A subseasonal Earth system prediction framework with CESM2

Jadwiga H. Richter[1], Anne A. Glanville[1], James Edwards[1], Brian Kauffman[1], Nicholas A. Davis[2], Abigail Jaye[3], Hyemi Kim[4], Nicholas M. Pedatella[5], Lantao Sun[6], Judith Berner[3,1], Who M. Kim[1], Stephen G. Yeager[1], Gokhan Danabasoglu[1], Julie M. Caron[1], and Keith W. Oleson[1]

[1] Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado

[2] Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, Colorado

[3] Mesoscale and Microscale Meteorology Laboratory, National Center for Atmospheric Research, Boulder, Colorado

[4] School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York

[5] High Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado

[6] Colorado State University, Fort Collins, Colorado

Key points: (140 char max each)

- A subseasonal research framework with CESM2(CAM6) and CESM2(WACCM6) is described
- Subseasonal prediction skill of CESM2(CAM6) and CESM2(WACCM6) is similar to that of CESM1(CAM5) and operational models
- The new framework facilitates predictability research for multiple aspects of the Earth system, including the mesosphere

Corresponding Author: Jadwiga H. Richter, jrichter@ucar.edu

Abstract

A framework to enable Earth system predictability research on the subseasonal timescale is developed with the Community Earth System Model, version 2 (CESM2) using two model configurations that differ in their atmospheric components. One configuration uses the Community Atmosphere Model, version 6 (CAM6) with its top near 40 km, referred to as CESM2(CAM6). The other employs the Whole Atmosphere Community Climate Model, version 6 (WACCM6) whose top extends to ~ 140 km in the vertical and it includes fully interactive tropospheric and stratospheric chemistry (CESM2(WACCM6)). Both configurations were used to carry out subseasonal reforecasts for the time period 1999 to 2020 following the Subseasonal Experiment’s (SubX) protocol. CESM2(CAM6) and CESM2(WACCM6) show very similar subseasonal prediction skill of 2-meter temperature, precipitation, the Madden-Julian Oscillation (MJO), and North Atlantic Oscillation (NAO) to the Community Earth System Model, version 1 with the Community Atmosphere Model, version 5 (CESM1(CAM5)) and to operational models. CESM2(CAM6) and CESM2(WACCM6) reforecast sets provide a comprehensive dataset for predictability research of multiple Earth system components, including three-dimensional output for many variables, and output specific to the mesosphere and lower-thermosphere (MLT) region. We show that MLT variability can be predicted ~ 10 days in advance of sudden stratospheric warming events. Weekly real-time forecasts with CESM2(WACCM6) contribute to the multi-model mean ensemble forecast used to issue the NOAA weeks 3-4 outlooks. As a freely available community model, both CESM2 configurations can be used to carry out additional experiments to elucidate sources of subseasonal predictability.

Plain Language Summary

Sources of subseasonal (i.e., timescale of three to four weeks) predictability for surface temperature, precipitation, and extreme events associated with subseasonal modes of variability are not well understood. In addition, there has been little exploration of the predictability of land, sea-ice, the stratosphere, and the mesosphere lower-thermosphere region. We describe here a subseasonal prediction research framework based on two configurations of the Community Earth System Model, version 2 (CESM2) that differ in their atmospheric components. Both configurations demonstrate subseasonal prediction skill comparable to that of operational models. Reforecasts carried out with two configurations of CESM2 provide a comprehensive dataset for predictability research of multiple aspects of the Earth system, including the mesosphere and lower thermosphere region. Real-time forecasts with these models contribute to the multi-model mean ensemble forecast used to issue the National Oceanic and Atmospheric Administration (NOAA) weeks 3-4 outlooks.

1 Introduction

Interest and demand for skillful subseasonal predictions (i.e., targeting three to four weeks) of the Earth system has grown in the recent decade. Multiple economic sectors such as agriculture, energy, and water management could benefit from improved subseasonal predictions (White et al. 2017). Such a need is a strong motivator of research of sources and limits of subseasonal predictability, including identifying windows of opportunity for increased forecast skill (Mariotti et al., 2020; NAS 2016). The international subseasonal-to seasonal (S2S) project and database (Vitart et al., 2017; Vitart and Robinson 2018) and the National Oceanic and Atmospheric Administration (NOAA) SubX project (Pegion et al., 2019) have been instrumental in providing real-time forecasts and reforecasts (forecasts initialized during the historical period) carried out with multiple operational and research models that serve as a community basis for research on predictability on S2S timescales.

Subseasonal prediction research has been focused mostly on prediction of the lowermost atmosphere, in particular surface temperature and precipitation, and extreme events associated with these, such as heat waves, droughts, heavy rainfall and cold outbreaks (Ford et al., 2018; de Andrade et al., 2019; Xiang et al., 2020). Substantial effort has also been invested in assessing predictability of dominant modes of variability on the subseasonal timescale, such as the Madden Julian Oscillation (MJO) and the North Atlantic Oscillation (NAO) as these can be drivers for extreme weather (e.g., Stan et al. 2017; Vitart et al., 2017; Kim et al., 2018; Lim et al., 2018; Sun et al., 2020; Yamagami & Matsueda, 2020). A few recent studies have started examining the predictability of sea-ice and noted a wide range of sea-ice prediction skill, with a multimodel mean forecast being skillful out to 5 months (Wayand et al., 2019; Zampieri et al., 2018). There has also been some exploration of the subseasonal predictability of various land model variables, such as soil moisture and snowpack (Hanchen et al., 2019; Diro & Lin, 2020), however predictability of other characteristics of land has not been explored. Several studies have looked at the predictability of the stratosphere, mainly at the predictability of sudden stratospheric warmings (SSWs), as they can significantly impact surface extreme weather especially over Eurasia (Tripathi et al., 2015), and the predictability of the quasi-biennial oscillation (QBO) which impacts the MJO (Lim et al., 2019; Kim et al., 2019a). Furthermore, there have only been limited efforts aimed at addressing the predictability of variability at higher altitudes (i.e., mesosphere, thermosphere, and ionosphere) as models used in S2S prediction typically do not extend into that region of the atmosphere. These prior studies have been limited as they used short reforecast periods (Wang et al., 2014; Pedatella et al., 2018a; Pedatella et al., 2019).

Variability of the mesosphere and lower-thermosphere (MLT) region drives a significant portion of near-Earth space weather, which can cause adverse effects on communications and navigation systems, and understanding the predictability in the MLT is thus an important component of enhancing space weather forecasting (Jackson et al., 2019).

Stratosphere-troposphere interactions provide a potential source of predictability on the S2S timescale because of their persistent and slow varying circulation anomalies (NAS, 2016). Increased predictability is believed to primarily come from SSWs which are followed by tropospheric circulation anomalies resembling the negative phase of the NAO. The QBO has been shown to lead to enhanced predictability on seasonal timescales (e.g., Boer & Hamilton, 2008; Marshall & Scaife, 2009), and is predictable out to several years ahead (Scaife et al., 2014b). Hence, a model that represents the QBO and SSWs well could potentially have more skill on the subseasonal timescale.

Richter et al. (2020) described the utility of the Community Earth System Model, version 1, with the Community Atmosphere Model version 5 as its atmospheric component (CESM1(CAM5)), a predecessor of CESM2, as a subseasonal prediction research model and demonstrated that the prediction skill of key surface variables with that model was comparable to the National Center for Environmental Prediction (NCEP) Climate Forecast System, version 2 (CFSv2) operational model. Here, we describe a new community resource for research on subseasonal predictability of multiple components of the Earth system: a subseasonal prediction system based on CESM2 with two configurations that differ in their atmospheric components. One configuration uses the Community Atmosphere Model, version 6 (CAM6), referred to as CESM2(CAM6). The other employs the Whole Atmosphere Community Climate Model, version 6 (WACCM6), and is referred to as CESM2(WACCM6). CESM2 is the newest version of the

NCAR coupled Earth system model used for the Coupled Model Intercomparison Project phase 6 (CMIP6) simulations (Danabasoglu et al., 2020). Both configurations of CESM2 include prognostic atmospheric, land, ocean and sea-ice components and resolve the interactions between them. Both configurations of the model include prognostic aerosols and CESM2(WACCM6) also includes fully interactive tropospheric and stratospheric chemistry. CESM2(WACCM6) has a very good representation of SSWs and an internally generated QBO, hence it potentially could be more skillful, especially during SSW events, than models with smaller vertical domains. Another unique aspect of CESM2(WACCM6) is the extension of the model domain into the lower thermosphere, enabling investigations into the predictability at MLT altitudes. SSW events are now recognized to have impacts throughout the whole atmosphere (Baldwin et al., 2020; Pedatella et al., 2018b), including the mesosphere, thermosphere, and ionosphere, where they influence the day-to-day weather of the near-Earth space environment. It is, therefore, important to understand the predictability of the SSW effects in the middle and upper atmosphere.

Weekly real-time forecasts are being generated since September 2020 with CESM2(WACCM6) and since April 2021 with CESM2(CAM6), and they contribute to the multi-model mean ensemble used to issue the experimental NOAA weeks 3-4 outlooks. The motivation behind the inclusion of CESM2(WACCM6) into this NOAA Climate Test Bed project is to examine how much improvement in surface prediction skill can be gained from the inclusion of a well-represented stratosphere, especially during the boreal winter, when the impacts of SSWs on the surface climate and impacts of the QBO on the MJO are the largest.

We describe here the S2S prediction framework, reforecasts, and near-real time forecasts with CESM2(CAM6) and CESM2(WACCM6) including the extensive output of atmospheric, land, ocean, and sea-ice models, with several key atmospheric variables reaching into the MLT region. CESM2 is a community model and is freely available to the broader community. The reforecast sets described here are designed to serve as a basis for future experiments with CESM2(CAM6) and CESM2(WACCM6) investigating sources of subseasonal predictability.

2 Model and System Description

2.1 Model Description

Subseasonal reforecasts and forecasts described here use the default released version of CESM2. CESM2 is an open-source, comprehensive Earth system model designed primarily for the studies of Earth's past, present and future climates. CESM2 includes ocean, atmosphere, land, sea-ice, land-ice, river, and wave model components and is thoroughly documented in Danabasoglu et al. (2020). The standard CESM2 uses a nominal 1° horizontal resolution (1.25° in longitude and 0.9° in latitude in its atmospheric components). CAM6 is the default atmospheric model. It has 32 vertical levels with the model lid near 2 hPa (~ 40 km). CAM6 uses the Zhang and McFarlane (1995) convection parameterization, the Cloud Layers Unified By Binormals (CLUBB; Golaz et al., 2002; Larson, 2017) unified turbulence scheme, and the updated Morrison-Gettelman microphysics scheme (MG2; Gettelman & Morrison, 2015). A form drag parameterization of Beljaars et al. (2004) and an anisotropic gravity wave drag scheme following Scinocca and McFarlane (2000) replace the turbulent mountain stress parameterization that was used in CESM1. The aerosols in CAM6 are represented using the Modal Aerosol Model version 4 (MAM4) as described in Liu et al. (2016).

CESM2(WACCM6) uses WACCM6 or the "high-top" version of the atmospheric model, which is documented in detail in Gettleman et al. (2019). WACCM6 has the same horizontal resolution as CAM6, however it has 70 vertical levels with a top near 4.5×10^{-6} hPa (~ 140 km). The representation of atmospheric physics is identical to that in CAM6, with the only exception being the representation of non-orographic gravity waves, which follows Richter et al. (2010) with changes to tunable parameters described in Gettleman et al. (2019). The higher model lid and parameterization of non-orographic gravity waves in WACCM6 allow for a better representation of middle atmospheric dynamics as compared to CAM6 and the simulation of an internally-generated QBO. Another key difference between CAM6 and WACCM6 is in the representation of chemistry. The comprehensive chemistry module in WACCM6 includes interactive tropospheric, stratospheric, and lower thermospheric chemistry (TSMLT) with 228 prognostic chemical species, described in detail in Gettleman et al. (2019). Differences in the representation of aerosols and chemistry between CAM6 and WACCM6 do not significantly impact the mean surface and tropospheric climate in historical simulations. However, CESM2(WACCM6) simulations have a more realistic representation of polar climate as compared to CESM2(CAM6) as shown in Gettleman et al. (2019).

CESM2(CAM6) and CESM2(WACCM6) use identical ocean, land, sea-ice, land-ice, river-transport, and wave models. The ocean model is based on the Parallel Ocean Program version 2 (POP2; Smith et al., 2010; Danabasoglu et al., 2012), but contains many advances since its version in CESM1. As described in Danabasoglu et al. (2020), these include a new parameterization for mixing effects in estuaries, increased mesoscale eddy (isopycnal) diffusivities at depth, use of prognostic chlorophyll for shortwave absorption, use of salinity-dependent freezing-point together with the sea-ice model, and a new Langmuir mixing parameterization in conjunction with the new wave model component. Several numerical improvements were also implemented as described in Danabasoglu et al. (2020). The horizontal resolution of POP2 is uniform in the zonal direction (1.125°), and varies from 0.64° (occurring in the Northern Hemisphere) to 0.27° at the Equator. In the vertical, there are 60 levels with a uniform resolution of 10 m in the upper 160m. The ocean biogeochemistry is represented using the Marine Biogeochemistry Library (MARBL), essentially an updated implementation of what has been known as the Biochemistry Elemental Cycle (Moore et al., 2002; 2004; 2013). CESM2 includes version 3.14 of the NOAA WaveWatch-III ocean surface wave prediction model (Tolman, 2009). CICE version 5.1.2 (CICE5; Hunke et al., 2015) is used to represent sea-ice in CESM2 and uses the same horizontal grid as POP2. The vertical resolution of sea-ice has been enhanced to eight layers, from four in CESM1; the snow model resolves three layers, and the melt pond parameterization has been updated (Hunke et al., 2013).

Both CESM2 configurations use the recently developed Community Land Model version 5 (CLM5) described in detail in Lawrence et al., (2019). As compared to CLM4, CLM5 includes improvements to soil hydrology, spatially explicit soil depth, dry surface layer control on soil evaporation, updated ground-water scheme, as well as several snow model updates. CLM5 includes a global crop model that treats planting, harvest, grain fill, and grain yields for six crop types (Levis et al., 2018), a new fire model (Li et al., 2013; Li & Lawrence, 2017), multiple urban classes and updated urban energy model (Oleson & Feddema, 2019), and improved representation of plant dynamics. The river transport model is the Model for Scale Adaptive River Transport (MOSART; H. Y. Li et al., 2013). The Community Ice Sheet Model Version 2.1 (CISM2.1; Lipscomb et al., 2019) is used to represent the ice sheets, although in the configuration of this model ice sheets are assumed to be fixed.

2.2 Initialization

S2S reforecasts with CESM2(CAM6) and CESM2(WACCM6) use the same land initial conditions, but differ in atmosphere, ocean, and sea-ice initialization. These differences are due to the different location of the two atmospheric models' lids and also due the inclusion of CESM2(WACCM) forecasts in NOAA's experimental Week 3-4 outlooks since September 2020, necessitating real-time forecasting ability with that model and completion of reforecasts with the same model set-up at that time. Initialization procedures for each model component are described below and summarized in Table 1.

Land initial conditions for CESM2(CAM6) and CESM2(WACCM6) reforecasts were produced using the stand-alone CLM5. The stand-alone CLM5 simulation employed a setup consisting of biogeochemistry-driven crops and glacial observations. A 700-year spin-up was performed using 6-hourly atmospheric variables (precipitation, temperature, wind speed, shortwave and longwave radiation, etc.) from the NCEP CFSv2 reanalysis (Saha et al. 2014). Near present-day (year 2000) greenhouse gas forcings were used continuously throughout the spin-up, while atmospheric forcings from NCEP CFSv2 were cycled between 1979-1985 until a steady state was achieved (~100 cycles). After soil moisture and temperatures stabilized with respect to the 1979-1985 climate state, the CLM5 continued to be forced with NCEP CFSv2 up through present day (no longer cyclically), and initial condition files were output for use in reforecasts each Monday.

CESM2(CAM6) atmosphere was initialized using the NCEP CFSv2 reanalysis interpolated to the CAM6 grid. Initialized fields include the zonal and meridional wind, temperature, specific humidity, surface pressure and surface geopotential. An ensemble is generated using the random field perturbation method at initial time which was shown to be as effective as other more sophisticated methods to generate model spread by Magnusson et al. (2009) and was utilized successfully in S2S reforecasts with CESM1(CAM5) (Richter et al. 2020).

Ocean and sea-ice initial conditions for CESM2(CAM6) come from a reforecast ocean-sea-ice coupled configuration of CESM2(CAM6) forced with the adjusted Japanese 55-year reanalysis project state fields and fluxes (JRA55-do forcing; Tsujino et al., 2018). We call this JRA55-do forced ocean simulation (JRA55-do FO). This simulation was integrated through five cycles of the 1958 - 2009 forcing, with the last cycle extended through 2019. This procedure follows the protocol for the CMIP6-endorsed Ocean Model Intercomparison Project phase 2 (OMIP2; Griffies et al., 2016; Tusjino et al., 2020), and is the same as was done for S2S reforecasts with CESM1(CAM5) (Richter et al., 2020).

The initialization of the atmosphere, ocean, and sea-ice in CESM2(WACCM6) is not as straightforward as for CESM2(CAM6) as the model's lid located near ~ 140 km extends above the currently available atmospheric reanalyses and the JRA55-do was only available through 2019 with a yearly update frequency in early 2020 (time of model set-up and running of reforecasts), which prohibited its use in near real-time forecasts. To generate realistic initial conditions for the entire atmospheric domain, first a specified dynamics (SD) simulation with fully coupled CESM2(WACCM6) was carried out (WACCM6-SD) in which the atmospheric dynamics were nudged to the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) (Gelaro et al. 2017) with a 1-hourly nudging timescale from 1999 to 2020. 1-hourly nudging ensured that the dynamics in the lower atmosphere are very close to the MERRA-2 reanalysis, which is important for tropospheric subseasonal prediction. The ocean in this WACCM6-SD simulation is initialized from the JRA55-do FO simulation (as done for

CESM2) in year 1998, and then it is left to evolve with atmospheric fluxes from the MERRA-2 reanalysis for 5 years. In this set-up, the ocean state drifts from the observed state and the JRA55-do simulation, hence every 5 years the ocean in the SD simulation is reinitialized with the ocean state from the JRA55-do forced ocean simulation. Hence ocean reinitialization occurred on January 1 of 1998, 2003, 2008, 2013, and 2018. We have developed the ability to update the JRA55-do in August of 2020, hence a final ocean reinitialization occurred on August 31 of 2020 in order to prime the real-time application which began at that time. Daily atmospheric, ocean and sea-ice, initial conditions were output from the WACCM6-SD simulation for use in reforecasts. The random field perturbation method was applied to the atmospheric conditions to generate ensemble spread in the same way as was done in CESM2(CAM6).

Figure 1 shows correlation and root-mean-square error (RMSE) maps between the sea surface temperature (SST) in JRA55-do FO (used to initialize CESM2(CAM6) reforecasts) and HadSST observations (Figs. 1a and 1c) and between SSTs in WACCM6-SD (used to initialize CESM2(WACCM6) reforecasts) and HadSST (Figs. 1b and 1d). Over the 1999 - 2019 period, the correlation between JRA55-do FO and WACCM6-SD and observations is close to 1 over the majority of ocean areas, with the exception of reduced values of the correlation coefficient in the Tropics and south of 50^{o}S. The correlation coefficients are lower in those regions in the WACCM6-SD as compared to the JRA55-do FO simulation. The RMSE distribution (Figure 1c,d) is also very similar between JRA55-do FO and WACCM6-SD, with the largest RMSE differences between the two simulations in the Tropics. The larger RMSE in WACCM6-SD as compared to the JRA55-do FO could be related to differences in variability in MERRA-2 as compared to JRA55-do. This greater Tropical drift away from the observed SSTs is illustrated clearly in Figure 1e which shows the El Nino Southern Oscillation (ENSO) index in the JRA55-

do FO, WA CCM6-SD, and HadISST. JRA55-do follows the observed ENSO index closely, however there are a few instances when the ENSO index in WACCM-SD departs significantly from observations. This includes the period from ~ 2015 to 2016 and 2019 to 2020.

For real-time forecasts with CESM2(WACCM6), the same initialization procedure was used as for reforecasts except that the CESM2(WACCM6) run was nudged to the NASA Forward Processing for Instrument Teams (FP-IT) reanalysis instead of MERRA-2, as the FP-IT reanalysis is available in near real-time.

2.3 Protocol and output

The S2S reforecasts were carried out following the SubX protocol (Pegion et al. 2019) with weekly initializations every Monday from 1999 to 2020 for CESM2, and for every Monday between September and March for CESM2(WACCM6). An 11-member ensemble was carried out for the CESM2(CAM6) and a 5-member ensemble was carried out for the CESM2(WACCM6) reforecasts. The computational cost of CESM2(WACCM6) is nearly eight times the cost of CESM2(CAM6), hence carrying out more ensemble members and all start dates was computationally prohibitive with CESM2(WACCM6). Near real-time forecasts began with CESM2(WACCM6) in September of 2020, and in April 2021 with CESM2(CAM6), both with a 21-member ensemble.

The S2S reforecast set with CESM2(CAM6) and CESM2(WACCM6) have extremely comprehensive output for the atmosphere, land, ocean, and sea-ice components of the model to enable studies of predictability of the broader Earth system, including the MLT region. Output is available from the NCAR Climate Data Gateway (see links in Acknowledgements). The complete list of output variables is shown in Tables S1- S7. Because the reforecasts follow the

SubX protocol, a portion of the output also follows that protocol, and a number of model native fields are renamed and reformatted to match the SubX priority 1, 2, and 3 (p1, p2, and p3, respectively) variables. In addition to these variables which are all two-dimensional (on a lat/lon grid), more daily averaged variables are saved for every model component. A handful of atmosphere-relevant variables are saved at 6-hourly intervals for applications such as tropical cyclone tracking. In addition, a limited number of 3-dimensional fields is stored at 14 pressure levels for CESM2(CAM6) and at 22 levels for CESM2(WACCM6) (see Table S4 for exact levels). Finally, for CESM2(WACCM6), diurnal and semidiurnal tide coefficients are stored at 8 levels at and above 10 hPa, permitting the evaluation of migrating and nonmigrating tides in the MLT. Because CESM2 includes an interactive crop model, the output list for the land model includes variables such as gross and primary production which are very unique to this dataset.

3 Results

The subseasonal prediction skill of CESM2(CAM6) and CESM2(WACCM6) in the S2S reforecasts is evaluated for key surface variables (temperature, precipitation), dominant subseasonal modes (MJO and NAO) as well as stratosphere-troposphere coupling. Subsequently we briefly examine MLT predictability during SSWs in CESM2(WACCM6). We compare the tropospheric prediction skill to that from reforecasts carried out with the default version of CESM1(CAM5) utilizing the 30-level version of CAM5 (Richter et al. 2020) for the common period of 1999 to 2015. As the reforecasts with the default (30-level) version of CESM1(CAM5) used a 10-member ensemble, we use here a 10-member average of CESM2(CAM6) as well, because ensemble size does affect skill (e.g., Sun et al., 2020). Therefore, in selected figures, we also show CESM1(CAM5) and CESM2(CAM6) skill based on a 5-member ensemble, because that is what the CESM2(WACCM6) skill assessment is based on. Richter et al. (2020) showed that the 2-meter temperature and precipitation skill of CESM1(CAM5) was very similar to the NOAA operational CFSv2 model and higher than those of most other models participating in SubX. Surface temperature and precipitation prediction skill is similar between the CFSv2 model and the European Centre for Medium-Range Weather Forecasts (ECMWF)Variable Resolution Ensemble Prediction System monthly forecast system (Wang & Robertson 2019), hence broadly speaking skill similar to CESM1(CAM5) implies prediction skill comparable to other operational models.

3.1 2-meter temperature and precipitation prediction skill

Figures 2a-c show the anomaly correlation coefficients (ACC) for 2-meter (2m) temperature for December, January, and February (DJF) for weeks 1-2, 3-4, and 5-6 for CESM2. The NOAA Climate Prediction Center (CPC) Global Daily Temperature dataset at the 0.5^ox0.5^o resolution is used as a verification dataset. Both for observations and simulations, the average daily temperature is calculated as the average of the daily maximum and minimum temperature. Similarly to what was done for CESM1 in Richter et al. (2020), ACC values are shown in colors only when they are significantly different from zero at the 95% confidence level or for ACC > 0.2. The significance level is calculated using a total sample size of 221, based on 13 start dates per year over 17 years (1995 to 2015) considered here. Subsequently, we assume a 2-week decorrelation time, and resulting in 110.5 independent samples. There are hence 108 degrees of freedom (number of independent samples minus 2), leading to a correlation equal or greater than 0.2 being significant at the 95% level using a two-tailed Student's t-test (Wilks, 2011). Because Richter et al. (2020) showed that nearly all the values over this threshold exceed the persistence forecast, a persistence forecast is not repeated here. Figures 2a-c show declining ACC values with forecast lead time reflecting a loss of deterministic skill with increasing forecast lead time. The globally averaged DJF ACC for 2m temperature over all land areas is ~ 0.3 for weeks 3-4 and 0.2 for weeks 5-6 with higher values over the northern part of South America (ACC of ~ 0.5 to 0.6 through weeks 5-6) and the lowest values over north and north-eastern Asia. The differences of DJF 2m temperature ACC between CESM2(CAM6) and CESM1(CAM5) and between CESM2(CAM6) and CESM2(WACCM) are given in Figs. 2d-f and Figs. 2g-i, respectively. Only values that exceed the 95% confidence level using the Fisher z transform (e.g., Zar, 2014) are shown. Figures 2d-2f show that DJF ACC for 2m temperature in CESM2(CAM6) is overall very similar to that of CESM1(CAM5) for the majority of the world's land regions, with the only exceptions being regions of decreased skill over parts of north-east and southernmost Asia, and southernmost part of India for weeks 3-4 and 5-6. Figures 2g-2i reveal that the DJF 2m temperature ACC for CESM2(WACCM6) is also very similar to that of CESM2, demonstrating that the whole atmosphere version of CESM2 does not fundamentally change the surface prediction skill of the model. There are a few land regions for which the DJF 2m temperature ACC is statistically significantly different for CESM2(WACCM6) as compared to CESM2, most evident in weeks 5-6. These include parts of North America for which CESM2(CAM6) is showing higher skill than CESM2(WACCM6), and eastern Asia where higher skill is seen in CESM2(WACCM6) as compared to CESM2(CAM6). A detailed investigation (beyond the scope of this paper) is needed to elucidate whether these differences can be attributed to differences either in the representation of the stratosphere or in ocean and atmosphere initialization procedures between the two configurations.

Figure 3 shows ACC of 2m temperature over land for June, July, August (JJA) average. Comparison to CESM2(WACCM6) is not possible for this season due to the limited range of reforecast start dates for that model version. The overall ACC of JJA 2m temperature is a little smaller as compared to that for DJF. The ACC values are the largest in northern South America and tropical Africa for weeks 3-4 and weeks 5-6 (Figs. 3b,c). The differences between ACC in CESM2(CAM6) and CESM1(CAM5) are very small as shown in Figs 3d-3f. Figures S1 and S2 show the 2m temperature ACC for March, April, May (MAM) and September, October, and November (SON) averages respectively. In MAM, CESM2(CAM6) shows a statistically significant degradation of 2m temperature prediction skill over Eurasia and Alaska by ~0.2 for weeks 3-4 and weeks 5-6 over CESM1(CAM5). In SON, there is very little difference between the 2m temperature ACC for CESM2(CAM6) and CESM1(CAM5), as well as between CESM2(WACCM6).

Figure 4 compares the DJF and JJA 2m temperature ACC averaged over all land areas and over North America. DJF ACC of 2m temperature is ~ 0.6 for weeks 1-2, ~ 0.3 for weeks 3-4, and < 0.2 for weeks 5-6 for global land for all the CESM versions considered here (Figure 4a). DJF ACC of 2m temperature over North America is ~0.7 for weeks 1-2, ~0.3 for weeks 3-4, and ~0.15 for weeks 5-6. JJA ACCs of 2m temperature for both global and North America land are ~0.1 lower for weeks 1-2 and 3-4 as compared to DJF, while they are comparable to those of DJF for weeks 5-6. There are overall small differences in 2m temperature ACCs between the various model versions considered, as well as between ACCs calculated for an ensemble size of 5 vs 10 for CESM1(CAM5) and CESM2(CAM6) for both DJF and JJA. Although there are small differences between CESM1(CAM5) and CESM2(CAM6) in DJF and JJA ACCs over North America and global land, the application of the Fisher z transform to these values showed that none of the differences between ACC values depicted by individual bars in Figure 4 are statistically significant.

Figures 5a-c and 6a-c show the ACC for precipitation for DJF and JJA for CESM2(CAM6). Precipitation prediction skill at subseasonal timescales (Figs. 5b,c) is quite low as compared to the 2m temperature, with ACC values on average of ~ 0.1 for weeks 3-4 and < 0.05 for weeks 5-6 consistent with previous findings (Pegion et al., 2019, Richter et al., 2020). Similarly to 2m temperature skill, precipitation skill is slightly higher in northern South America and parts of Africa in weeks 3-4 in CESM2(CAM6) as compared to other land areas, reaching ACC values of 0.3-0.4 over small regions (Figs. 5b,6b). There is little difference in DJF ACC of precipitation between CESM2(CAM6) and CESM1(CAM5), and between CESM2(CAM6) and CESM2(WACCM6). In JJA (Figure 6), the overall precipitation skill over land is even lower than in DJF with the exception of Australia. In CESM2, for both weeks 3-4 and weeks 5-6 the ACC of precipitation is ~ 0.3 - 0.5 over most of Australia, showing that CESM2(CAM6) is skillful in that region. CESM1(CAM5) already had significant ACC over Australia in JJA (Richter et al., 2020), so this skill has increased in CESM2(CAM6) especially for weeks 5-6 (Figure 6f). Figure 7 summarizes the precipitation prediction skill for DJF and JJA for all the models considered in this study. Averaged over global land and North America the ACC of precipitation is greater than zero but smaller than 0.1 for weeks 3-4 and weeks 4-5. ACC values less than 0.1 imply that precipitation is generally not predictable on the subseasonal timescales, except for very few selected regions discussed above.

3.2 Spread and error characteristics for 2m temperature

To shed some light on the ensemble characteristics of our S2S forecasts, we compute the RMSE of the ensemble mean and the ensemble spread (Figure 8). Similarly to the ACC, the RMSE over North American land is markedly higher in winter than in summer and decreases slightly if the ensemble size is increased from 5 to 10 members. Unlike Figure 4, we do not detect the same rapid decrease in skill between week 1-2 and 3-4 forecasts. This points to the fact that week 1-2 reforecasts have a high pattern correlation with the verifying analysis but might have problems capturing the anomaly magnitudes.

The ensemble spread is computed as lead-time dependent standard deviation of all members around the ensemble mean and is shown as hatched bars in Figure 8. For a reliable ensemble system, the ensemble spread should inform the state-dependent predictability of the system and the spread and error of the ensemble mean should have the same magnitude (e.g. Leutbecher and Palmer, 2008). However, most ensemble systems are overconfident (e.g., Berner et al. 2015, Leutbecher et al. 2017) and the spread predicting the uncertainty of the forecast is smaller than the RMSE.

Such underdispersion is also evident in our reforecasts. In weeks 1-2, regardless of the season, or land area average, the spread is under-dispersive by at least 40% (Fig 8.). The underdispersion improves for longer lead times but forecasts remain markedly overconfident for all experiments. The differences between the different CESM configurations are small for JJA, but for DJF, CESM1(CAM5) creates consistently more spread than CESM2(CAM6) or CESM2(WACCM) over North American land. Increasing the ensemble size has a more pronounced effect on the spread than the RMSE error. This indicates that the value of the ensemble might lie in the improved representation of uncertainty rather than improved deterministic skill.

3.3 MJO and NAO prediction skill

The MJO and the NAO are key drivers of extreme weather on subseasonal timescales and believed to be key sources of subseasonal predictability. To evaluate the MJO prediction skill, the Real-time Multivariate MJO (RMM; Wheeler & Hendon, 2004) index is calculated with the 200 hPa and 850 hPa daily zonal wind from ECMWF Reanalysis v5 (ERA5; Hersbach et al., 2020) and the Outgoing Longwave Radiation (OLR) from NOAA Advanced Very High-Resolution Radiometer (Liebmann & Smith 1996). Predicted RMM indices are calculated by projecting the forecast anomalies for those fields onto the associated observed EOF eigenvectors (Kim et al., 2018). Then, the RMM index bivariate ACCs are computed between the predicted and observed RMM1 and RMM2 indices as a function of forecast lead days. The MJO prediction skill is assessed during boreal winter with the reforecasts initialized during November-March (NDJFM). Due to the limited sample size, all days are selected as MJO events without any discrimination of the initial MJO amplitude. Figure 9 shows ACC as a function of forecast lead days where ACC of 0.5 is explicitly denoted as it is often used as a skill threshold (e.g., Rashid et al., 2011). The figure clearly demonstrates that the MJO in CESM2(CAM6) and CESM2(WACCM6) is predictable out to 25 days, which is longer than the predictability of the MJO for most of the SubX models (not shown), but less than than the MJO predictability of out to 33 days in the ECMWF-CY43R system (Kim et al., 2019b). The ACC of the MJO in CESM1(CAM5) is slightly higher compared to CESM2(CAM6) and CESM2(WACCM6), however, none of the skill differences are statistically significant based on the Fisher z transform. There is also very little difference in the overall MJO skill between CESM2(CAM6) and CESM2(WACCM6) (when the same ensemble size is considered) indicating that neither the extension of the model top into the middle atmosphere nor the different ocean initialization in CESM2(WACCM6) as compared to CESM2(CAM6) affect MJO prediction skill.

The NAO is a key driver of winter extreme weather over Europe and North America (Hurrell, 1995; Scaife et al., 2008). It is predictable on weather (< 2 weeks) timescales and seasonal timescales (e.g., Riddle et al., 2013, Scaife et al. 2014a), however its predictability on subseasonal timescales is less certain and has not been explored extensively. Zuo et al. (2016) found the NAO to be predictable only out to ~ 9 days using the Beijing Climate Center Atmospheric General Circulation Model version 2.2 (BCC AGCM2.2). Pegion et al. (2019) showed that NAO skill was high (ACC > 0.5) through week 2 in all the SubX models. Richter et al. (2020) found that the ACC of NAO in CESM1(CAM5) was 0.5 at week 3 and 0.4 at week 4 (10-member ensemble). Sun et al. (2020) found that an increase in ensemble size to 20 enhances the NAO prediction skill, with an NAO ACC of 0.51 for weeks 3 to 6 in boreal winter in CESM1(CAM5). The prediction skill of the NAO in the various CESM configurations is shown in Figure 10. The NAO index was obtained by first calculating EOF analysis of ERA-Interim monthly (NDJFM) sea level pressure (SLP) anomalies over the Atlantic sector (20°N– 80°N, 90°W–40°E) and treating the leading EOF pattern as the NAO. The NAO index was then calculated by projecting the SLP anomaly in the reanalysis and reforecasts that were initialized during NDJFM onto the leading EOF. The week 3-4 NAO ACC is above or close to 0.5 for all the CESM versions considered here, similar to the skill in ECMWF and NCEP reforecasts (Wang & Robertson, 2018). ACC of CESM1(CAM5) based on a 10-member ensemble and CESM2(WACCM6) based on a 5-member ensemble have the highest skill at week 3-4, however, with the current reforecast sample size, these skill values are not significantly different than the ACC for CESM2(CAM6) or CESM1(CAM5) based on a 5-member ensemble. The NAO skill for CESM2(WACCM6) is very close to the NAO skill for CESM1(CAM5) at weeks 5-6, and substantially higher than that for CESM2(CAM6) with a 5-member ensemble. This could possibly be attributed to a better resolved stratosphere in CESM2(WACCM6), but as with other comparisons shown throughout this manuscript, due to the limited sample size, these differences are not statistically significant.

3.4 Stratosphere-troposphere coupling

The stratosphere, and in particular, stratosphere-troposphere coupling during SSWs may be an important source of subseasonal predictability. SSWs are associated with enhanced surface pressure over the polar cap, and they tend to be followed by warm temperatures over Northeastern Canada and Greenland, cold temperatures over Eurasia, and enhanced precipitation over Western Europe (Butler et al., 2017, Domeisen & Butler 2020, Baldwin et al., 2020). This coupling between tropospheric weather and sudden warmings is often summarized by the time evolution of the annular modes (Baldwin and Dunkerton 2001), or nearly equivalently, the standardized polar cap geopotential anomalies (Figure 11). During the onset of an SSW, anomalously positive geopotential anomalies descend from the middle to the lower stratosphere, where they can linger for over one month. Their descent to the surface manifests itself as changes to the Arctic Oscillation (AO) or the NAO over the Atlantic sector.

Here, the standardized polar cap geopotential anomalies in MERRA-2 (Figure 11a) and in CESM2(WACCM6) and CESM2(CAM6) reforecasts that predicted a major SSW within 7 days of the SSW central date in MERRA-2 reanalysis (Figs. 11b,c) are composited with respect to the central date of the observed or reforecasted SSW. We emphasize that only reforecasts that predicted an SSW were selected to assess the models' ability to capture surface impacts. The central date of an SSW is the first day when the zonal-mean zonal wind at 60^{o}N and 10 hPa becomes negative, with 14 SSW events in the reforecast period. The central dates of the observed events are: (1) Feb 26, 1999; (2) Feb 11, 2001; (3) Dec 30, 2001; (4) Feb 17, 2002; (5) Jan 18, 2003; (6) Jan 5, 2004; (7) Jan 21, 2006; (8) Feb 24, 2007; (9) Feb 22, 2008; (10) Jan 24, 2009; (11) Feb 9, 2010; (12) Jan 6, 2013; (13) Feb 12, 2018; and (14) Jan 2, 2019. Figure 12 shows that while the magnitude of the positive geopotential anomalies during the SSW events is comparable between MERRA-2 and the CESM2(WACCM6) and CESM2(CAM6) reforecasts, the positive anomalies in the lower stratosphere do not linger as long in the reforecasts, only out to day 35 and 39, respectively. However, the positive surface geopotential anomalies linger for 4 weeks after the central date of a SSW in the reforecasts and in MERRA-2, indicating that the coupling of the events with the troposphere is comparable (Baldwin et al., 2021). The squared pattern correlations of the composited geopotential anomalies between the CESM2(WACCM6) and CESM2(CAM6) reforecasts and MERRA-2 are similarly high at 0.80 and 0.77 respectively. In contrast, the averages of the individual reforecast pattern correlations with their respective SSW event in MERRA-2 are substantially lower with 0.39 for CESM2(WACCM) and 0.25 for CESM2(CAM6). In summary, CESM2(WACCM6) reforecasts of the polar cap geopotential anomalies following an SSW are somewhat more consistent with those of MERRA-2 than in CESM2(CAM6) reforecasts.

3.5 Mesosphere and lower thermosphere prediction

Initial investigations by Wang et al. (2014) and Pedatella et al. (2018b) demonstrated the potential to predict the MLT variability during the 2009 SSW event, though these studies were limited to a single event. The CESM2(WACCM6) reforecasts provide the opportunity to perform more detailed investigations into the MLT predictability during SSWs. Davis et al. (2021) showed that SSW predictability at lead times of one to two weeks is enhanced in reforecasts initialized with weaker stratospheric jets. Figure 12 presents an analysis of SSW predictability using a composite of the zonal-mean temperature between 70º-90ºN from 14 major SSW events that are captured in the time periods of CESM2(WACCM6) reforecasts and SSWs from WACCM Specified Dynamics simulations with thermosphere-ionosphere eXtension (WACCMX-SD; Liu et al., 2018) are used for verification (Figure 12a). Figures 12b-e show the composites for reforecasts initialized 15, 10, 5, and 0 days prior to the SSW central date. Note that the results in Figure 12 are based on compositing the reforecasts regardless of whether they successfully forecast a SSW, and we consider reforecasts initialized within +/- 3 days of the specified lag for the composites (i.e., a lag of -10 includes reforecasts initialized 7-13 days prior to the SSW).

Several distinct features of the middle atmosphere (stratosphere ~ 100 to 0.5 hPa or ~10 to 50 km; mesosphere: 0.5 hPa to 10^{-3} hPa ~ 50 to 90 km, lower thermosphere: above 10^{-3} hPa) response to SSWs can be seen in Figure 12a. This includes a mesosphere cooling that begins right after the central date of the SSW between ~10^{-1} and 10^{-3} hPa that accompanies the warming in the stratosphere, as well as the reformation of the stratopause at high altitudes following the SSW. We note that formation of an elevated stratopause following an SSW does not always occur (Chandran et al., 2013), though it is present in the vast majority of the events considered here, thus appearing in the composite analysis. The CESM2(WACCM6) reforecasts indicate that the formation of an elevated stratopause and the mesosphere cooling can be predicted ~10-15 days in advance of the SSW, though the altitude of the elevated stratopause is too low in these early predictions. The reforecasts initialized closer to the SSW (Figure 12d) and near the SSW onset (Figure 12e) capture the mesosphere cooling and elevated stratopause with higher fidelity when comparing to WACCMX-SD. These results provide an initial demonstration that the MLT variability can be predicted ~5-15 days in advance of SSWs. The MLT variations during SSW generate subsequent variations in the ionosphere and thermosphere, and the results in Figure 12 thus suggest that it may be possible to also forecast the upper atmosphere variability ~10 days in advance of an SSW.

3.6 Limitations of current framework for chemistry prediction

As CESM2(WACCM6) includes a comprehensive tropospheric and middle atmospheric chemistry module, we were hopeful that the current model framework could also be used to explore the predictability of stratospheric chemistry such as water vapor and ozone. However, we have discovered that nudging CESM2(WACCM6) to MERRA-2 with a 1-hourly timescale introduces significant deviations between modeled and observed water vapor. This is illustrated in Figure 13 which shows the time evolution of the stratospheric tropical water vapor, also known as the “tape recorder” (Mote et al. 1996), for WACCM6-SD simulation (used to initialize CESM2(WACCM6) reforecasts) and Microwave Limb Sounder (MLS) observations (Lambert et al., 2015). Figure 13a reveals that stratospheric water vapor concentrations in WACCM6-SD are approximately double the observed concentration. Additionally, the water vapor tape recorder indicates faster ascent in WACCM6-SD, such that the simulated water vapor leads the observations as seen in the 100 hPa and 70 hPa time series (Figure 13b).

We have performed several sensitivity experiments with WACCM6-SD, including an experiment in which we lowered the nudging top from 60 km to 50 km and another experiment in which we increased the nudging timescale from 1 to 2 hours. We found that the first experiment had no effect on the simulation of water vapor, whereas the second experiment decreased the value of tropical lower stratospheric water vapor by about 15%, making the time evolution of water vapor closer to observations. It is possible that the 2-hour nudging results in a colder and higher tropopause or weaker recirculation of water vapor-rich air from the midlatitudes, both of which would reduce water vapor within the tape recorder. It is also possible that temperature nudging acts as diabatic heating and artificially changes the strength of the meridional circulation (Miyazaki et al., 2005). This could decrease the transit time of water vapor-rich tropospheric air through the tropical tropopause layer, thereby decreasing the amount of dehydration that can occur. An even longer nudging timescale in the stratosphere may improve the representation of stratospheric chemistry in the S2S reforecasts/forecasts with CESM2(WACCM6) and we will explore this in the future further.

4 Summary and Conclusions

We have described here a fully coupled Earth system subseasonal prediction framework with CESM2(CAM6) and CESM2(WACCM6) developed for research purposes. CESM2(CAM6) and CESM2(WACCM6) are the newest versions of the NCAR Earth system model used in CMIP6, and the two configurations differ in the atmospheric model components. CESM2(CAM6) has a top near 40 km, whereas CESM2(WACCM6) extends up to ~ 140 km and includes fully interactive tropospheric and stratospheric chemistry. Both configurations include prognostic aerosols. Subseasonal reforecasts were carried out following the SubX protocol for years 1999 - 2020 with weekly start dates for each year for CESM2, and with weekly start dates only between September and March for CESM2(WACCM6). Near real-time forecasts with the model have been running since September 2020 for CESM2(WACCM6) and since April 2021 for CESM2.

We demonstrated that the prediction skill of 2m temperature and precipitation as well as of the MJO and NAO are comparable to the prediction skill for these variables in CESM1 and similar to the skill seen in some operational models (NOAA's CFSv2 and ECMWF). The high subseasonal prediction skill of this research framework, along with extensive output obtained for all model components, makes it an excellent tool for studies of subseasonal predictability. We demonstrated that stratospheric-tropospheric coupling during SSW events is well represented in CESM2(CAM6) and CESM2(WACCM6), which implies that both configurations will likely capture well surface impacts of these events. This will be investigated in future studies. CESM2(WACCM6) can also be used for predictability research of the dynamics of the stratosphere and the mesosphere and lower-thermosphere region. We further demonstrated that variability in the MLT region is predictable ~ 10 days in advance of SSWs.

In general, the subseasonal prediction skill of tropospheric atmospheric variables is very similar between CESM2(CAM6) and CESM2(WACCM6). Therefore, the differences either in ocean and atmosphere initialization procedures or differences in model lids and representation of the stratosphere have not translated into many significant differences in prediction skills of the variables examined here. Nevertheless, the noted differences in skill include higher DJF 2m temperature skill in eastern Asia in CESM2(WACCM6) as compared to CESM2(CAM6), and higher 2m temperature skill in parts of North America in CESM2(CAM6), both for weeks 5-6. Stratospheric-tropospheric coupling is well-represented in both models, however, the polar cap geopotential anomalies following an SSW are more consistent with observations in CESM2(WACCM6) as compared to CESM2. The impact of this difference on predictability of surface extreme weather associated with SSWs will be investigated in future work.

CESM2(CAM6) and CESM2(WACCM6) are freely available for use by the community.

The reforecast sets described here are publicly available and are designed to serve as a basis for future experiments elucidating sources of subseasonal predictability. The near real-time forecasts are contributing to the NOAA week 3-4 outlook. The extensive output from the atmospheric, land, ocean and sea-ice components of the model may open new avenues of research.

Acknowledgements and Data

This work was supported by the National Oceanic and Atmospheric Administration's Weather Program Office/Climate Test Bed Program, and by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement 1852977. Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via NSF Interagency Agreement 1844590. Portions of this study were supported by the NOAA Office of Weather and Air Quality Research Programs (OWAQ) (Subseasonal to Seasonal) under grant NA19OAR4590156.Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. We thank the National Aeronautics and Space Administration (NASA) for making the FP-IT data available to us for this project. MERRA-2 is available from NASA's Global Modeling and Assimilation office at https://gmao.gsfc.nasa.gov/ reanalysis/MERRA-2/. CPC Global Temperature data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their Web site at https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html. CESM2(CAM6) and

CESM2(WACCM6) reforecast outputs are available for download from the NCAR Climate Data Gateway and can be accessed via the following DOI's: https://doi.org/10.5065/0s63-m767 and https://doi.org/10.5065/ekns-e430 respectively.

References

Beljaars, A. C. M., Brown, A. R., & Wood, N. (2004). A new parameterization of turbulent orographic form drag. *Quarterly Journal of the Royal Meteorological Society*, 130, 1327–1347. https://doi.org/10.1256/qj.03.73

Baldwin, Mark P., & T. J. Dunkerton (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, 294.5542, 581-584, 10.1126/science.1063315.

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., et al. (2021). Sudden stratospheric warmings. *Reviews of Geophysics*, 59, e2020RG000708. https://doi.org/10.1029/2020RG000708

Berner, J., Fossell K. R., Ha S.-Y., Hacker J. P. & Snyder C. (2015). Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Monthly Weather Review,* 143.4, 1295-1320. https://doi.org/10.1175/MWR-D-14-00091.1

Boer, G. J., & Hamilton, K. QBO influence on extratropical predictive skill, *Clim. Dyn*., 31, 987–1000 (2008).

Butler, A. H., Sjoberg, J. P., Seidel, D. J., & Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth System Science Data*, 9, 63–76, https://doi.org/10.5194/essd-9-63-2017.

Chandran, A., Collins, R. L., Garcia, R. R., Marsh, D. R., Harvey, V. L., Yue, J., & de la Torre, L. (2013). A climatology of elevated stratopause events in the whole atmosphere community climate model, *J. Geophys. Res. Atmos.*, 118, 1234- 1246, doi:10.1002/jgrd.50123.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., et al. (2012). The CCSM4 ocean component. *Journal of Climate*, 25, 1361–1389. https://doi.org/10.1175/JCLI-D-11-00091.1

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001916. https://doi.org/ 10.1029/2019MS001916

Davis, A. N., J. H. Richter, J. Edwards, & A. A. Glanville (2021): A positive zonal wind feedback on sudden stratospheric warming development revealed by CESM2 (WACCM6) reforecasts, *Geophysical Research Letters,* 48, e2020GL090863. https://doi.org/10.1029/2020GL090863

de Andrade, F. M., Coelho, C. A. S., & Cavalcanti, I. F. A. (2019). Global precipitation hindcast quality assessment of the subseasonal to seasonal (S2S) prediction project models. *Climate Dynamics*, *52*(9-10), 5451–5475. https://doi.org/10.1007/s00382-018-4457-z

Diro, G. T., & Lin, H. (2020). Subseasonal Forecast Skill of Snow Water Equivalent and Its Link with Temperature in Selected SubX Models, *Weather and Forecasting*, *35*(1), 273-284.

Domeisen, D.I.V. & Butler, A.H. (2020). Stratospheric drivers of extreme events at the Earth's surface. *Communications Earth and Environment*, 1**,** 59, https://doi.org/10.1038/s43247-020-00060-z

Ford, T.W., Dirmeyer, P.A. & Benson, D.O. Evaluation of heat wave forecasts seamlessly across subseasonal timescales. *npj Climate and Atmospheric Science* 1, 20 (2018). https://doi.org/10.1038/s41612-018-0027-7

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Gettelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, 28, 1268–1287. https://doi.org/10.1175/JCLI-D-14-00102.1

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., et al. (2019). The Whole Atmosphere Community Climate Model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, 124(23), 12,380–12,403. https://doi.org/10.1029/2019JD030943

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Sciences*, 59, 3540–3551.

Griffies, S. M., Danabasoglu G., Durack P. J., Adcroft A. J., Balaji V., Böning C. W., et al. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the ocean model intercomparison project. *Geoscience Model Development*, 9, 3231–3296, https://doi.org/10.5194/gmd-9-3231-2016.

Zhu H., H. Chen, Y. Zhou & X. Dong (2019). Evaluation of the subseasonal forecast skill of surface soil moisture in the S2S database, *Atmospheric and Oceanic Science Letters,* 12:6, 467-474, DOI: 10.1080/16742834.2019.1663123

Hersbach, H., Bell, B., Berrisford, P., Hirahara S., Horányi A., Muñoz-Sabater J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of Royal Meteorological Society*, 146, 1999–2049, https://doi.org/10.1002/qj.3803

Hunke, E. C., Hebert, D. A., & Lecomte, O. (2013). Level-ice melt ponds in the Los Alamos sea ice model, CICE. *Ocean Modelling*, 71, 26–42. https://doi.org/10.1016/j.ocemod.2012.11.008

Hunke, E. C, Lipscomb, W. H., Turner, A. K., Jeffery, N., & Elliott, S. (2015). CICE: The Los Alamos Sea Ice Model. Documentation and Software User's Manual. Version 5.1. T-3 Fluid Dynamics Group, Los Alamos National Laboratory, Tech. Rep. LA-CC-06-012.

Hurrell, J. W., (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science,* 269, 676–679, https://doi.org/10.1126/science.269.5224.676.

Jackson, D. R., Fuller-Rowell, T. J., Griffin, D. J., Griffith, M. J., Kelly, C. W., Marsh, D. R., & Walach, M.-T. (2019). Future directions for whole atmosphere modeling: Developments in the context of space weather. *Space Weather*, 17, 1342– 1350. https://doi.org/10.1029/2019SW002267

Kim, H., Vitart, F., & Waliser, D. E. (2018). Prediction of the Madden–Julian Oscillation: A Review. *Journal of Climate*, *31*(23), 9425–9443. https://doi.org/10.1175/JCLI-D-18-0210.1

Kim, H., Richter, J. H., & Martin, Z. (2019a). Insignificant QBO-MJO prediction skill relationship in the SubX and S2S subseasonal reforecasts. *Journal of Geophysical Research: Atmospheres*, *124*, 12,655–12,666. https://doi.org/10.1029/2019JD031416

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

Lambert, A., Read, W. & Livesey, N. (2015). MLS/Aura Level 2 Water Vapor (H2O) Mixing Ratio V004, Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, 10.5067/Aura/MLS/DATA2009

Larson, V. E., (2017). CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere. arXiv:1711.03675v2 [physics.ao-ph].

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model Version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11, 4245–4287. https://doi.org/10.1029/2018MS001583

Leutbecher, M., & Tim N. Palmer (2008). Ensemble forecasting. *Journal of computational physics*, 3515-3539.

Leutbecher, M., Lock S.-J., Ollinaho P., Lang, S. T.K., Balsamo, G., Bechtold P., et al. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society* 143.707 (2017): 2315-2339. https://doi.org/10.1002/qj.3094

Levis, S., Badger, A., Drewniak, B., Nevison, C., & Ren, X. L. (2018). CLMcrop yields and water requirements: Avoided impacts by choosing RCP 4.5 over 8.5. *Climatic Change*, 146, 501–515. https://doi.org/10.1007/s10584-016-1654-9

Li, F., Levis, S., & Ward, D. S. (2013). Quantifying the role of fire in the Earth system—Part 1: Improved global fire modeling in the Community Earth System Model (CESM1). *Biogeosciences*, 10, 2293–2314. https://doi.org/10.5194/bg-10-2293-2013

Li, F., & Lawrence, D. M. (2017). Role of fire in the global land water budget during the twentieth century due to changing ecosystems. *Journal of Climate*, 30, 1893–1908. https://doi.org/10.1175/JCLI-D-16-0460.1

Lim, Y., Son, S., & Kim, D. (2018). MJO Prediction Skill of the Subseasonal-to-Seasonal Prediction Models, *Journal of Climate*, *31*(10), 4075-4094. Retrieved Feb 14, 2021, from https://journals.ametsoc.org/view/journals/clim/31/10/jcli-d-17-0545.1.xml

Lim, Y., Son, S.-W., Marshall, A. G., Hendon, H. H., & Seo, K.-H. (2019). Influence of the QBO on MJO prediction skill in the subseasonal- to-seasonal prediction models. *Climate Dynamics*, 1–15.

Lipscomb, W. H., Price, S. F., Hoffman, M. J., Leguy, G. R., Bennett, A. R., Bradley, S. L., et al. (2019). Description and evaluation of the Community Ice Sheet Model (CISM) v. 2.1. *Geoscientific Model Development*, 12, 387–424. https://doi.org/10.5194/gmd-12-387-2019

Liu, X., Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., et al. (2016). Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within Version 5.3 of the Community Atmosphere Model. *Geoscientific Model Development*, 9, 505–522. https://doi.org/10.5194/gmd-9-505-2016

Liu, H.-L., Bardeen, C. G., Foster, B. T., Lauritzen, P., Liu, J., Lu, G., Wang, W. et al., (2018). Development and validation of the Whole Atmosphere Community Climate Model with thermosphere and ionosphere extension (WACCM-X 2.0). *Journal of Advances in Modeling Earth Systems*, 10, 381– 402. https://doi.org/10.1002/2017MS001232

Marshall, A. G., & Scaife, A. A. (2010): Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution, *Journal of Geophysical Res*earch, 115, D16114, doi:10.1029/2009JD012643

Miyazaki, K., Iwasaki, T., Shibata, K., Deushi, M., & Sekiyama, T. T. (2005). The impact of changing meteorological variables to be assimilated into GCM on ozone simulation with MRI CTM, *Journal of the Meteorological Society of Japan. Ser. II*, 83, 909-918, https://doi.org/10.2151/jmsj.83.909.

Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., & Fung, I. Y. (2002). An intermediate complexity marine ecosystem model for the global domain. *Deep Sea Research*, 49, 403–462. https://doi.org/10.1016/S0967-0645(01)00108-4

Moore, J. K., Doney, S. C., & Lindsay, K. (2004). Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochemical Cycles,* 18, GB4028. https://doi.org/10.1029/2004GB002220

Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., & Misumi, K. (2013). Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 scenarios. *Journal of Climate*, 26, 9291–9312. https://doi.org/10.1175/JCLI-D-12-00566.1

Mote, P. W., Rosenlof, K. H., McIntyre, M. E., Carr, E. S., Gille, J. C., Holton, J. R., et al. (1996). An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, *Journal of Geophysical Research*, 101(D2), 3989– 4006, doi:10.1029/95JD03422.

NAS, 2016: Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. The National Academies Press, 350 pp., https://doi.org/10.17226/21873.

Oleson, K. W., & Feddema, J. (2019). Parameterization and surface data improvements and new capabilities for the Community Land Model Urban (CLMU). *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2018MS001586

Pedatella, N. M., Liu, H.-L., Marsh, D. R., Raeder, K., Anderson, J. L., Chau, J. L., et al. (2018a). Analysis and hindcast experiments of the 2009 sudden stratospheric warming in WACCMX+DART. *Journal of Geophysical Research: Space Physics*, 123, 3131– 3153. https://doi.org/10.1002/2017JA025107

Pedatella, N. M., J. L. Chau, H. Schmidt, L. P. Goncharenko, C. Stolle, K. Hocke, V. L. Harvey, B. Funke, and T. A. Siddiqui (2018). *EOS*. https://eos.org/features/how-sudden-stratospheric-warming-affects-the-whole-atmosphere

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

Rashid, H. A., Hendon, H. H., Wheeler, M. C., & Alves, O. (2011). *Climate Dynamics*, 36(3-4), 649–661. https://doi.org/10.1007/s00382-010- 0754-x

Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a Physically Based Gravity Wave Source Parameterization in a General Circulation Model. *Journal of the Atmospheric Sciences*, *67*(1), 136–156. https://doi.org/10.1175/2009JAS3112.1

Richter, J. H., Pegion K., Sun L, Kim H., Caron J. M., Glanville A., et al. (2020). Subseasonal prediction with and without a well-represented stratosphere in CESM1. *Weather and Forecasting*, doi: https://doi.org/10.1175/WAF-D-20-0029.1.

Riddle, E. E., Butler, A. H., Furtado, J. C., Cohen, J. L., & Kumar, A. (2013). CFSv2 ensemble prediction of the wintertime Arctic Oscillation, *Climate Dynamics*, 41, 1099–1116.

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

Scaife, A. A., C. K. Folland, L. V. Alexander, A. Moberg, & J. D. Knight (2008). European climate extremes and the North Atlantic Oscillation. *Journal of Climate*, 21, 72–83, https://doi.org/ 10.1175/2007JCLI1631.1.

Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters, *Geophysical Research Letters*, 41, 2514–2519.

Scaife, A.A., Athanassiadou M., Andrews M., Arribas A., Baldwin M., Dunstone N., et al. (2014b). Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales, *Geophysical Research Letters*, 41, 1752-1758, 10.1002/2013GL059160 (2014b)

Scinocca, J., & N. Mcfarlane (2000). The parametrization of drag induced by stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological Society*, 126, 2353–2394. https://doi.org/10.1256/smsqj.56801

Smith, R., Jones P., Briegleb B., Bryan F., Danabasoglu G., Dennis J., et al. (2010). The Parallel Ocean Program (POP) reference manual, Ocean component of the Community Climate System Model (CCSM), *LANL Technical Report*, LAUR-10-01853, 141 pp.

Sun, L., Perlwitz, J., Richter, J. H., Hoerling, M. P., & Hurrell, J. W. (2020). Attribution of NAO predictive skill beyond 2 weeks in boreal winter. *Geophysical Research Letters*, 47, e2020GL090451. https://doi.org/ 10.1029/2020GL090451

Stan C., Straus, D. M., Frederiksen, J. S., Lin, H., Maloney, E. D., & Schumacher, C., et al. (2017). Review of Tropical-Extratropical Teleconnections on Intraseasonal Time Scales. *Review of Geophysics*, 55, 902-937.

Tolman, H. L. (2009). User manual and system documentation of WAVEWATCH III TM version 3.14. *Technical note, MMAB Contribution,* 276, p.220.

Tripathi, O. P., Baldwin, M., Charlton-Perez, A., Charron, M., Eckermann, S.D., Gerber, E., et al. (2015). The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quarterly Journal of the Royal Meteorological Society*, 141: 987-1003. https://doi.org/10.1002/qj.2432

Tsujino, H., Urakawa S., Nakano H., Small R. J., Kim W. M., Yeager S. G., et al. (2018). JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modelling*, 130, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002.

Tsujino, H., Urakawa L. S., Griffies S. M., Danabasoglu G., Adcroft A. J., Amaral A. E., et al. (2020). Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2). *Geoscience Model Development*, 13, 3643–3708, https://doi.org/10.5194/ gmd-13-3643-2020.

Vitart, F., C. Ardilouze, A. Bounet, A. Brookshaw, M. Chen, C. Codorean, et al. (2017). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society,* 98, 163–173. doi:10.1175/ BAMS-D-16-0017.1.

Vitart, F. (2017). Madden–Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, *143*, 2210–2220. https://doi.org/10.1002/qj.3079

Vitart, F., & A. W. Robertson (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. npj *Climate Atmospheric Science*, 1, 3, https://doi.org/10.1038/S41612- 018-0013-0.

Wang H., Akmaev, R. A., Fang, T.-W., Fuller-Rowell, T. J., Wu, F., Maruyama, N., & Iredell, M. D. (2014). First forecast of a sudden stratospheric warming with a coupled whole-atmosphere/ionosphere model IDEA, *Journal of Geophysical Research: Space Physics*, 119, 2079– 2089, doi:10.1002/2013JA019481.

Wang, L., & Robertson, A. W. (2019). Week 3–4 predictability over the United States assessed from two operational ensemble prediction systems. *Climate Dynamics*, 52(9-10), 5861–5875. https://doi.org/10.1007/s00382-018-4484-9.

Wayand, N. E., Bitz, C. M., & Blanchard-Wrigglesworth, E. (2019). A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophysical Research Letters*, *46*, 3298–3307. https://doi.org/10.1029/2018GL081565

White, C. J., and Carlsen H., Robertson A. W., Klein R. J. T., Lazo J. K., Kumar A., et al. (2017). Potential applications of Subseasonal-to-Seasonal (S2S) predictions. *Meteorological Applications*, 24, 315–325, https://doi.org/10.1002/met.1654.

Wilks, D. S. (2011). Statistical Methods in the Atmospheric Sciences. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.

Xiang, B., Sun, Y. Q., Chen, J.-H., Johnson, N. C., & Jiang, X. (2020). Subseasonal prediction of land cold extremes in boreal wintertime. *Journal of Geophysical Research: Atmospheres*, *124*, e2020JD032670. https://doi.org/ 10.1029/2020JD032670

Yamagami, A., & Matsueda, M. (2020). Subseasonal forecast skill for weekly mean atmospheric variability over the Northern Hemisphere in winter and its relationship to midlatitude teleconnections. *Geophysical Research Letters*, *47*, e2020GL088508. https://doi. org/10.1029/2020GL088508

Zampieri, L., Goessling, H. F., & Jung, T. (2018). Bright prospects for Arctic Sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, *45*, 9731–9738. https://doi.org/10.1029/2018GL079394

Zar, J.H., (2014). Spearman Rank Correlation: Overview. Wiley StatsRef: Statistics Reference Online. doi:10.1002/9781118445112.stat05964

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Center general circulation model. *Atmosphere-Ocean*, *33*, 407–446.

Zuo J., H-L. Ren, J. Wu, Y. Nie, Q. Li (2016). Subseasonal variability and predictability of the Arctic Oscillation/North Atlantic Oscillation in BCC_AGCM2.2, Dynamics of Atmospheres and Oceans, 75, pp 33-45, https://doi.org/10.1016/j.dynatmoce.2016.05.002

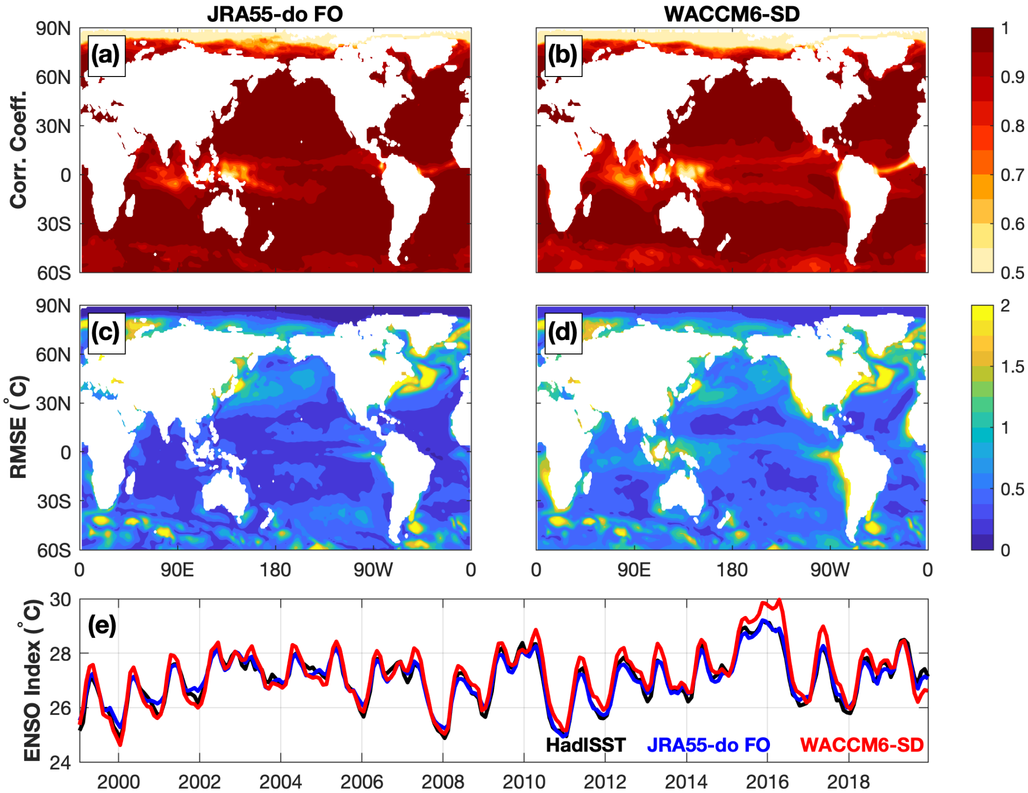

Figure 1: Correlation coefficients for SST from (a) JRA55-do FO (b) WACCM6-SD in comparison with HadISST observations. RMSE for (c) JRA55-do FO and (d) WACCM6-SD when compared with the same observations. (e) The ENSO Nino3.4 Index (e) for all the datasets. All calculations use monthly data from 1999-2019.

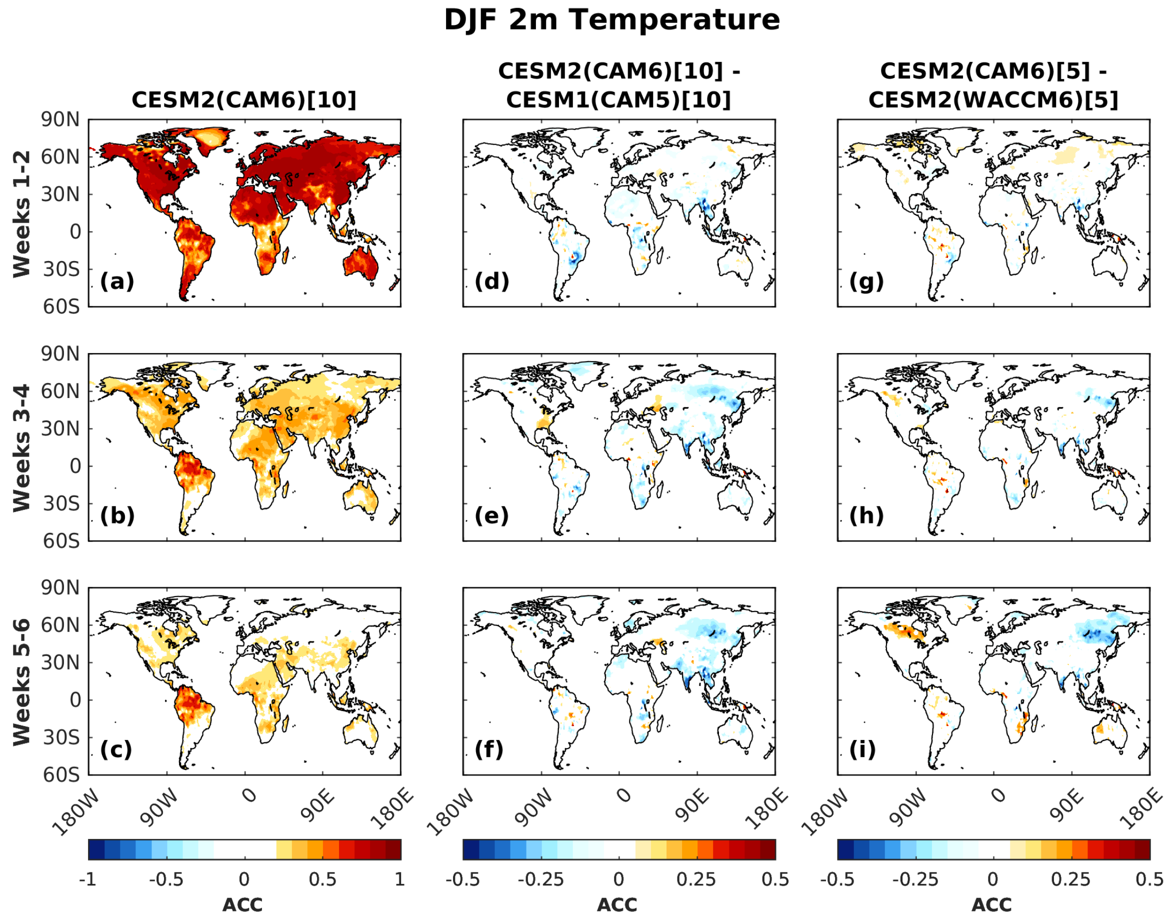

Figure 2: The DJF 2m temperature ACC for CESM2(CAM6) over land for (a) weeks 1-2 (day 1-14 averaged), (b) weeks 3-4 (day 15-28 averaged), and (c) weeks 5-6 (day 29-42 averaged). Using the same biweekly separation, panels (d)-(f) show the difference in ACC for CESM2(CAM6) minus CESM1(CAM5) and panels (g)-(i) show the difference in ACC for CESM2(CAM6) minus CESM2(WACCM6). Data in the difference plots that fall below the 95% confidence level using a Fisher z transformation are omitted. Note the different colorbar ranges. All calculations use daily data from 1999-2015. The number of ensemble members used in the analysis is given in the column titles in the square brackets.

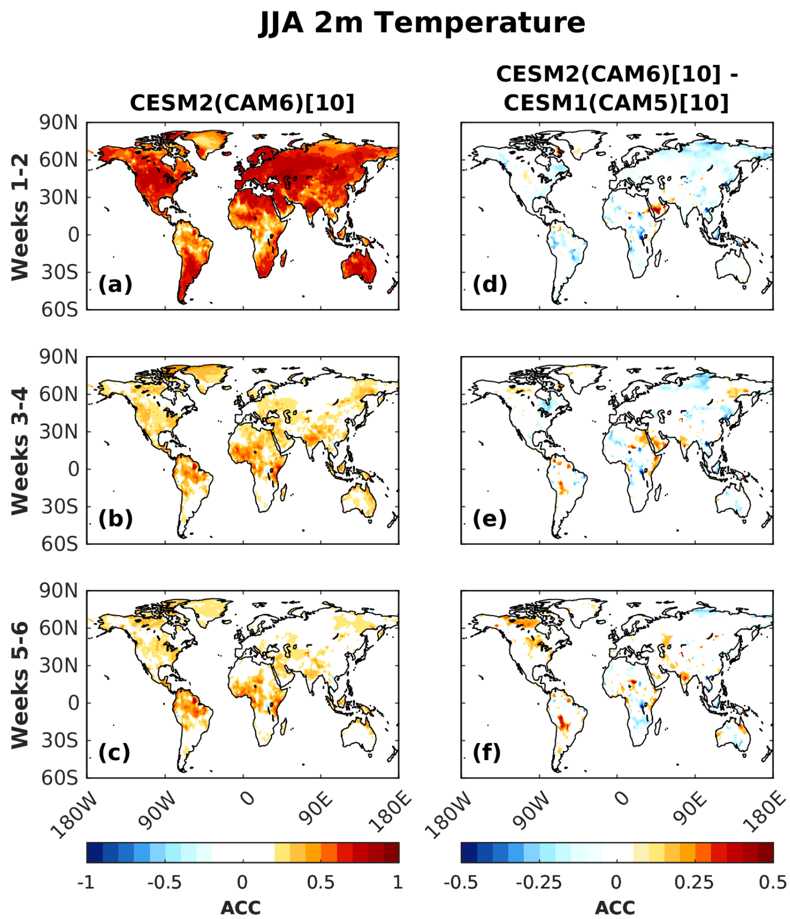

Figure 3: Same as Figure 2 but for JJA. Note, there is no CESM2(WACCM6) data for April - August.

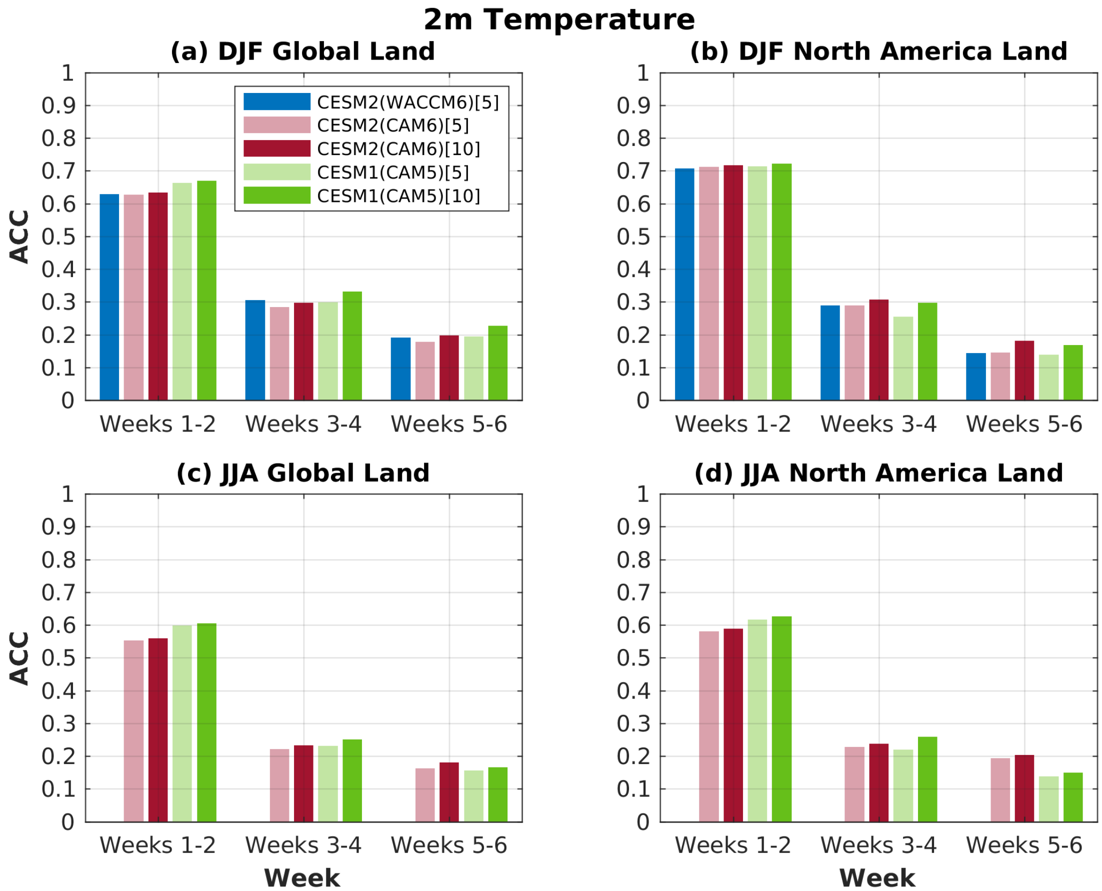

Figure 4: DJF 2m temperature ACC averaged over (a) the global land and (b) North American land. Panels (c) and (d) are the same as panels (a) and (b) but for JJA. Note, there is no CESM2(WACCM6) data for April - August. All calculations use daily data from 1999-2015. In the legend, the number of ensemble members used in the calculations is shown in square brackets.

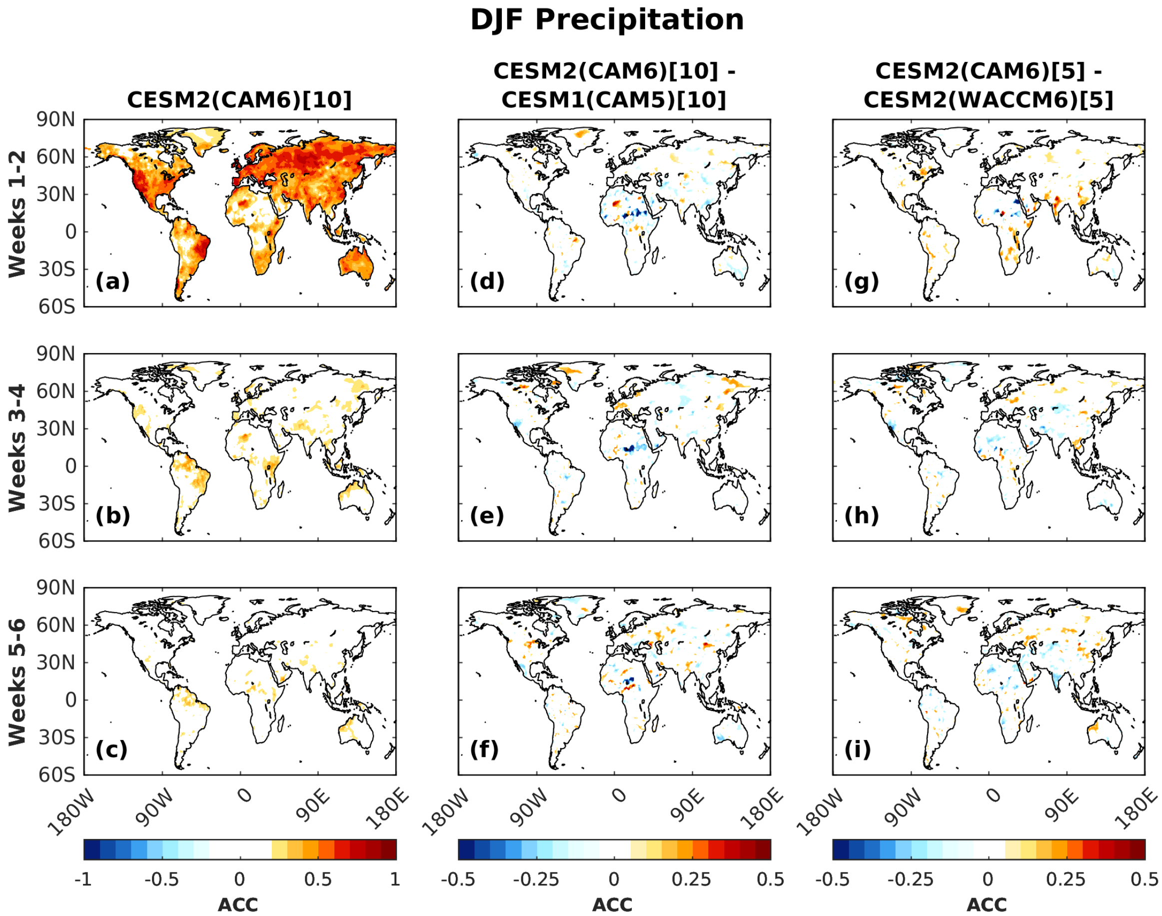

Figure 5: Same as Figure 2 but for precipitation.

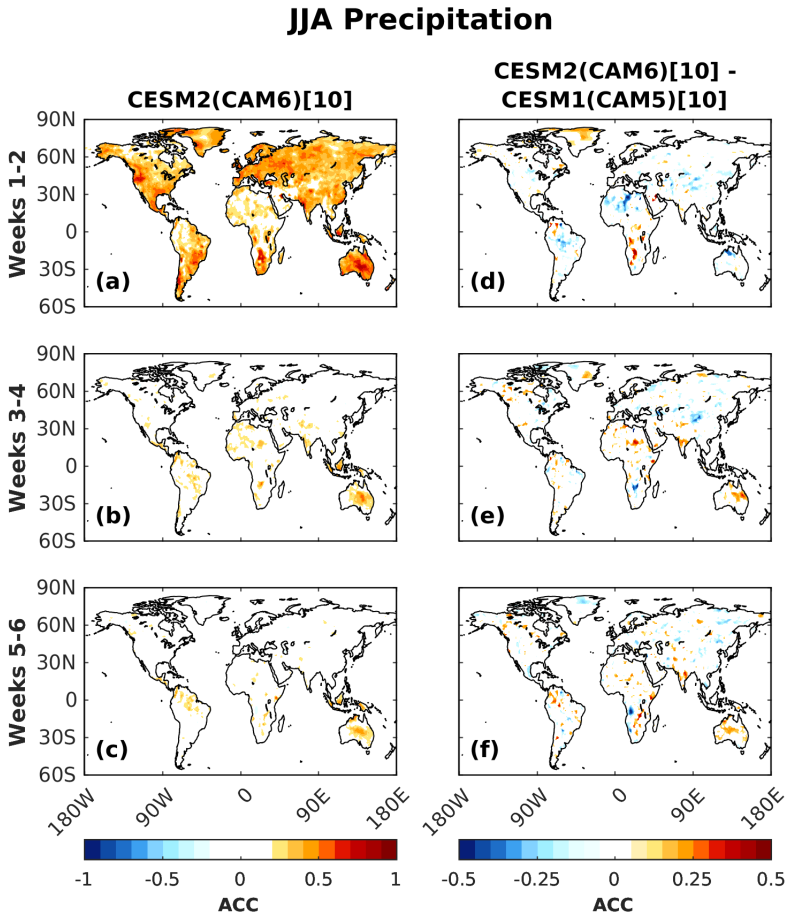

Figure 6: Same as Figure 3 but for precipitation.

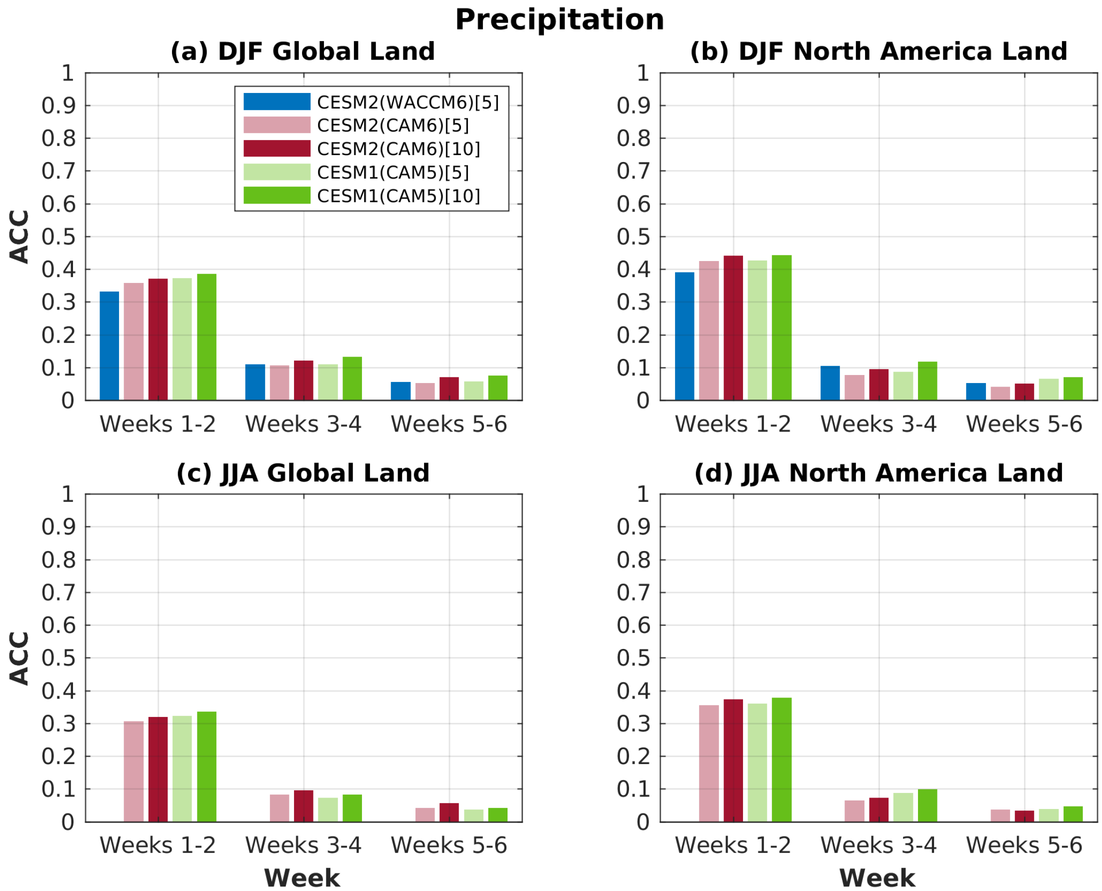

Figure 7: Same as Figure 4 but for precipitation.

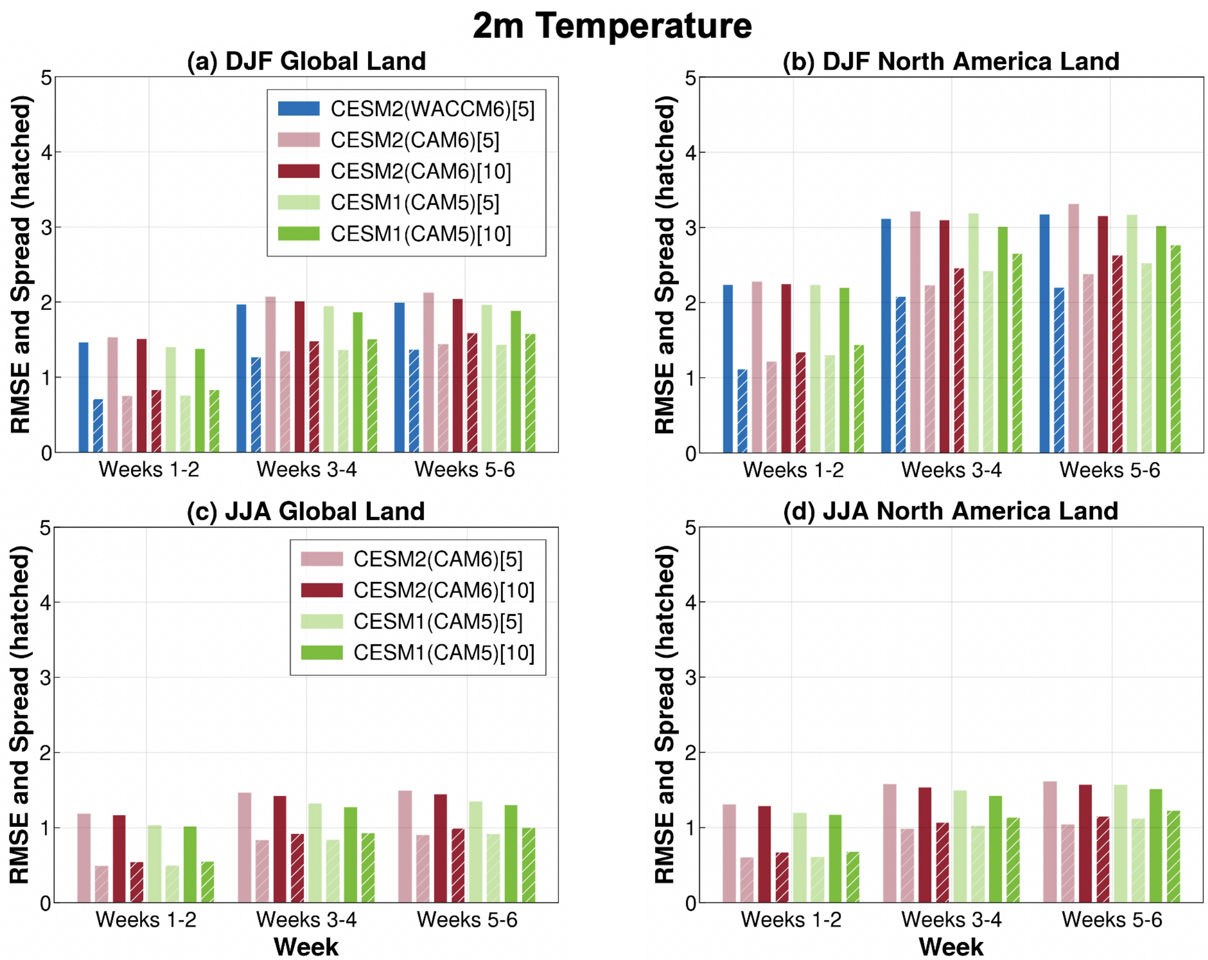

Figure 8. RMSE (solid bars) and spread (hatched bars) for 2m temperature for DJF (top) and JJA (bottom). Metrics are shown for global land (left) and North American land (right).

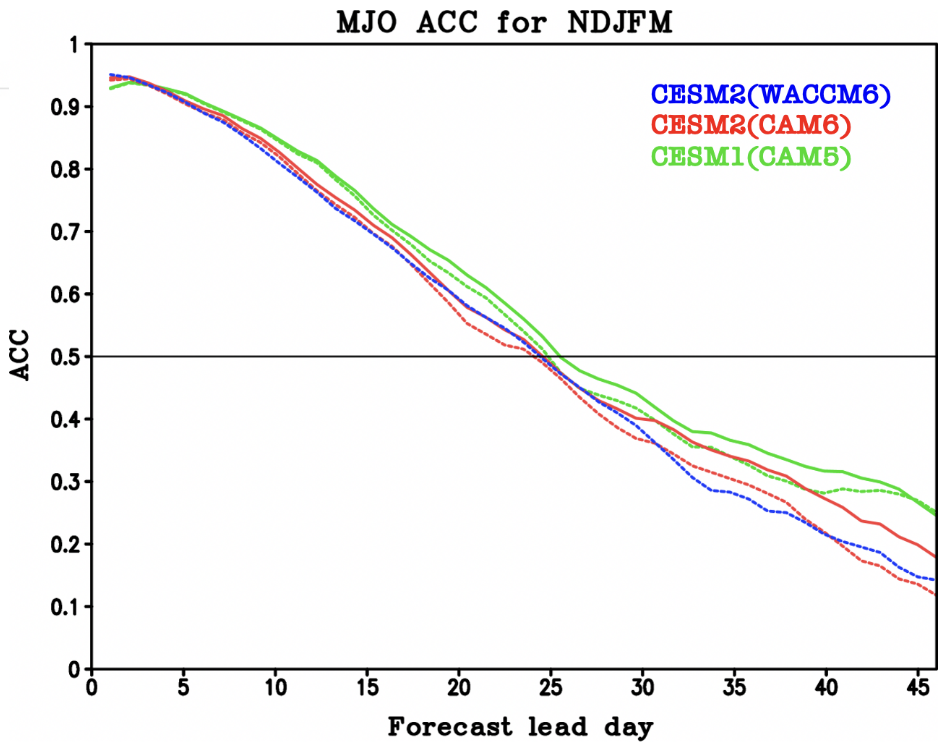

Figure 9: ACC for MJO for NDJFM from CESM1(CAM5), CESM2(CAM6), and CESM2(WACCM6). Solid (dashed) lines indicate the average of 10 (5) ensemble members.

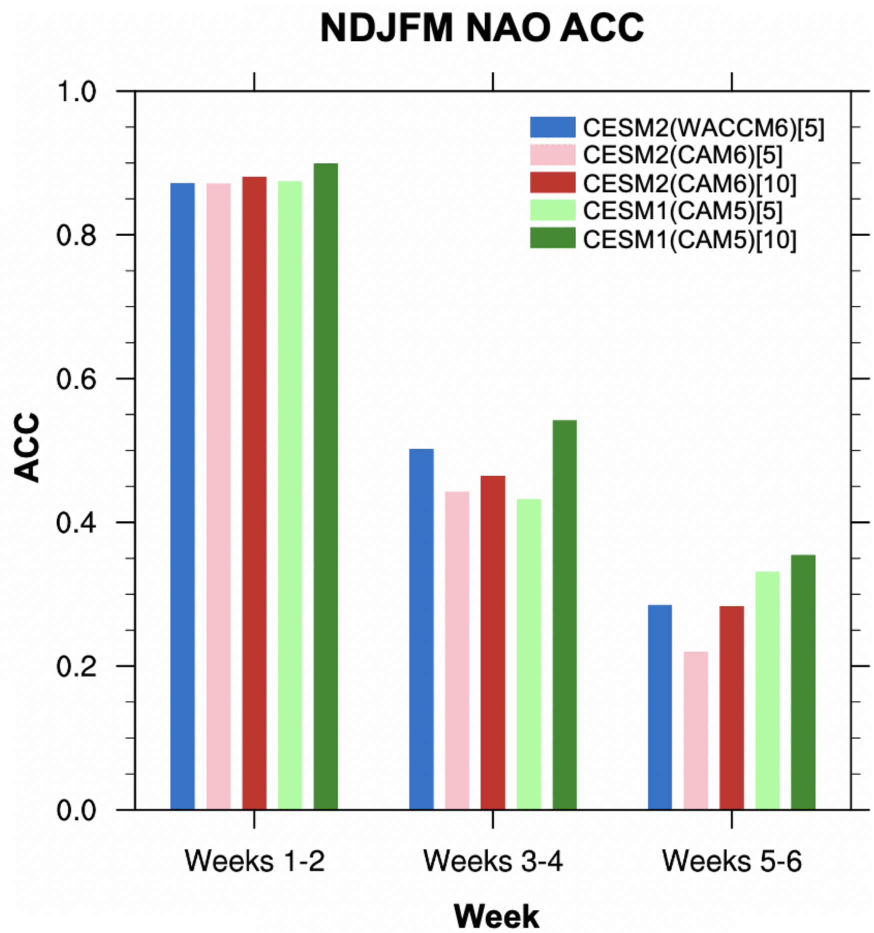

Figure 10: Biweekly NAO ACC in NDJFM from CESM1(CAM5), CESM2(CAM6), and CESM2(WACCM6). The number of ensemble members used in the analysis is given in the square brackets.

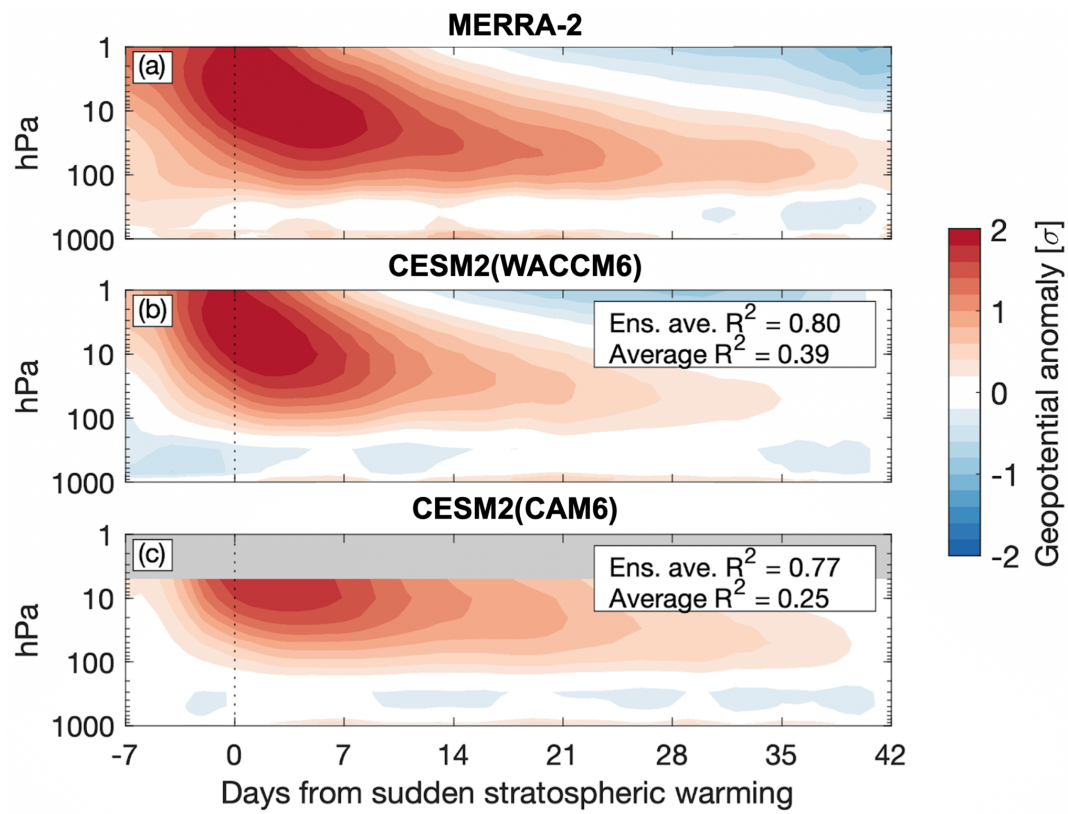

Figure 11: Standardized polar cap geopotential anomalies composited around the central SSW date for (a) MERRA-2, (b) CESM2(WACCM6) reforecasts, and (c) CESM2 reforecasts, shaded every 0.2 standard deviations. The squared pattern correlation between the ensemble average of the reforecasts and MERRA-2, as well as the average of all squared correlations between each individual ensemble member and MERRA-2 for every event, are displayed in the upper right of each panel.

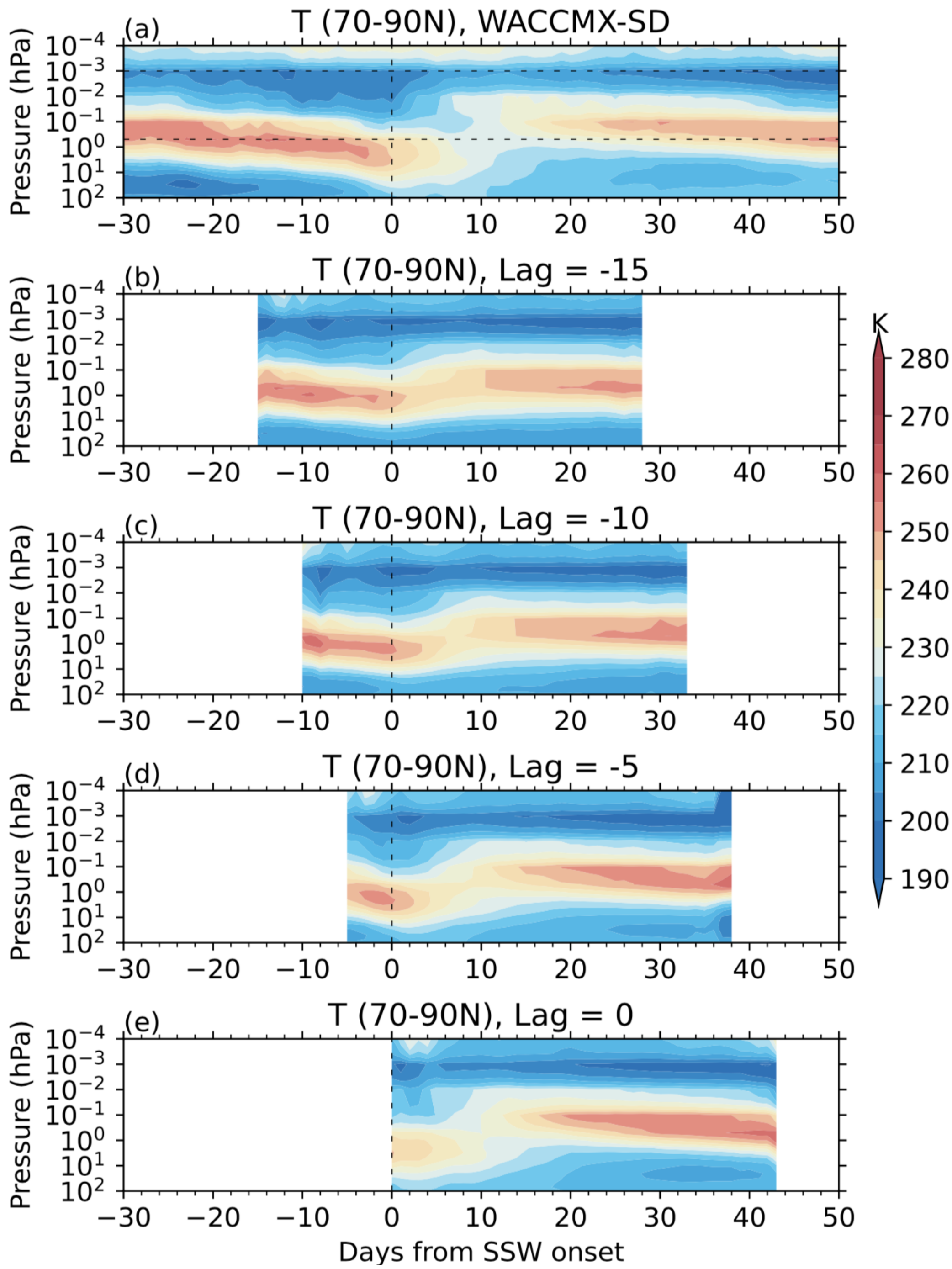

Figure 12: Composite of zonal-mean temperature (T) between 70°-90°N for 14 major SSW events in (a) WACCMX-SD, and CESM2(WACCM6) reforecasts initialized at a lag of (b) -15, (c) -10, (d) -5, and (e) 0 days prior to the SSW central date. The SSW onset date is defined as the zonal-mean zonal wind reversal at 60N and 10 hPa. The horizontal dashed lines in panel (a) mark the stratopause (~ 0.5 hPa), and mesopause (~10^{-3} hPa respectively).

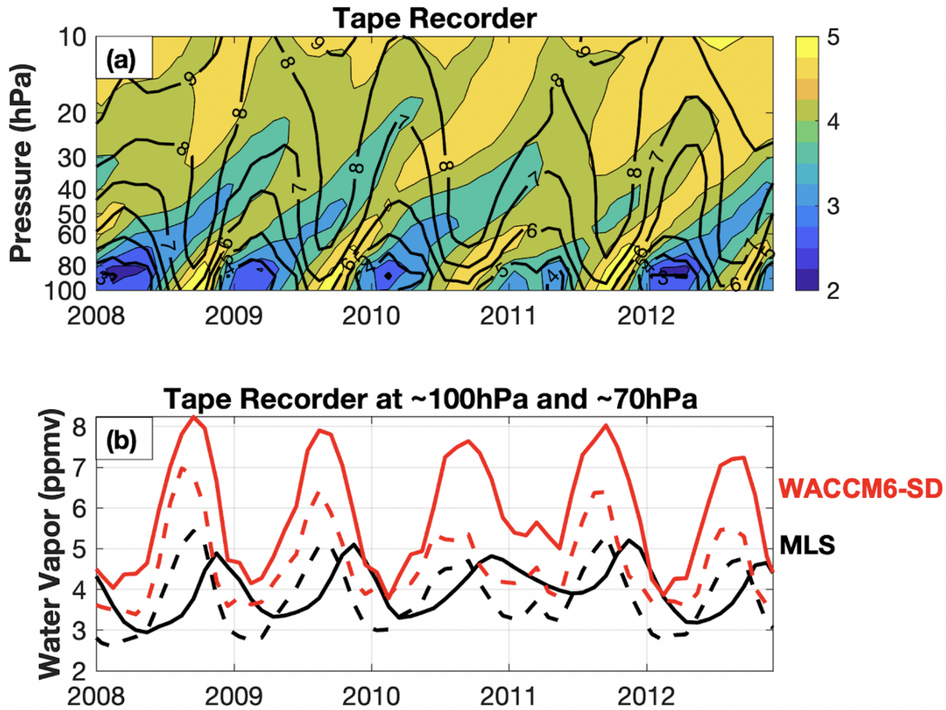

Fig 14: The 10ºS-10ºN tape recorder of water vapor (a) for MLS observations (filled color contours) and WACCM6-SD (back contour lines). The time series of the tape recorder at ~100hPa (dashed lines) and ~70hPa (solid lines) for MLS (black) and WACCM6-SD (red).

	Atmosphere	Land	Ocean & Sea-ice	Reforecast Period	Initialization Frequency	# Ens Members Reforecasts	# Ens Members Forecasts
CESM2 (CAM6)	CFSv2	CLM5 spun up with CFSv2	JRA55-do forced ocn/sea-ice	All months, 1999 - 2020	Every Monday*	11	21**
CESM2 (WACCM6)	WACCM6-SD run nudged to MERRA-2	CLM5 spun up with CFSv2	Hybrid: JRA55-do every 5 yrs / MERRA-2 forced ocn	Sep-Mar 1999 - 2020	Every Monday*	5	21***

Table 1: Summary of initialization methods for S2S reforecasts with CESM2(CAM6) and CESM2(WACCM6). *Reforecasts are started every Monday, except for leap years, in which case the reforecast was carried out on a Sunday; **Real-time forecasts with CESM2(CAM6) stared April 2021; *** Real-time forecasts with CESM2(WACCM6) started in September 2020.