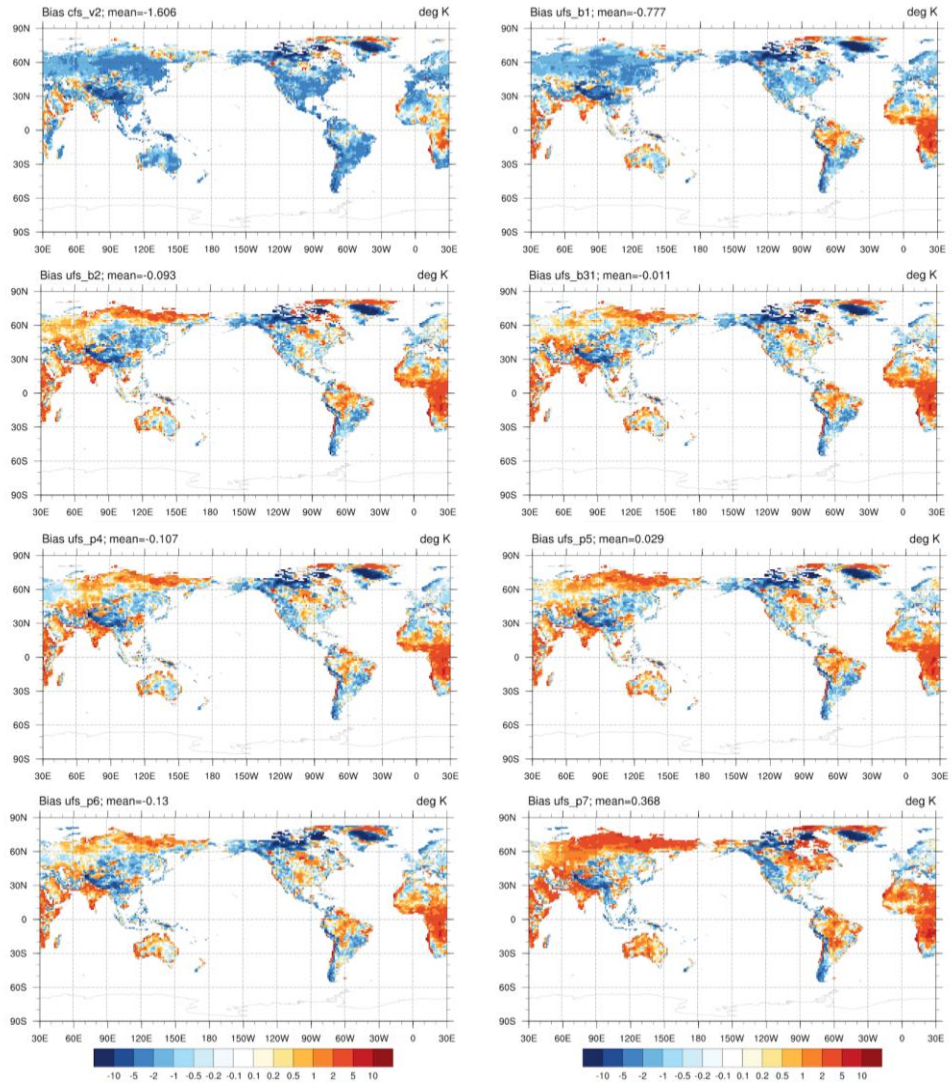

Figure 12.3: Annual average week 3&4 daily maximum 2-m air temperature bias wrt CPC in: CFSv2 (top left), P1 (top right), P2 (second row, left), P3.1 (second row, right), P4 (third row, left), P5 (third row, right), P6 (last row, left), P7 (last row, right)

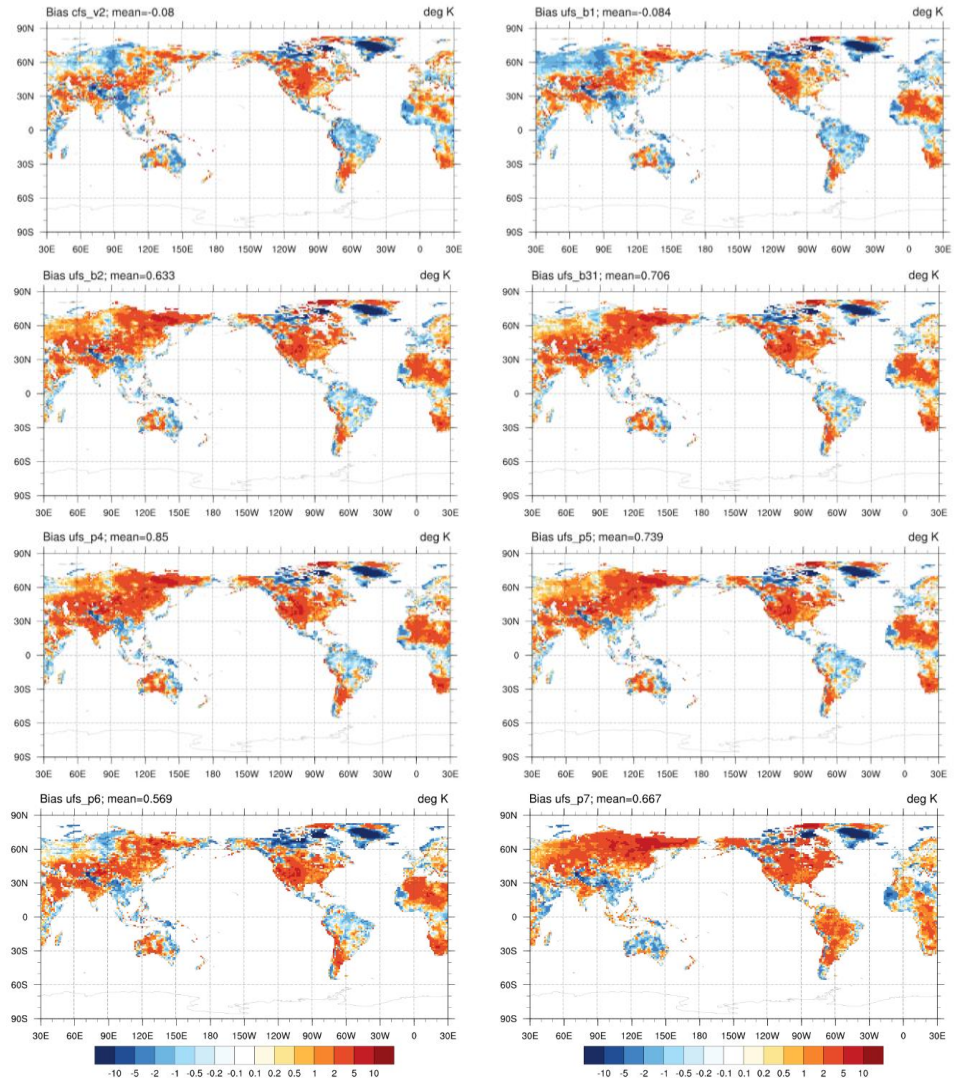

Figure 12.4: Annual average week 3&4 daily minimum 2-m air temperature bias wrt CPC in: CFSv2 (top left), P1 (top right), P2 (second row, left), P3.1 (second row, right), P4 (third row, left), P5 (third row, right), P6 (last row, left), P7 (last row, right)

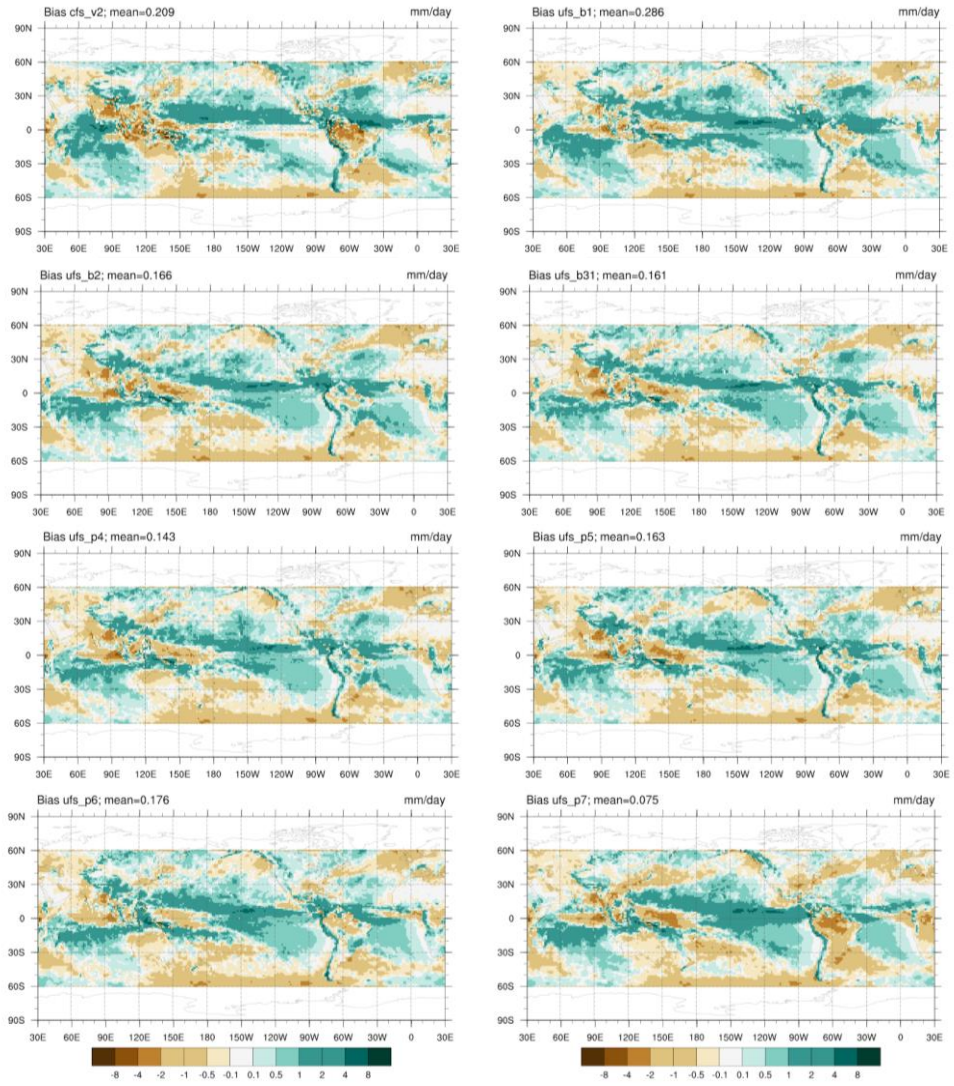

Figure 12.5: Annual average week 3&4 daily precipitation rate bias wrt IMERG in: CFSv2 (top left), P1 (top right), P2 (second row, left), P3.1 (second row, right), P4 (third row, left), P5 (third row, right), P6 (last row, left), P7 (last row, right)

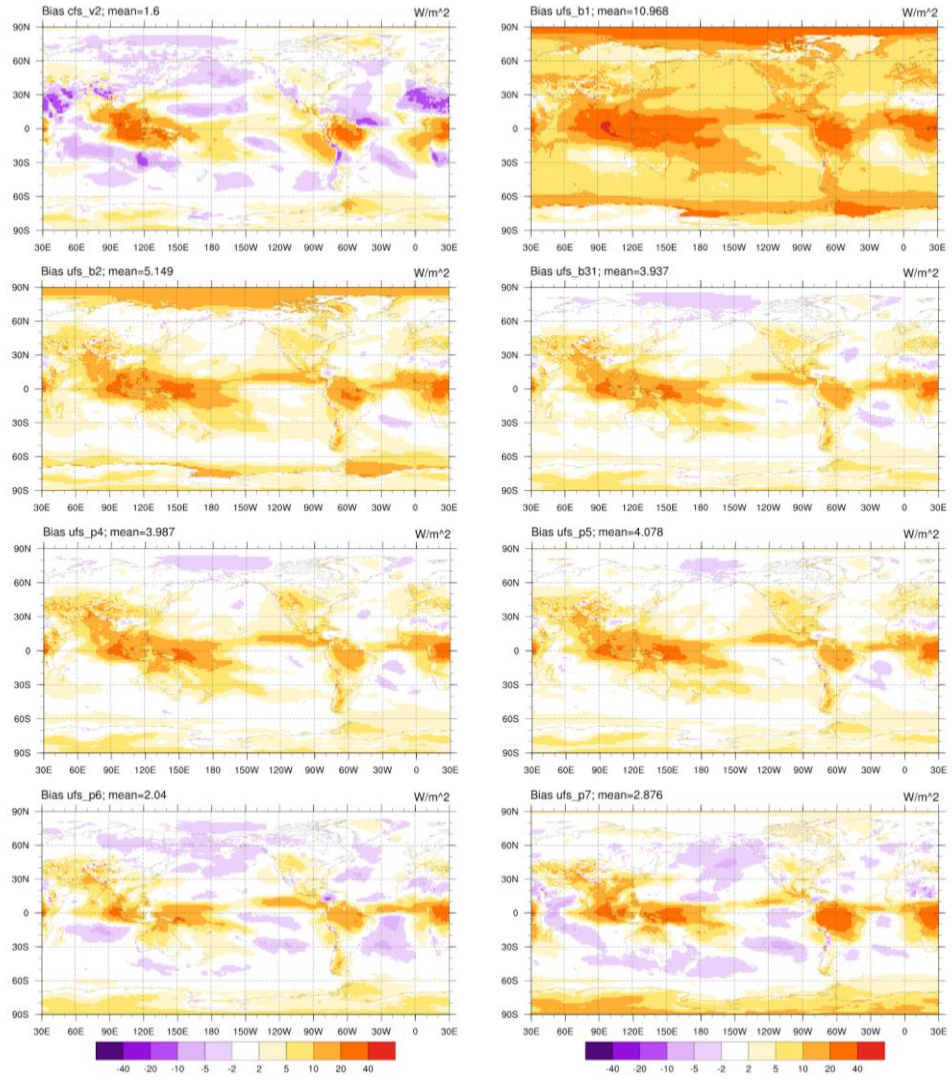

Figure 12.6: Annual average week 3&4 OLR bias wrt NOAA-CDR in: CFSv2 (top left), P1 (top right), P2 (second row, left), P3.1 (second row, right), P4 (third row, left), P5 (third row, right), P6 (last row, left), P7 (last row, right)

Appendix 1: Reference data sets

The datasets used in prototype evaluations as either sources of observations or proxy for observations include the following:

Table A1.1 Reference datasets used for validation and verification

Type	Dataset	Variable(s)	Link	Reference
Gridded datasets derived from in situ data	ARGO	mixed layer depth	http://mixedlayer.ucsd.edu/data/Argo_mixedlayers_monthlyclim_12112019.nc	Holte *et al.* (2017)
	CPC Global 0.5 Degree Maximum and Minimum Daily Temperature	daily min/max 2-meter air temperature	ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/	Shi (2019)
	CPC Global 0.5 Degree Unified Rain Gauge data	precipitation	ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/	Xie *et al.* (2010)
Gridded datasets derived from remote sensing	AVISO	sea surface height	https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-sea-level-global	Taburet *et al*. (2021)
	CERES EBAF	monthly surface radiative fluxes	https://ceres.larc.nasa.gov/data/#ebaf-level-3	Kato *et al.* (2018)
	CERES SYN1°	daily surface radiative fluxes	https://ceres.larc.nasa.gov/data/#syn1deg-level-3	Doelling *et al.* (2016)
	GLEAM	evapotranspiration	https://gmd.copernicus.org/articles/10/1903/2017	Martens *et al.* (2017)

	IMERG	precipitation	https://doi.org/10.5067/GPM/IMERGDF/DAY/06	Huffman *et al.* (2019)
	MODIS	aerosol optical depth	https://atmosphere-imager.gsfc.nasa.gov/products/daily	Levi *et al.* (2013)
	OSCAR	pentad surface ocean currents	https://doi.org/10.5067/OSCAR-03D01	Bonjean and Lagerloef (2002)
	NASA-Team	ice concentration	https://nsidc.org/data/nsidc-0051	Cavalieri *et al.* (1996)
	NOAA-CDR	outgoing longwave radiation	https://www.ncei.noaa.gov/data/outgoing-longwave-radiation-daily/access/	Lee and NOAA CDR Program, (2011)
	NOAA/NSDIC-CDR version 3	ice concentration	https://nsidc.org/data/G02202/versions/3	Meier *et al.* (2017)
	SMAP	surface ocean salinity	ftp://podaac.jpl.nasa.gov/allData/smap/	Entekhabi *et al.* (2014)
	SMOS	surface ocean salinity	https://www.cen.uni-hamburg.de/en/icdc/data/ocean/esa-cci-sea-surface-salinity.html#zugang	Mecklenburg *et al.* (2012)
Analyses and reanalyses	CFSR	various atmospheric fields	https://rda.ucar.edu/datasets/ds094.0/	Saha *et al.* (2010)
	ECMWF-ORAS5	various ocean fields	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-oras5	Zuo *et al.* (2018)
	EN4	monthly subsurface	https://www.metoffice.gov.uk/hadobs/en4/	Good *et al.* (2013)

		ocean temperature		
	ERA5	various atmospheric fields	https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5	Hersbach *et al.* (2020)
	GODAS	pentad subsurface ocean currents	https://psl.noaa.gov/data/gridded/data.godas.html	Behringer and Xue (2004)
	MERRA-2	atmospheric optical depth	https://disc.gsfc.nasa.gov/datasets/M2T1NXAER_5.12.4/summary	GMAO (2015)
	OSTIA	sea surface temperature	https://doi.org/10.5067/GHOST-4FK01	Stark *et al.* (2007)
	SPEAR-ECDA	various ocean fields	Provided by GFDL	Lu *et al.* (2020)
In situ observations	ESRL Global Monitoring Laboratory, South Pole Observing Station	land skin temperature & surface radiative fluxes	https://gml.noaa.gov/dv/data/?site=spo	N/A
	Surface Radiation Network (SURFRAD)	land skin temperature	https://gml.noaa.gov/grad/surfrad/	Augustine *et al.* (2000)

Appendix 2: Evaluation of CPC-3DVar Ocean Initial Conditions against Existing Ocean Reanalyses

The CPC-3DVar Ocean Initial Conditions were used for P2–P7. The following analysis was performed using the first day of integration of one of the later prototypes (P5). Due to the minimal change in ocean state over the course of a single day, it is equally valid for all prototypes and reflects the state of the initial conditions.

A2.1 Biases in CPC-3DVar Ocean Initial Conditions

The globally averaged (75°S–60°N) monthly temperature and salinity profiles from the CPC-3DVar ocean ICs averaged for the period April 2011-March 2018 show a slightly warmer and saltier upper ocean (20–50m) than the ocean reanalyses (GFDL's SPEAR-ECDA and ECMWF's ORAS5) (Fig. A2.1). Below 50m, the globally averaged (75°S–60°N) subsurface in CPC-3DVar is saltier than ORAS5, but slightly fresher than ECDA (Fig. A2.1).

Consistent with a warmer upper ocean, the mixed layer depth (MLD) is also shallower in the ICs, with largest differences (shallower) in the Southern Ocean (Fig. A2.2). Individual basin-averages reveal the MLD differences are most prominent in the Indian Ocean. The shallower MLD In the sub-tropics is most prominent in the winter months (OND in NH, and MJJ in SH). Tropical regions show shallower MLD irrespective of seasons, revealing a climatological bias consistent with the double-ITCZ pattern. Differences between the reanalyses are exacerbated in the eastern upwelling boundary regions, and the Southern Indian Ocean revealing inadequate physics or observations in these regions. Large differences in MLD are seen in the western Pacific warm pool (MLD in ECDA > ORAS5 by ~30m)—a region dominated by strong convection and barrier layer formation; and in south-eastern Pacific (MLD in ECDA > ORAS5 by ~15m)—a region commonly affected by biases in simulating stratocumulus clouds. Comparing the spatial pattern of temperatures between the two reanalyses reveal warmer upper 50m (~0.8°C) along the eastern boundary consistent with a shallower MLD (~10m) in ORAS5, and a warmer upper 50m in the Southern Indian Ocean consistent with a shallower MLD in ECDA (Fig. A2.3).

Below 100m, the differences between the two reanalyses' temperatures are most prominent at high latitudes in the Southern Ocean, and North Atlantic (~50°N), with ECDA showing warmer temperatures in the eastern and western ends of the North Atlantic subpolar gyre, around the Agulhas and Falkland Current systems compared to ORAS5 (Fig. A2.3). These are prominent eddy formation regions largely driven by the wind driven circulations, and are poorly resolved in climate models.

Comparing the spatial patterns in salinity reveals that the ocean ICs are largely fresher than the reanalyses in the upper 50m, but saltier in regions of source of freshwaters, such as: the

Maritime Continent, along the ITCZ, along eastern equatorial coastal upwelling, and along the northern boundaries of South America in the tropical Atlantic (Fig. A2.4). The region of deep water formation in the North Atlantic has fresh bias compared to both reanalyses, which could indicate less sinking and less deep water formation in the ocean ICs. However, disagreements between the reanalyses are also evident, with saltier waters in ECDA in regions north of Kuroshio region —an active typhoon formation region in western Pacific; north-west Atlantic—deep water formation region and Gulf Stream, western tropical north Atlantic—an active region of tropical cyclone and freshwater input from Amazon basin rivers, the Maritime Continent and warm pool regions, and along the Antarctic continental boundaries (bottom panel of Fig. A2.4). It is possible that compared to ORAS5, ECDA might not be simulating the freshwaters from rivers into the open ocean correctly. The spatial differences between ocean ICs and the reanalyses also hold at larger depths (100–400m), although the magnitude of the differences reduce.

A2.2 Temporal Variability in CPC-3DVar Ocean Initial Conditions

Temporal variability in the ocean ICs is assessed via correlations in temperature and salinity fields from ECDA and ORAS5 for the period 2011–2018. MLD correlations show largest correlations (>0.8) in the tropical Pacific and eastern subtropical Pacific, and smallest correlations (<0.2) in the tropical Atlantic, Southern Ocean, Kuroshio and Gulf Stream region, and Indian Ocean relative to either reanalysis (Fig. A2.5). Correlations are better with ORAS5 in the southeastern tropical Pacific and equatorial Indian Ocean, where the two reanalyses agree poorly among themselves. However in regions where the two reanalyses are in relative disagreement (correlated by <0.5), such as the tropical North Atlantic, the extra-tropical South Pacific, and the North Pacific, the ocean ICs show weak correlations to either reanalyses, indicating room for improvement in these regions.

The upper 300m heat content of individual basin averages in the ocean ICs generally agrees well with both reanalyses (not shown). After detrending the heat content time-series the correlations with either reanalysis remain very high in the Pacific, but are low for all other basin averages (Fig. A2.6). This indicates that much of the change in upper ocean heat content over the period 2011–2018 is related to the global warming related changes that are prominent in the Atlantic, Indian Ocean and Southern Ocean. On detrending, the upper ocean heat content of the ocean ICs is closer to that of ECDA than that of ORAS5, with poorest correlations with ORAS5 in the Southern Ocean.

The correlations of depth-averaged temperatures in ocean ICs with both reanalyses decrease with depth. Spatially, except for the tropical Indo-Pacific, correlations drastically decrease in 100–200m averaged temperatures from the SST, and further decrease in 200–300m averages (not shown). Correlations of depth-averaged temperatures in ocean ICs with either reanalysis are stronger in the off-equatorial regions of western Pacific, and the northwestern tropical Atlantic, likely due to the correct positioning of the Gulf Stream. Zonally averaged sections show high correlations for the tropical subsurface temperatures to 500m depth (Fig. A2.7). However,

moving polewards, higher correlations are mostly restricted to the upper 100–200m depths in the southern hemisphere. Northward of ~50°N higher correlations extend below 500m, reflecting the region of deep mixing in the North Atlantic. The temperature correlations are generally higher in regions of sub-tropical subduction zones, and regions of deep mixing. In general the temperatures at depths in ocean ICs agree better with ECDA than with ORAS5. Meridional averaged sections show larger depths of high correlations in the Pacific basin, deepening eastward, and shallower depths of high correlations in the Indian Ocean and Atlantic (Fig. A2.8).

Salinity correlations are generally lower than temperature correlations, indicating wide disparity in variability from the surface to deeper ocean (Fig. A2.9, A2.10). Regions dominated by evaporation–precipitation processes are mostly where the salinity correlations are the poorest between the ocean ICs and either reanalyses. Zonally averaged sections show deeper correlations in the tropics, and shallower depths correlations at the equator and high latitudes. Meridional sections show much poorer correlations below 200m.

A2.3 Atlantic MOC in CPC-3DVar Ocean Initial Conditions

On longer timescales, reconstruction of the meridional overturning circulation is assessed on computing the stream function in the ocean ICs and comparing with that from the reanalyses. Although the period of 7 years (2011–2018) is too short given the time-scale of Atlantic Meridional Overturning Circulation (AMOC), such comparison provides a glimpse of the initialization of large scale climate indices used in the ocean ICs for the subseasonal forecasts. The mean MOC stream function of the North Atlantic shows a maximum northward transport near 25°N and 35°N–40°N in ORAS5 , whereas the maximum is between 35°N–40°N in ECDA (Fig. A2.11). The northward transport in the ocean ICs (CPC-3DVar) is indeed as in the reanalyses, although weaker, with multiple maxims between 16°N–22°N, 25°N and 30°N–35°N.

A2 Figures

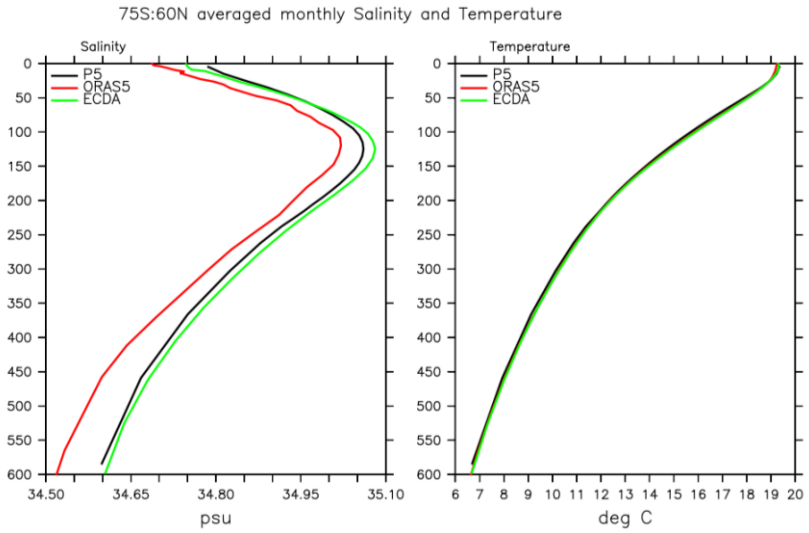

Figure A2.1: Globally averaged (75°S-60°N) mean salinity and temperature profiles in the upper 600m from Day 1 of P5 and monthly fields from ORAS5 and SPEAR-ECDA for the period April 2011-Mar 2018.

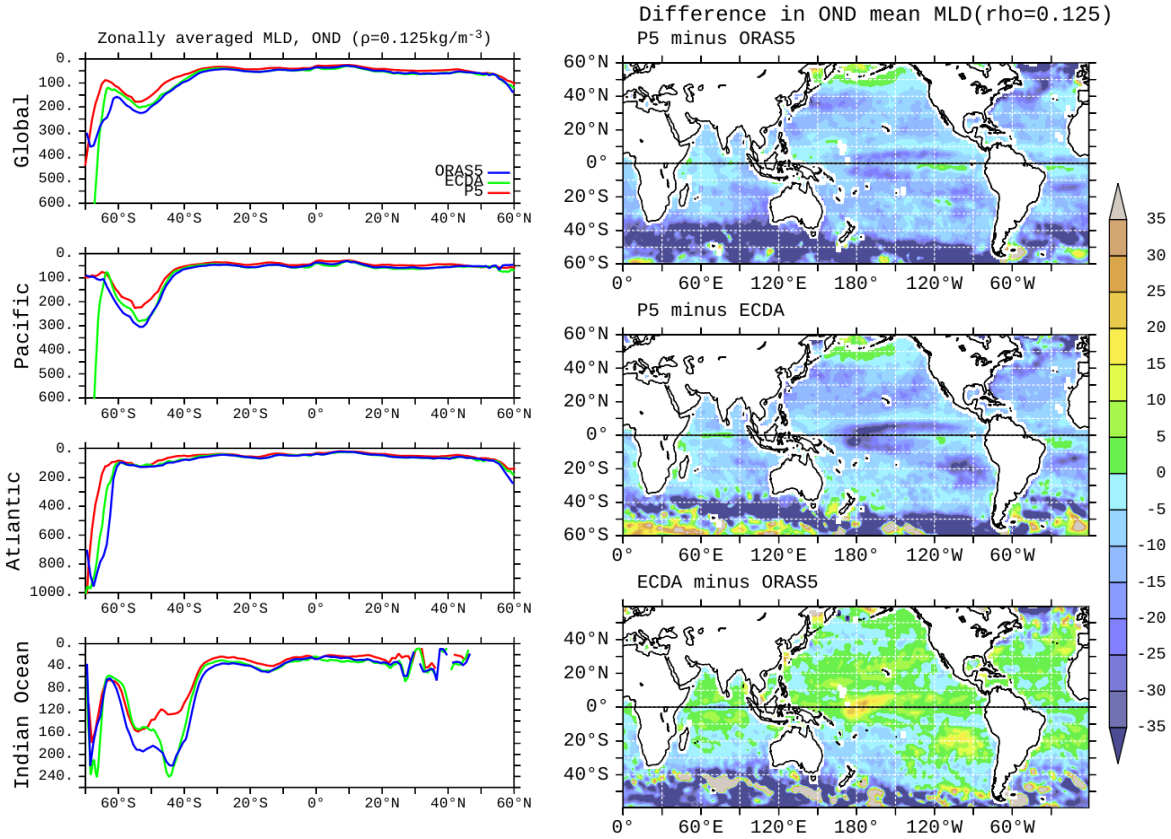

Figure A2.2: Zonally averaged seasonal MLD (OND) for the global ocean, Pacific, Atlantic and Indian Ocean (left). Spatial map of the difference between MLD in ICs and ORAS5 (top), IC and ECDA (middle) and that among the reanalyses (bottom) for OND averages (right).

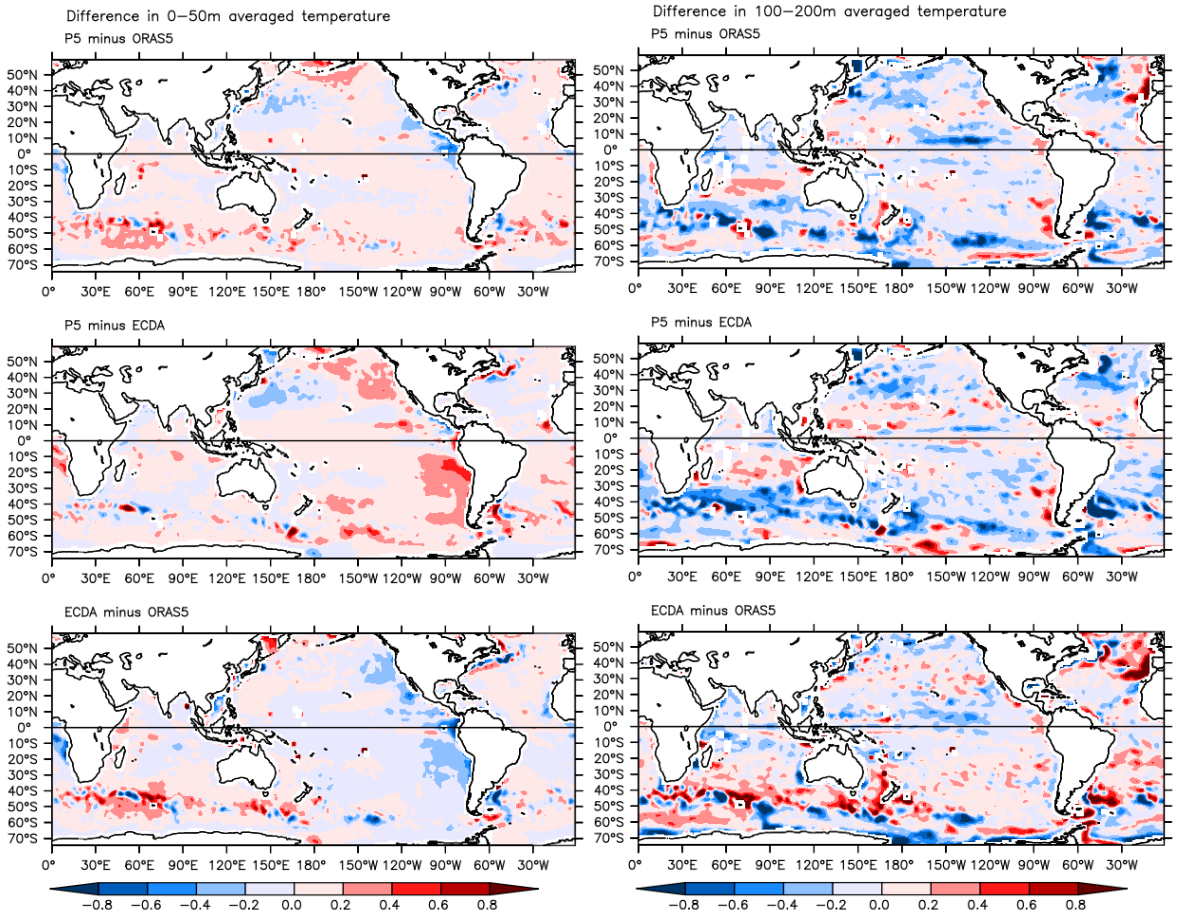

Figure A2.3: Bias in temperature averaged in the upper 50m (left), and 100–200m (right) in ICs from ORAS5 (top), and ECDA (middle), along with the difference between the reanalyses ECDA minus ORAS5 (bottom) for the period April 2011–Mar 2018.

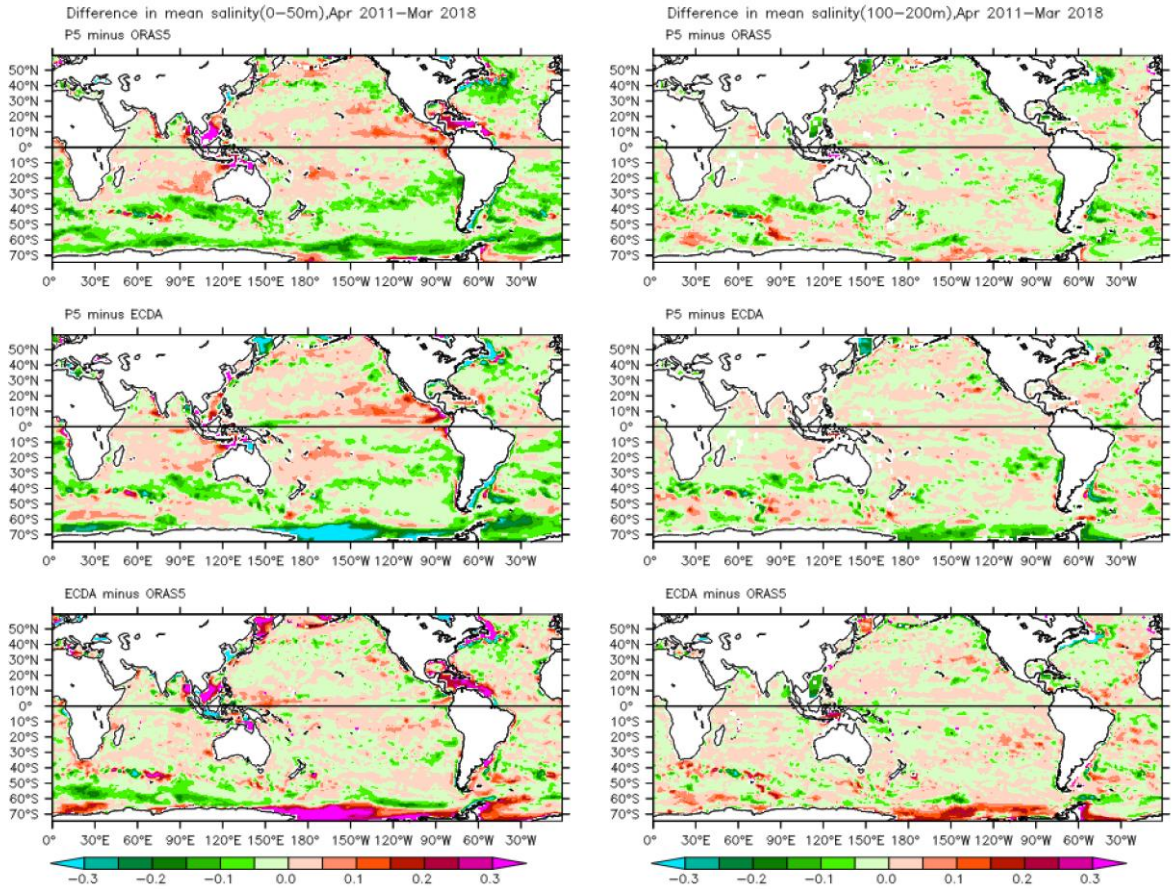

Figure A2.4: Bias in salinity averaged in the upper 50m (left), and 100–200m (right) in ICs from ORAS5 (top), and ECDA (middle), along with the difference between the reanalyses ECDA minus ORAS5 (bottom) for the period April 2011–Mar 2018.

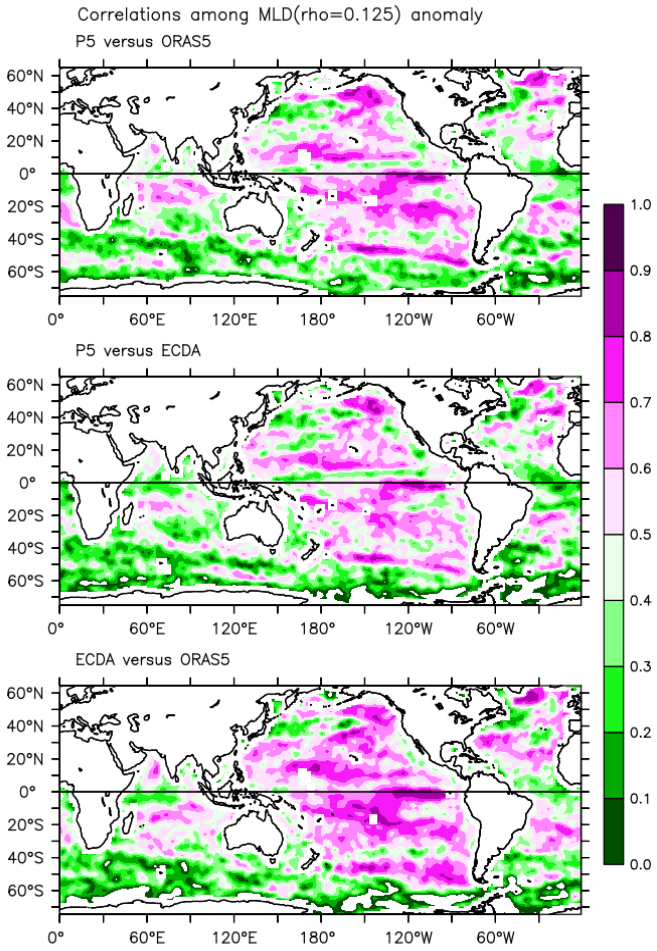

Figure A2.5: Correlation of monthly MLD from ICs with ORAS5 (top), ICs with ECDA (middle), and among the two reanalyses ECDA and ORAS5 (bottom) for the period April 2011–Mar 2018. MLD computed using a density criterion of ρ =0.125km/m^3.

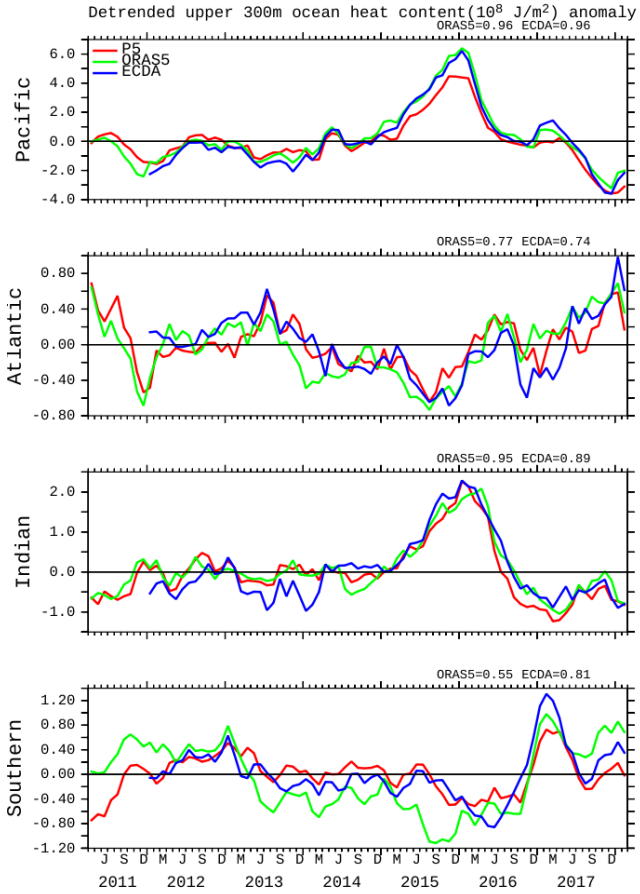

Figure A2.6: Time series of detrended upper 300m ocean heat content in ICs (red), ORAS5 (green) and ECDA (blue) averaged for the Pacific, Atlantic, Indian Ocean and Southern Ocean basins separately covering the period April 2011–Mar 2018.

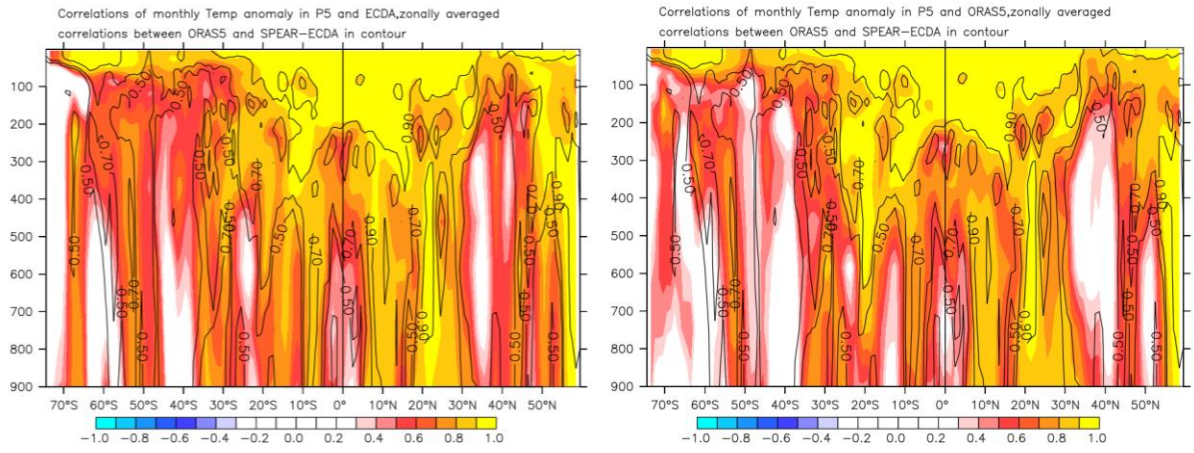

Figure A2.7: Correlation in zonally averaged monthly temperature from ICs and ECDA (left) and ICs and ORAS5 (right), contoured over by the correlation between the two reanalyses.

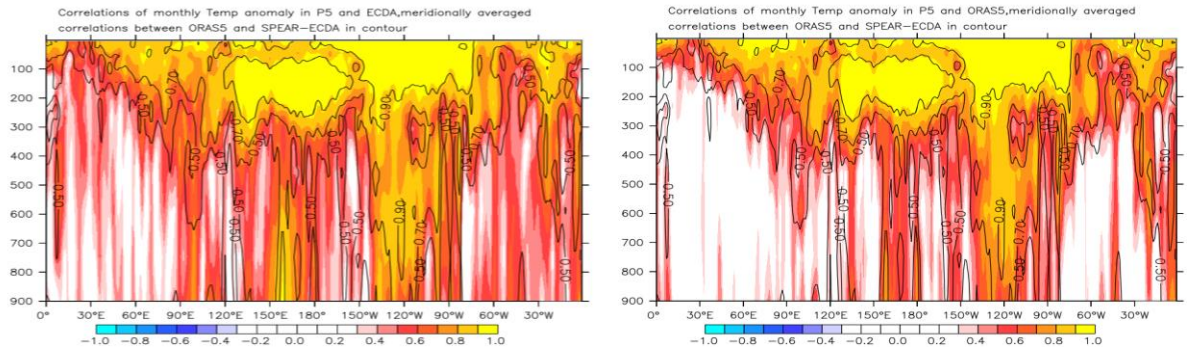

Figure A2.8: Correlation in meridionally averaged over 70°S-60°N monthly temperature from ICs and ECDA (left) and ICs and ORAS5 (right), contoured over by the correlation between the two reanalyses.

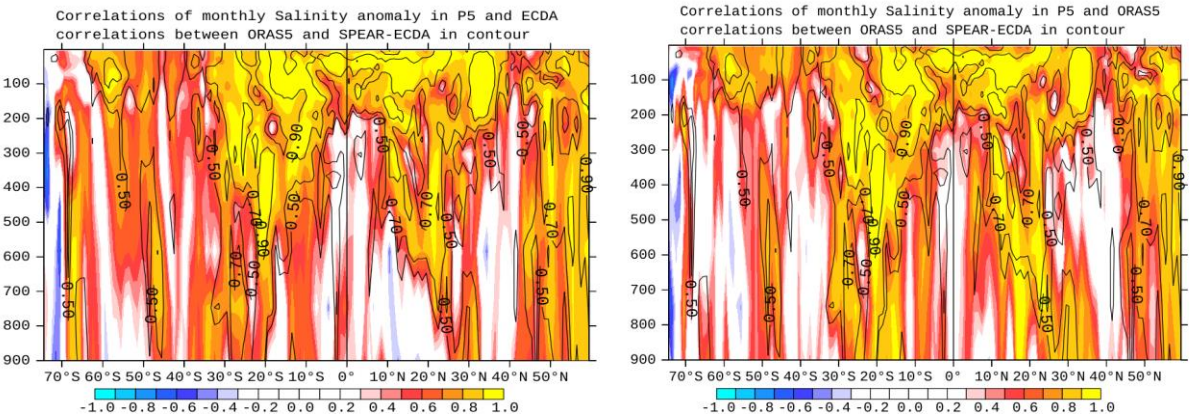

Figure A2.9: Correlation in zonally averaged monthly salinity from ICs and ECDA (left) and ICs and ORAS5 (right), contoured over by the correlation between the two reanalyses.

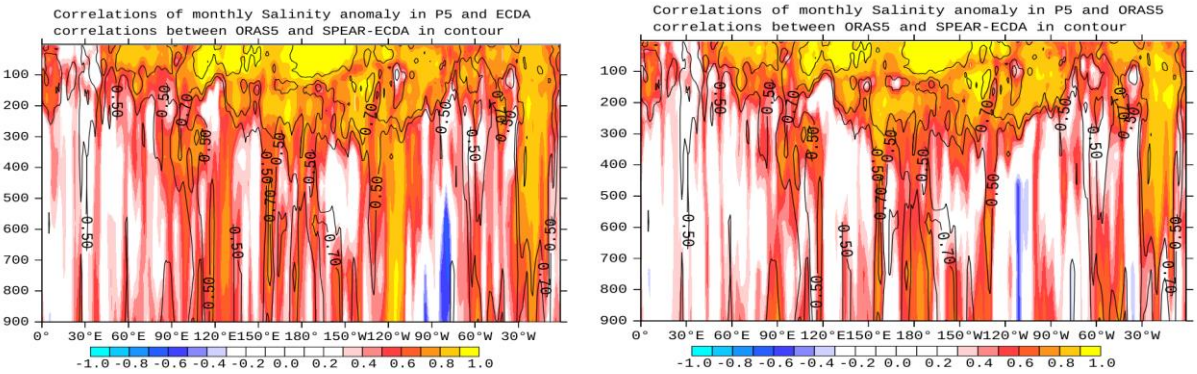

Figure A2.10: Correlation in meridionally averaged over 70°S-60°N monthly salinity from ICs and ECDA (left) and ICs and ORAS5 (right), contoured over by the correlation between the two reanalyses.

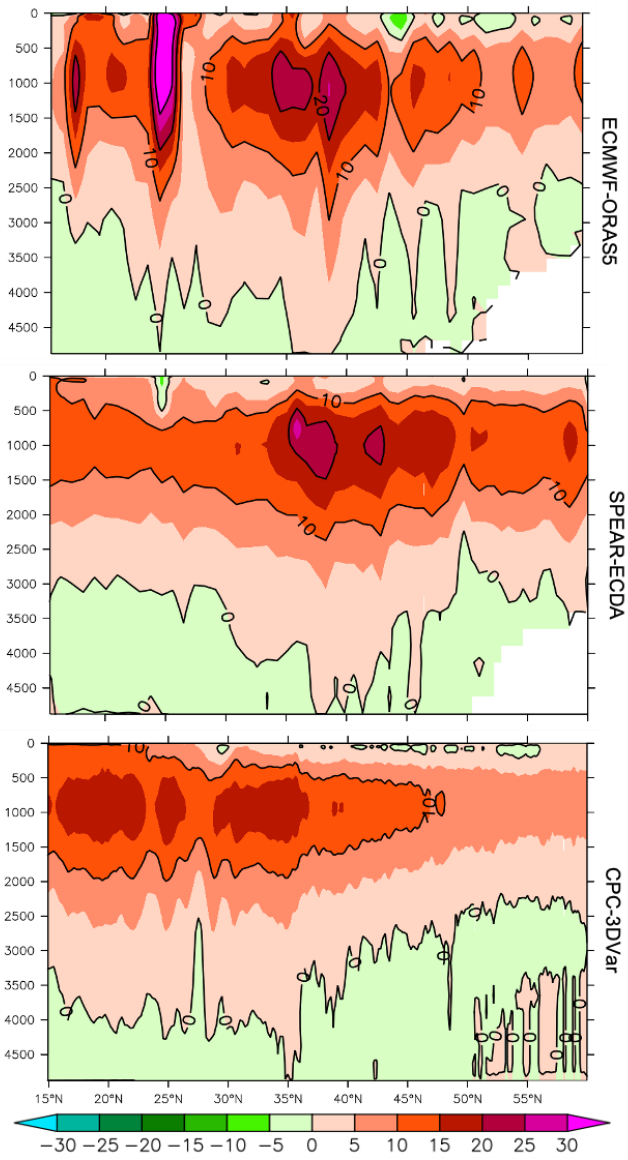

Figure A2.11: Mean meridional overturning circulation streamfunction (Sv) averaged over the North Atlantic for the period 2011-2018 in ECMWF-ORAS5 (top), SPEAR-ECDA(middle), and CPC-3DVar (bottom), showing maximum transport in dark red to pink shades.