Subseasonal Earth System Prediction with CESM2

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

Abstract: Prediction systems to enable Earth system predictability research on the subseasonal time scale have been developed with the Community Earth System Model, version 2 (CESM2) using two configurations that differ in their atmospheric components. One system 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, and it includes fully interactive tropospheric and stratospheric chemistry [CESM2(WACCM6)]. Both systems are utilized to carry out subseasonal reforecasts for the 1999–2020 period following the Subseasonal Experiment’s (SubX) protocol. Subseasonal prediction skill from both systems is compared to those of the National Oceanic and Atmospheric Administration CFSv2 and European Centre for Medium-Range Weather Forecasts (ECMWF) operational models. CESM2(CAM6) and CESM2(WACCM6) show very similar subseasonal prediction skill of 2-m temperature, precipitation, the Madden-Julian oscillation, and North Atlantic Oscillation to its previous version and to the NOAA CFSv2 model. Overall, skill of CESM2(CAM6) and CESM2(WACCM6) is a little lower than that of the ECMWF system. In addition to typical output provided by subseasonal prediction systems, CESM2 reforecasts provide comprehensive datasets 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 from CESM2(WACCM6). It is shown that sudden stratosphere warming events, and the associated variability in the MLT, can be predicted ~10 days in advance. Weekly real-time forecasts and reforecasts with CESM2(CAM6) and CESM2(WACCM6) are freely available.

Significance Statement: We describe here the design and prediction skill of two subseasonal prediction systems based on two configurations of the Community Earth System Model, version 2 (CESM2): CESM2 with the Community Atmosphere Model, version 6 [CESM2(CAM6)] and CESM 2 with Whole Atmosphere Community Climate Model, version 6 [CESM2(WACCM6)] as its atmospheric component. These two systems provide a foundation for community-model based subseasonal prediction research. The CESM2(WACCM6) system provides a novel capability to explore the predictability of the stratosphere, mesosphere, and lower thermosphere. Both CESM2(CAM6) and CESM2(WACCM6) demonstrate subseasonal surface prediction skill comparable to that of the NOAA CFSv2 model, and a little lower than that of the ECMWF forecasting system. CESM2 reforecasts provide a comprehensive dataset for predictability research of multiple aspects of the Earth system, including the whole atmosphere up to 140 km, land, and sea ice. Weekly real-time forecasts, reforecasts, and models are publicly available.

Keywords: Forecast verification/skill; Forecasting; Seasonal forecasting

1. Introduction

Interest and demand for skillful subseasonal predictions (i.e., targeting 3–4 weeks) of the Earth system have 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 for research on 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 Robertson 2018) and the 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 time scales. The existing databases and subseasonal prediction...
models, however, do not easily allow researchers to run their own experiments that would elucidate sources of predictability, and hence there is a need for a subseasonal prediction research models accessible to the broader community.

In addition, 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 time scale, 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 and Matsueda 2020). A few recent studies have started examining the predictability of sea ice (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 (Zhu et al. 2019; Diro and Lin 2020); however, predictability of other characteristics of land has not been explored. Several studies have looked at the skill in predicting sudden stratospheric warmings (SSWs), as they can significantly impact surface extreme weather especially over Eurasia (Tripathi et al. 2015): role of stratospheric initial conditions on the NAO (Nie et al. 2019); and the predictability of the quasi-biennial oscillation (QBO) which impacts the MJO (Lim et al. 2019; Kim et al. 2019a). However, 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. Prior studies of the mesosphere and lower-thermosphere (MLT) variability have been limited as they used short reforecast periods (Wang et al. 2014; Pedatella et al. 2018a,b).

Richter et al. (2020) described the utility of the Community Earth System Model, version 1 (CESM1), with the Community Atmosphere Model version 5 (CAM5) as its atmospheric component [CESM1(CAM5)], a predecessor of CESM, version 2 (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 Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (CFSv2) operational model. Here we describe the design and prediction skill of two subseasonal prediction systems based on two configurations of CESM2: one with CAM, version 6 [CESM2(CAM6)] and another with Whole Atmosphere Community Climate Model, version 6 CESM2(WACCM6) as its atmospheric component. These prediction research systems provide a new resource for research on subseasonal predictability of multiple components of the Earth system that is freely available to the community. CESM2 is the newest version of the coupled Earth system model developed at National Center for Atmospheric Research in collaboration with the community and 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, incorporating the interactions between them, and prognostic aerosols with CESM2(WACCM6) also having 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. SSW events are now recognized to have impacts throughout the whole atmosphere (Baldwin et al. 2021; Pedatella et al. 2018), including the mesosphere, thermosphere, and ionosphere, where they influence the day-to-day weather of the near-Earth space environment.

We compare here the prediction skill of several key variables for these models to the NOAA CFSv2 and ECMWF operational systems. CESM2(CAM6) and CESM2(WACCM6) utilize different initialization methods for the atmospheric and ocean components due to the different vertical atmospheric model domains and timing of various model developments. Hence, the simulations presented here are not designed to isolate the role of a well-resolved stratosphere in the absence of any other changes as was done in Richter et al. (2020), but primarily to provide a validation and description of the new systems, as well as provide an illustration of the capabilities of CESM2(WACCM6) in looking at middle atmosphere predictability. The reforecast sets described here are designed to serve as a basis for future experiments with CESM2(CAM6) and CESM2(WACCM6) to investigate sources of subseasonal predictability.

2. Model and system description

a. 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. It 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 in all its components (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). It 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 and 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 Gettelman et al. (2019). WACCM6 has the same horizontal resolution as CAM6; however, it has 70 vertical levels with a top near
4.5 \times 10^{-6} \text{ hPa} \sim 140 \text{ km). The representation of atmospheric physics is identical to that in CAM6, with the only exception being the representation of nonorographic gravity waves, which follows Richter et al. (2010) with changes to tunable parameters described in Gettelman et al. (2019). The higher model lid and parameterization of nonorographic 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 Gettelman 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 Gettelman 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 160 m. 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 its previous version, 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 and Lawrence 2017), multiple urban classes and updated urban energy model (Oleson and Feddema 2019), and improved representation of plant dynamics. The river transport model is the Model for Scale Adaptive River Transport (MOSART; 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.

On the NCAr Cheyenne Supercomputer, CESM2(CAM6) and CESM2(WACCM6) use \sim 3500 and \sim 28 000 cpu hours per model year, respectively. The subseasonal reforecast set with CESM2(CAM6) cost \sim 5 million Cheyenne cpu hours, which is a substantial allocation.

b. Initialization

CESM2 was primarily developed for studies of climate variability and change and was subsequently adapted for decadal, seasonal, and subseasonal predictions, with the goal of the system to eventually provide seamless Earth system prediction from weeks to decades. Hence, CESM2 does not employ data assimilation as is done in many operational S2S models (Johnson et al. 2005; Sandery et al. 2020), but instead the initialization procedures are built on the methods used for decadal prediction with CESM1 (Yeager et al. 2018). As such, subseasonal 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 locations of the two atmospheric model’s lids and also due to the inclusion of CESM2(WACCM6) forecasts in NOAA’s experimental Week 3–4 outlooks since September 2020, necessitating real-time forecasting ability. The SubX protocol requires that reforecasts and forecasts with a given model are completed with the same model setup, and hence CESM2(WACCM6) reforecasts and forecasts are run with initial ocean states that were available when the system was developed in 2019. 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 a stand-alone CLM5 simulation which employed a setup consisting of biogeochemistry-driven crops and glacial observations. A 700-yr spinup 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 spinup, while atmospheric forcings from NCEP CFSv2 were cycled between 1979 and 1985 until an approximate steady state was achieved (∼100 cycles). After soil moisture and temperatures stabilized with respect to the 1979–85 climate state, the CLM5 continued to be forced with NCEP CFSv2 up through the 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 a random field perturbation method at initial time which was shown to be as effective as other more sophisticated methods to generate model spread by Magnussen et al. (2009) and was utilized successfully in S2S reforecasts with CESM1(CAM5) (Richter et al. 2020). The random field perturbation method involves adding and subtracting a scaled difference between two random daily atmospheric conditions generated by CESM2 (CAM6) for the same month as the forecast initialization date using a historical simulation. 500 such differences were generated. At each hindcast initialization, these differences were chosen, randomly as perturbations and, after scaling, they were then added to and subtracted from the CFSv2 reanalysis interpolated onto the CAM6 grid. The perturbations are applied to the zonal and meridional wind components, temperature, specific humidity, and surface pressure. The first ensemble member is just the interpolated analysis without any perturbations. For subsequent ensemble members, a scaling factor of 0.15 was used to scale the perturbations, with each perturbation generating two ensemble members (one is added and one is subtracted), as was done in CESM1 (Richter et al. 2020). The scaling factor of 0.15 was determined to be optimal in producing sufficient, but not too large ensemble spread (Richter et al. 2020).

Ocean and sea ice initial conditions for CESM2(CAM6) come from an ocean–sea ice coupled configuration of CESM2 forced with the adjusted Japanese 55-year Reanalysis product state fields and fluxes (JRA55-do forcing; Tsujino et al. 2018). This simulation is referred to as JRA55-do forced ocean simulation (JRA55-do FO). This procedure constrains ocean temperature and salinity in the surface layer, but it does not directly constrain the sea surface height field. The simulation was integrated through five cycles of the 1958–2009 forcing, with the last cycle extended through 2019, following the protocol for the CMIP6-endorsed Ocean Model Intercomparison Project phase 2 (OMIP2; Griffies et al. 2016; Tsujino 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 setup 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 (zonal and meridional winds and temperature) were nudged to the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) (Gelaro et al. 2017) with a 1-hourly nudging time scale from 1999 to 2020. The 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 random field perturbation method was then used to generate the ensemble spread similarly to what was done for CESM2 (CAM6), with the random perturbations coming from a CESM2 (WACCM6) simulation. 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 setup, the ocean state drifts from the observed state and the JRA55-do simulation, hence every 5 years the ocean in the SD simulation is re-initialized with the ocean state from the JRA55-do FO simulation. Consequently, ocean re-initialization occurred on 1 January 1998, 2003, 2008, 2013, and 2018. We have developed the ability to update the JRA55-do in August 2020, and a final ocean re-initialization occurred on 31 August 2020 to bring the real-time system’s ocean as close to observation as possible at that time. Daily atmospheric, ocean, and sea ice initial conditions were output from the WACCM6-SD simulation for use in reforecasts.

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 HadISST observations (Rayner et al. 2003) (Figs. 1a and 1c) and between SSTs in WACCM6-SD (used to initialize CESM2(WACCM6) reforecasts) and HadISST (Figs. 1b and 1d). Over the 1999–2019 period, the correlation between

<table>
<thead>
<tr>
<th>Atmosphere</th>
<th>Land</th>
<th>Ocean and sea ice</th>
<th>Reforecast period</th>
<th>Initialization frequency</th>
<th>No. of ensemble members reforecasts</th>
<th>No. of ensemble members forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>CESM2(CAM6)</td>
<td>CLM5 spun up with CFSv2</td>
<td>JRA55-do forced ocean/sea ice</td>
<td>All months, 1999–2020</td>
<td>Every Monday</td>
<td>11</td>
<td>21*</td>
</tr>
<tr>
<td>CESM2(WACCM6)</td>
<td>CLM5 spun up with CFSv2</td>
<td>JRA55-do forced ocean</td>
<td>Sep–Mar 1999–2020</td>
<td>Every Monday</td>
<td>5</td>
<td>21*</td>
</tr>
</tbody>
</table>

* Reforecasts are started every Monday, except for leap years, in which case the reforecast is carried out on a Sunday.

1 Real-time forecasts with CESM2(CAM6) started in April 2021.

1 Real-time forecasts with CESM2(WACCM6) started in September 2020.
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°S, but particularly in the Arctic Ocean. The correlation coefficients are lower in those regions in the WACCM6-SD as compared to the JRA55-do FO simulation: averaged over 10°S–10°N (90°–50°S) the correlation coefficient is 0.86 (0.81) for WACCM6-SD and 0.91 (0.86) for JRA55-do FO. The RMSE distribution (Figs. 1c and 1d) is also very similar between JRA55-do FO and WACCM6-SD, with the largest RMSE differences between the two simulations appear to occur in the tropics. The larger RMSE in WACCM6-SD as compared to 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 Fig. 1e, which shows the El Niño Southern Oscillation (ENSO) index in the JRA55-do FO, WACCM6-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 noticeably from observations. This includes the period from ~2015 to 2016 and from 2019 to 2020. In short, there is a greater ocean drift in the CESM2(WACCM6) reforecasts which results from the drift generated in the initial conditions generating procedure.

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.

c. Protocol and output

The subseasonal reforecasts were carried out following the SubX protocol (Pegion et al. 2019) with weekly initializations every Monday from 1999 to 2020 for CESM2(CAM6), and for every Monday between September and March for CESM2(WACCM6). An 11-member ensemble was carried out for CESM2(CAM6) and a 5-member ensemble was carried out for CESM2(WACCM6) reforecasts. Because the computational cost of CESM2(WACCM6) is 8 times the cost of CESM2(CAM6), 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 2020, and in April 2021 with CESM2(CAM6), both with a 21-member ensemble. The subseasonal reforecast set with CESM2(CAM6) and CESM2(WACCM6) have extremely comprehensive output for the atmosphere, land, ocean, and sea ice components 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 the acknowledgments). 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 latitude–longitude 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 three-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 subseasonal 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 earlier reforecasts carried out with the default version of CESM1(CAM5) that utilizes the 30-level version of CAM5 (Richter et al. 2020) for the common period of 1999–2015. As the reforecasts with 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; Meehl et al. (2021)]. Therefore, in select figures, we also show CESM1(CAM5) and CESM2(CAM6) skill comparisons based on a 5-member ensemble, because that is what the CESM2(WACCM6) skill assessment is based on. Because all the CESM model versions described here use slightly different initial conditions, the comparisons shown can be used broadly to compare skill across these predictions systems, but not to isolate the role of model physics or model top on prediction skill. For reference, the prediction skill of the NOAA operational CFSv2 model (years 1999–2010, 4-member ensemble) and of the ECMWF Variable Resolution Ensemble Prediction System monthly forecast system (Wang and Robertson 2019; years 1996–2016, 11-member ensemble) are also presented.

a. 2-m temperature and precipitation prediction skill

Figures 2a–c show the anomaly correlation coefficients (ACC) for 2-m temperature over land for the December–January–February (DJF) means for weeks 1–2, 3–4, and 5–6 for CESM2(CAM6). The NOAA Climate Prediction Center (CPC) Global Daily Temperature dataset at the 0.5° × 0.5° 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 per the SubX protocol. Similar to what was done for CESM1(CAM5) 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 (1995–2015) considered here. Subsequently, we assume a 2-week decorrelation time, resulting in 110 independent samples. Thus, there are 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 2-m temperature over all land areas is ∼0.6 for weeks 1–2, ∼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 ∼0.5–0.6 through weeks 5–6) and the lowest values over north and northeastern Asia. The differences of DJF 2-m temperature ACC between CESM2(CAM6) and CESM1(CAM5) and between CESM2(CAM6) and CESM2(WACCM6) 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–f show that DJF ACC for 2-m temperature in CESM2(CAM6) is overall very similar to that of CESM1(CAM5) for a majority of the world’s land regions, with the only exceptions being regions of decreased skill over parts of northeast and southernmost Asia, and southernmost part of India for weeks 3–4 and 5–6. Figures 2g–i reveal that the DJF 2-m temperature ACC for CESM2(WACCM6) is also very similar to that of CESM2(CAM6), demonstrating that the whole atmospheric version of CESM2 does not fundamentally change the surface prediction skill of the model. There are a few land regions for which the DJF 2-m temperature ACC is statistically significantly different for CESM2(WACCM6) as compared to CESM2(CAM6), 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 systems.

Figure 3 shows ACC of 2-m temperature over land for the June–July–August (JJA) averages. 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 2-m temperature is a little smaller compared to that for DJF. The ACC values are the largest in northern South America and tropical Africa for weeks 3–4.
The differences between ACC in CESM2(CAM6) and CESM1(CAM5) are very small as shown in Figs. 3d–f. Similarly to what was shown in Fig. 2, there is some degradation in skill from CESM1 to CESM2, primarily in parts of Eurasia. This is likely due to changes in model physics between CESM1 and CESM2. Figures S1 and S2 in the online supplemental material show the 2-m temperature ACC for the March, April, May (MAM) and September–October–November (SON) averages, respectively. In MAM, CESM2(CAM6) shows a statistically significant degradation of 2-m 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 2-m temperature ACC for CESM2(CAM6) and CESM1(CAM5), as well as between CESM2(WACCM6).

Figure 4 compares the DJF and JJA 2-m temperature ACC averaged over all land areas and over North America. DJF ACC of 2-m 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 (Fig. 4a). DJF ACC of 2-m 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 2-m 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. Overall, the 2-m temperature prediction skill of all the CESM versions considered here is very similar to NOAA’s CFSv2 model for all weeks, however, consistently lower than the ECMWF forecasting system. The ECMWF model simulates initial uncertainties using singular vectors and ensemble of data assimilation, represents model uncertainties due to physical parameterizations using a stochastic scheme, and has much finer horizontal resolution (~16 km up to day 15, and ~32 km after that) as compared to CESM (~100-km resolution).
These factors are likely the main causes of the differences between ECMWF and CESM subseasonal skill.

There are overall small differences in 2-m temperature ACCs between the various CESM versions considered, as well as between ACCs calculated for an ensemble size of 5 versus 10 for CESM1(CAM5) and CESM2(CAM6) for both DJF and JJA. Such differences, however, as also depicted by the individual bars in Fig. 4, are not statistically significant based on the Fisher z transform analysis. A caveat here is that this significance test depends on the sample size. For a typical season, there are ∼220 time samples from the CESM reforecasts. We investigated how big of a time sample size would be needed for the ACC differences to be statistically significant between CESM2(CAM6) and other versions of the model, and this number would need to be ∼800. This indicates that a hindcast record of length that is 4 times larger than current (∼80 years) would be needed in order for the shown differences to be statistically significant, which is currently not plausible due to limitations in computing and also initial data availability going that far back in time.

Figures 5a–c and 6a–c show the ACC for precipitation for DJF and JJA, respectively, for CESM2(CAM6). Precipitation prediction skill at subseasonal time scales (Figs. 5b,c) is quite low as compared to the 2-m 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.)
Similar to 2-m 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 (Fig. 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 (Fig. 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 usually 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 time scales, except for very few selected regions discussed above. As in 2-m temperature, prediction skill of precipitation of the CESM model versions is very similar to the NOAA CFSv2 and lower than that of ECMWF model.

b. Spread and error characteristics for 2-m 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 (Fig. 8). Unlike the ACC which evaluates pattern correlation of predicted and observed anomalies, RMSE quantifies the amplitude of error in predicted anomalies. To minimize the effects of systematic biases, the lead-dependent climatology was removed prior to calculating the RMSE. Similar 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. Figures 4a and 4b show an increase in RMSE in DJF from weeks 1–2 to weeks 3–4 but do not show a rapid decrease in skill between weeks 3–4 and weeks 5–6, implying little loss in capturing anomaly magnitudes. In JJA, the RMSE magnitudes are even more consistent across the forecast period.

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 Fig. 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 largely evident in our reforecasts. In weeks 1–2, regardless of the season, or land area average, the spread is underdispersive by at least 40% (Fig. 8). The underdispersion improves for longer lead times but forecasts remain markedly overconfident for all experiments. The spread differences between the CESM versions are small for JJA, but for DJF, CESM1(CAM5) creates consistently more spread than CESM2(CAM6) or CESM2(WACCM) over North America.

c. MJO and NAO prediction skill

The MJO and the NAO are key drivers of extreme weather on subseasonal time scales and believed to be key sources of subseasonal predictability. To evaluate the MJO prediction skill, the real-time multivariate MJO (RMM; Wheeler and Hendon 2004) index is calculated with the 200- 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 and Smith 1996). Predicted RMM indices are calculated by projecting the forecast anomalies for those fields onto the associated observed empirical orthogonal functions (EOF) eigenvectors (Kim et al. 2018). Then, the RMM index bivariate ACCs are computed between the predicted and observed 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) and for CFSv2 (details of the dataset are in Kim et al. 2014), but less than the MJO predictability of out to 33 days in the ECMWF-CY43R system (details of the datasets are in 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 is 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) affects 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) and seasonal time scales (e.g., Riddle et al. 2013; Scaife et al. 2014); however, its predictability on subseasonal time scales 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

![Fig. 6. As in Fig. 3, but for precipitation.](image)

Unauthenticated | Downloaded 05/28/22 02:42 AM UTC
models participating in the SubX project. 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 of ensemble size to 20 in CESM1(CAM5) enhances the NAO prediction skill to 0.51 for weeks 3–6 in boreal winter. The prediction skill of the NAO in the various CESM versions is shown in Fig. 10. The NAO index was obtained by first performing an EOF analysis of ERA-Interim (Dee et al. 2011) monthly (NDJFM) sea level pressure (SLP) anomalies over the Atlantic sector (20°–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 weeks 3–4 NAO ACC is above or close to 0.5 for all the CESM versions considered here, similar to the skill in NOAA’s CFSv2 reforecasts and to the ECMWF model for CESM1 (a little lower for other CESM versions). ACC of CESM1(CAM5) based on a 10-member ensemble and CESM2(WACC6) based on a 5-member ensemble have the highest skill at weeks 3–4; however, with the current reforecast sample sizes, 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 (WACC6) is 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.

d. 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 and Butler 2020; Baldwin et al. 2021). This coupling between tropospheric weather and SSWs is often summarized by the time evolution of the annular modes (Baldwin and Dunkerton 2001), or nearly equivalently, the standardized polar cap geopotential anomalies (Fig. 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 (Fig. 11a) and in CESM2(WACC6) 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 re-forecasted 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°N and 10 hPa becomes negative, with 14 SSW events in the reforecast period. The central dates of the observed events are as follows: 1) 26 February 1999, 2) 11 February 2001, 3) 30 December 2001, 4) 17 February 2002, 5) 18 January 2003, 6) 5 January 2004, 7) 21 January 2006, 8) 24 February 2007, 9) 22 February 2008, 10) 24 January 2009, 11) 9 February 2010, 12) 6 January 2013, 13) 12 February 2018, and 14) 2 January 2019. Figure 11 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. They also do not extend all the way down to the tropopause as they do in MERRA-2. However, the positive surface geopotential anomalies linger for 4 weeks after the central date of an 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 0.8. In contrast, the averages of the individual reforecast pattern correlations with their respective SSW events in MERRA-2 are substantially lower with 0.4 for CESM2(WACCM6) and 0.2 for CESM2(CAM6). We expect the average of the individual correlation coefficients to be lower than the ensemble/event mean correlation coefficient as the geopotential of individual forecasts will deviate from observations, while the ensemble and event mean will resemble observations. The lead times at which SSWs are predicted in CESM2(WACCM6) are longer than in CESM2(CAM6): 11 days on average versus 8 days, respectively. In summary, CESM2(WACCM6) reforecasts of the polar cap geopotential anomalies following an SSW indicate that they are somewhat more consistent with those of MERRA-2 than in CESM2(CAM6) reforecasts. Together with higher SSW lead times we can expect better surface prediction from CESM2(WACCM6) associated with these events; however, a detailed analysis of the surface impacts is beyond the scope of this overview manuscript.

e. Mesosphere and lower thermosphere prediction

Initial investigations by Wang et al. (2014) and Pedatella et al. (2018a) 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 an 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 70° and 90°N from the 14 major SSW events from CESM2(WACCM6) in comparison with the SSWs from WACCM Specified Dynamics simulations with thermosphere-ionosphere xTension (WACCMX-SD; Liu et al. 2018) for verification. 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 Fig. 12 are based on compositing the reforecasts regardless of whether they successfully forecast an SSW, and 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 an SSW) are considered.

Several distinct features of the middle atmosphere (stratosphere: ∼100–0.5 hPa or ∼10–50 km; mesosphere: 0.5 to 10−3 hPa or ∼50–90 km; lower thermosphere: above 10−3 hPa) respond to SSWs can be seen in Fig. 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 (Fig. 12d) and near the SSW onset (Fig. 12e) capture the mesosphere cooling and elevated stratopause with higher fidelity when compared 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 Fig. 12 thus suggest that it may be possible to forecast the upper atmosphere variability ∼10 days in advance of an SSW.

f. Limitations of the current system for chemistry prediction

As CESM2(WACCM6) includes a comprehensive tropospheric and middle atmospheric chemistry module, we were hopeful that the current CESM2(WACCM6) subseasonal prediction system 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 time scale introduces significant deviations between modeled and observed water vapor. This is illustrated in Fig. 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 and 70 hPa time series (Fig. 13b).

We have performed several sensitivity experiments with WACCM6-SD, including an experiment in which we lowered the nudging top from 60 to 50 km and another experiment in which we increased the nudging time scale from 1 to 2 h. 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-h 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. Another source of error may be the meridional circulation, which is difficult to constrain through specified dynamics schemes (Orbe et al. 2020). While short nudging time scales may decrease the errors in the variability, they may do so at the expense of strengthening the residual 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 time scale in the stratosphere may improve the representation of stratospheric chemistry in the S2S reforecast and forecasts with CESM2(WACCM6) which will be explored in the future further.

4. Summary and conclusions

We have described here fully coupled Earth system subseasonal prediction systems based on 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. The two CESM2-based subseasonal prediction systems differ in initialization procedures for the atmosphere and the ocean, which largely follow previous approaches for decadal prediction simulations with CESM1. Subseasonal reforecasts were carried out following the SubX protocol for years 1999–2020 with weekly start dates for each year for CESM2(CAM6), 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(CAM6).

We demonstrated that the prediction skill of 2-m temperature and precipitation as well as of the NAO in the CESM2-based subseasonal prediction systems are comparable to the prediction skill for these variables in CESM1 and similar to the skill of NOAA’s CFSv2 model, and a little lower than that the skill of the ECMWF model. The MJO in CESM2 is predictable out to ~25 days—a few days longer than in the NOAA CFSv2 system, but less than that in the ECMWF system. The high subseasonal prediction skill of the CESM2 subseasonal systems, along with extensive output obtained for all model components, makes it an excellent tool for studies of subseasonal predictability. We further demonstrated that stratospheric–tropospheric coupling during SSW events is well represented in CESM2(CAM6) and CESM2(WACCM6), which implies that both prediction systems 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 also found 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 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 2-m temperature skill in eastern Asia in CESM2(WACCM6) as compared to CESM2(CAM6), and higher 2-m temperature skill in parts of North America in CESM2(CAM6), both for weeks 5–6. From the current design of the S2S systems it is difficult to isolate where exactly these differences in skill come from. Similarly, we cannot isolate whether the small differences in S2S prediction skill of surface variables between CESM1 and CESM2 come from the numerous changes in model physics or differences in changing atmospheric initial conditions from ERA-Interim to CFSv2 reanalysis. Nevertheless, from the similarity of forecast patterns and similarity of spatial distributions of ACC for surface variables for all these systems, it seems that small differences in initial conditions make little difference to the skill of subseasonal forecasts.

Stratospheric–tropospheric coupling is well represented in both CESM2-based subseasonal prediction systems; however, the polar cap geopotential anomalies following an SSW are more consistent with observations in CESM2(WACCM6) as compared to CESM2(CAM6). The impact of this difference on predictability of surface extreme weather associated with SSWs will be investigated in future work.

The CESM2-based subseasonal prediction systems do not utilize data assimilation as many other modern operational systems. Despite this, the prediction skill of these systems is quite good, but it is possible that additional skill and model shock reduction could be achieved with data assimilation. CESM2(CAM6) and CESM2(WACCM6) are freely available for use of the community. The reforecast sets and forecasts described here are also publicly available and are designed to serve as a foundation for future experiments elucidating sources of subseasonal predictability. The near-real-time forecasts are freely available and contribute to the information used by NOAA to issue weeks 3–4 outlook. The extensive output from the atmospheric, land, ocean, and sea ice components may open new avenues of research.

Acknowledgments. 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 and 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.

Data availability statement. MERRA-2 data are available from NASA’s Global Modeling and Assimilation office at https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. CPC Global Temperature data are provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, 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 DOIs: https://doi.org/10.5065/063-m767 and https://doi.org/10.5065/ekns-e430, respectively. Downloadable information for real-time forecasts for both systems
can be found on https://www.cesm.ucar.edu/working-groups/earth-system-prediction/simulations/subseasonal-reforecasts-forecasts/.

REFERENCES


Lawrence, D. M., and Coauthors, 2019: The Community Land Model version 5: Description of new features, benchmarking,


—, K. Lindsay, S. C. Doney, M. C. Long, and K. Misumi, 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. *J. Climate*, **26**, 9291–9312, https://doi.org/10.1175/JCLI-D-12-00566.1.


Vitart, F., and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. npj Climate Atmos. Sci., 1, 3, https://doi.org/10.1038/s41612-018-0013-0.


