

Evaluation of Processes Related to Stratosphere–Troposphere Coupling in GEFSv12 Subseasonal Hindcasts

ZACHARY D. LAWRENCE^{a,b}, DILLON ELSTURBY^{a,c}, AMY H. BUTLER^c, JUDITH PERLWITZ^b, JOHN R. ALBERS^{a,b},
LAURA M. CIASTO^d, AND ERIC RAY^{a,c}

^a Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado

^b NOAA/Physical Sciences Laboratory, Boulder, Colorado

^c NOAA/Chemical Sciences Laboratory, Boulder, Colorado

^d National Center for Environmental Prediction, Climate Prediction Center, College Park, Maryland

(Manuscript received 14 October 2022, in final form 4 April 2023, accepted 5 April 2023)

ABSTRACT: The representation of the stratosphere and stratosphere–troposphere coupling processes is evaluated in the subseasonal Global Ensemble Forecast System, version 12 (GEFSv12), hindcasts. The GEFSv12 hindcasts develop systematic stratospheric biases with increasing lead time, including a too strong boreal wintertime stratospheric polar vortex. In the tropical stratosphere, the GEFSv12 winds and temperatures associated with the quasi-biennial oscillation (QBO) tend to decay with lead time such that they underestimate the observed amplitudes; consistently, the QBO-associated mean meridional circulation is too weak. The hindcasts predict extreme polar vortex events (including sudden stratospheric warmings and vortex intensifications) about 13–14 days in advance, and extreme lower-stratospheric eddy heat flux events about 6–10 days in advance. However, GEFSv12’s ability to predict these events is likely affected by its zonal-mean circulation biases, which increases the rates of false alarms and missed detections. Nevertheless, GEFSv12 shows stratosphere–troposphere coupling relationships that agree well with reanalysis and other subseasonal forecast systems. For instance, GEFSv12 reproduces reanalysis relationships between polar vortex strength and the Northern Annular Mode in the troposphere. It also exhibits enhanced weeks 3–5 prediction skill of the North Atlantic Oscillation index when initialized during strong and weak polar vortex states compared to neutral states. Furthermore, GEFSv12 shows significant differences in Madden–Julian oscillation (MJO) amplitudes and enhanced MJO predictive skill in week 4 during easterly versus westerly QBO phases, though these results are sensitive to the level used to define the QBO. Our results provide a baseline from which future GEFS updates may be measured.

KEYWORDS: Stratospheric circulation; Stratosphere–troposphere coupling; Climate prediction; Ensembles; Forecast verification/skill

1. Introduction

A growing body of research has established that the stratosphere can influence tropospheric weather and climate on subseasonal-to-seasonal (S2S) time scales (see Butler et al. 2019; Domeisen and Butler 2020, for reviews). These relationships arise as a result of two-way coupling across atmospheric layers involving interactions between planetary-scale atmospheric waves and the mean circulation. The stratospheric circulation features that may subsequently influence the tropospheric state generally exhibit high persistence and predictability, at least relative to the troposphere (Domeisen et al. 2020a; Son et al. 2020). Therefore, the stratosphere is recognized as an important potential source of predictive skill for S2S forecasts (Merryfield et al. 2020; Domeisen et al. 2020b; Scaife et al. 2022), including as a driver of “forecast windows of opportunity” (Mariotti et al. 2020; Albers and Newman 2021).

The ability to leverage the stratosphere as a source of predictability may be limited by stratospheric biases and deficiencies that are relatively common in global models (Charlton-Perez et al. 2013; Richter et al. 2020; Rao et al. 2022; Serva et al. 2022), including in S2S and seasonal prediction systems (Lawrence et al. 2022; Rao et al. 2023). For example, many S2S systems exhibit biases in two key circulation features of the stratosphere, the stratospheric polar vortices (Vaugh and Polvani 2010; Vaugh et al. 2017), and the quasi-biennial oscillation (QBO) (Baldwin et al. 2001). There can be many reasons for such model errors (see Lawrence et al. 2022, for a more detailed discussion with a focus on S2S models), but generally they are tied to deficiencies in resolved and/or parameterized wave drag (McLandress et al. 2012; Garfinkel et al. 2022; Schwartz et al. 2022), low model lid heights and/or vertical resolution in the stratosphere (Charlton-Perez et al. 2013; Richter et al. 2014; Shaw et al. 2014), radiative effects from trace gases such as water vapor and ozone (Hogan et al. 2017; Bland et al. 2021), and even model numerics (Polichtchouk et al. 2019, 2021). Some modeling centers such as the European Centre for Medium-Range Weather Forecasts (ECMWF) have prioritized improving the representation of the stratosphere in their models because of the potential benefits for tropospheric prediction (Shepherd et al. 2018; Polichtchouk et al. 2021).

The National Oceanic and Atmospheric Administration (NOAA) Global Ensemble Forecast System version 12

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/MWR-D-22-0283.s1>.

Corresponding author: Z. D. Lawrence, zachary.lawrence@noaa.gov

DOI: 10.1175/MWR-D-22-0283.1

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

(GEFSv12) was released in 2020, five years after its predecessor. Major updates implemented in GEFSv12 include a change to the dynamical core from the previous hydrostatic Global Spectral Model to a finite volume-based (FV3) nonhydrostatic model (Lin 2004; Putman and Lin 2007), new and updated physics schemes, an increase in operational ensemble size from 20 to 30 members, and the extension of forecasts from 16 to 35 days (Zhou et al. 2022). The release of GEFSv12 also included a 30-yr set of GEFSv12 hindcasts from 1989 to 2019 (Guan et al. 2022), and a 20-yr GEFSv12 reanalysis that was used to initialize the hindcasts from 2000 to 2019 (Hamill et al. 2022). With these characteristics, GEFSv12 represents NOAA's latest operational subseasonal prediction system since the release of the Climate Forecast System version 2 (CFSv2) in 2011 (Saha et al. 2014).

Generally, the representation of the stratosphere and stratosphere–troposphere coupling processes has not been methodically assessed and documented for NOAA S2S prediction systems. The CFSv2 has seen widespread use in studies focused on aspects of stratospheric predictability and stratosphere–troposphere coupling (see, e.g., Zhang et al. 2013; Riddle et al. 2013; Yu et al. 2019; Miller and Wang 2019; Rao and Garfinkel 2020), including in model intercomparison studies such as those by Domeisen et al. (2020a,b) and Lawrence et al. (2022). Since the GEFSv12 is relatively newer, it has not yet seen similar use. Lawrence et al. (2022) showed that the “NOAA family” of systems including the GEFSv12 and CFSv2 show some similar and different stratospheric biases; for instance, both systems exhibit warm global and annual mean stratospheric temperature biases, but the GEFSv12 exhibits a strong Arctic polar vortex bias in boreal winter, while the CFSv2 exhibits a weak vortex bias instead.

The community-based Unified Forecast System (UFS) is intended to be the source system for NOAA operational forecasting models going forward, including for S2S applications (Cikanek et al. 2019; Jacobs 2021). In addition to being the most major update to NOAA's S2S forecasting capabilities in roughly a decade, the GEFSv12 is considered to be the first UFS-based application for NOAA S2S prediction. Thus, there is a significant need to evaluate processes relevant for S2S predictability such as stratosphere–troposphere coupling in the GEFSv12. Such evaluations can inform community-driven UFS development and provide benchmarks for future iterations of the GEFS. A better understanding of the interplay between stratospheric biases, coupling processes, and predictive skill in one S2S system such as GEFSv12 may also provide insight into the behavior and performance of other models.

Herein we provide a thorough evaluation of relevant stratospheric metrics and stratosphere–troposphere coupling processes in the GEFSv12 hindcasts. In section 2 we briefly describe the GEFSv12 hindcasts, the reanalysis data we use for comparison, and our methods. In section 3 we present our results, focusing on the zonal-mean state, extreme stratospheric events, and aspects of stratosphere–troposphere coupling. Finally, in section 4 we summarize and discuss our results.

2. Data and methods

a. GEFSv12 hindcasts

We evaluate the 35-day “phase 2” GEFSv12 hindcasts which cover the period from 2000 to 2019. These hindcasts were run once weekly (beginning 5 January 2000) with 11 members, for a total of 1043 initializations. The GEFSv12 hindcasts were output with higher temporal and spatial resolutions for earlier time steps and pressure levels. For our purposes, we thin all data to match the output grid for leads beyond 10 days, which is 0.5° in the horizontal with 6-hourly time steps. Pressure level fields are provided on 25 standard levels between 1000 and 1 hPa, with 10 between 100 and 1 hPa. Guan et al. (2022) provide further details on the GEFSv12 hindcast configuration, data archive, and performance.

b. GEFSv12 reanalysis (GEFS-R)

The GEFSv12 reanalysis (hereinafter, GEFS-R) was created to initialize the phase 2 GEFSv12 hindcasts. It was developed so the assimilation system and initial conditions for the hindcasts would be as consistent as possible with the operational system of the real-time GEFSv12 forecasts. This was done to minimize differences in short-term forecast biases that may otherwise arise from differing characteristics of the systems used to generate analyses for the hindcasts versus operational forecasts. Hamill et al. (2022) provide a more comprehensive description of GEFS-R and the motivations behind it.

Herein, we compare GEFS-R to another more commonly used reanalysis (ERA5, described in the next subsection) to better understand how biases in the GEFSv12 hindcasts may be tied to its initial conditions versus drift in the underlying model, including in uncertain regions such as the upper stratosphere. We also use GEFS-R for the period of 2000–19 as a comparison to ERA5's longer reanalysis period dating back to 1979 to demonstrate the sensitivity of stratosphere–troposphere coupling statistical relationships to the period of years considered.

The GEFS-R atmospheric fields are provided on a $0.25^\circ \times 0.25^\circ$ grid with 6-hourly frequency (0000, 0600, 1200, and 1800 UTC times). Pressure level fields are given on 31 standard levels between 1000 and 1 hPa, with 11 between 100 and 1 hPa.

c. ERA5 reanalysis

We use ERA5 (Hersbach et al. 2020) for direct comparisons and verification of the GEFSv12 hindcasts with reanalysis. For the 2000–06 period, we use the revised ERA5.1 data that includes reduced biases in stratospheric temperatures and humidity (see Simmons et al. 2020, for more details). We use ERA5 atmospheric fields on a $0.25^\circ \times 0.25^\circ$ grid with 3-hourly temporal resolution. Pressure level data are provided on 37 standard levels between 1000 and 1 hPa, with 11 levels between 100 and 1 hPa.

d. Data preprocessing

1) ZONAL-MEAN DATASETS

Variations in the stratospheric circulation can largely be understood through wave–mean flow interactions (Andrews et al. 1987).

For our study, we created a publicly available archive of GEFSv12 hindcast “zonal-mean datasets.” These include zonal means of wind components (u , v , ω), temperatures (T), geopotential heights, and eddy fluxes. Here, “eddy fluxes” refer to the meridional and vertical fluxes of heat and momentum, such as the eddy heat flux $\overline{v'T'}$, where \bar{x} refers to the zonal mean of field x , and $x' = x - \bar{x}$. These eddy fluxes are additionally computed for zonal wavenumbers 1–3.

2) REGIONAL AVERAGING AND ANOMALIES

Throughout the paper, we use diagnostics averaged over different regions or latitude bands. In such cases, we use cosine-latitude weighted averages to account for the area of grid cells.

Similarly, whenever anomalies are needed from the hindcasts, they are computed by removing a lead-dependent climatology. Because the GEFSv12 hindcasts were initialized once per week, they have nonuniform sampling across days of year in the 2000–19 period. Therefore, we smooth the lead-dependent climatologies according to the “SubX method” described in [Pegion et al. \(2019\)](#), see their appendix B). This entails applying a 31-day rolling triangular weighted average across the days of year of the raw climatologies.

3) WEATHER AND CLIMATE INDICES

Section 3 includes results that quantify the ability of GEFSv12 to predict aspects of weather and climate variability, including polar vortex geometry ([Vaughn 1997](#); [Seviour et al. 2013](#)), the North Atlantic Oscillation (NAO; [Hurrell et al. 2001](#)), Northern Annular Mode (NAM; [Thompson and Wallace 2000](#)), and the Madden–Julian oscillation (MJO; [Wheeler and Hendon 2004](#)). The full details for how we compute the indices that describe these phenomena from reanalysis and hindcast data are included in the [appendix](#).

e. Forecast evaluation

When referring to forecast lead times, we use the convention that weeks 1, 2, ..., refer to aggregates over days 1–7, 8–14, ..., etc., and similarly for pentads. In cases where we compare values with reanalysis, we equivalently sample the reanalysis data to match the valid times in the GEFSv12 hindcasts. For instance, a “week 3–5” aggregate of reanalysis data would correspond to using the GEFSv12 day 15–35 valid times within a specified season (e.g., only DJFM valid times), or from a set of initializations (e.g., only NDJF initializations) to sample the reanalysis.

1) EXTREME STRATOSPHERIC EVENTS

Our analysis considers GEFSv12’s ability to predict different extreme stratospheric events. We define such events using thresholds so that the target observations and forecasts are dichotomous, corresponding to whether the given threshold is met at a given lead time. For aggregate leads (pentads or weeks), we mark the entire period with “1” if the given threshold is met for any day in the bin (and “0” otherwise). In this way, we judge the performance of the hindcasts at predicting the presence of extreme stratospheric conditions

without requiring ensemble members to align perfectly in time for pentad or weekly time scales, in a manner appropriate for S2S predictions (so-called “fuzzy verification” [Ebert 2008](#); [Yu et al. 2019](#)). We then compile statistics about the hit (forecast and observations = 1), miss (forecast = 0, observations = 1), and false alarm (forecast = 1, observations = 0) rates. With these, we show the lead time at which the number of hits equals the sum of misses and false alarms, attributes diagrams, and receiver operator characteristic (ROC) curves ([Hsu and Murphy 1986](#); [Wilks 2019](#)).

We focus on four extreme stratospheric events: SSWs ([Baldwin et al. 2021](#)), vortex intensifications (VIs; [Limpasuvan et al. 2005](#); [Tripathi et al. 2015](#)), and extreme positive and negative heat flux events ([Shaw et al. 2014](#); [Dunn-Sigouin and Shaw 2015](#)). We define SSWs using the common threshold for 10-hPa 60°N zonal-mean zonal winds below 0 m s^{-1} during the NDJFM period. We then use the percentile of 0 m s^{-1} for SSWs in the 2000–19 NDJFM reanalysis record (~ 7.5 percentile) to define the equivalently rare extreme (92.5 percentile) to define VIs, which corresponds to winds above 49 m s^{-1} . We use the same percentiles to define the extreme heat flux events from 50 hPa, 45°–75°N wave-1 $\overline{v'T'}$ in the NDJFM reanalysis record, which gives thresholds of -2 and 38 K m s^{-1} . We only focus on wave-1 eddy heat fluxes, which generally dominate the total eddy heat flux and the occurrence of negative heat flux events ([Shaw and Perlwitz 2013](#); [Dunn-Sigouin and Shaw 2015](#)).

2) SKILL METRICS

We make use of common correlation metrics to examine the effect of background stratospheric conditions on GEFSv12 hindcasts of the NAO and MJO indices. The correlation skill score (CSS; [Charlton-Perez et al. 2021](#)) is

$$\text{CSS} = \frac{\sum_{i=1}^N \bar{x}(t_i, \tau) y(t_i + \tau)}{\sqrt{\sum_{i=1}^N \bar{x}^2(t_i, \tau) \sum_{i=1}^N y^2(t_i + \tau)}}, \quad (1)$$

where $\bar{x}(t_i, \tau)$ is an ensemble mean forecast for initialization t_i at lead time τ , $y(t_i + \tau)$ is the corresponding observation for the same valid time period, and N is the number of initializations. Here it is assumed that appropriate climatologies have been removed such that x and y are anomalies, or are representative of anomalies in the case of EOF-based indices. The subset correlation (SSC; [Charlton-Perez et al. 2021](#)) is

$$\text{SSC} = \frac{\sum_{i=1}^N \{\bar{x}(t_i, \tau) - [\bar{x}]\} \{y(t_i + \tau) - [y]\}}{\sqrt{\sum_{i=1}^N \{\bar{x}(t_i, \tau) - [x]\}^2 \sum_{i=1}^N \{y(t_i + \tau) - [y]\}^2}} \quad (2)$$

where $[\bar{x}]$ and $[y]$ are the means of the forecast and observation anomalies within the subset of N initializations, which are functions of τ only. Finally, for bivariate variables such as the

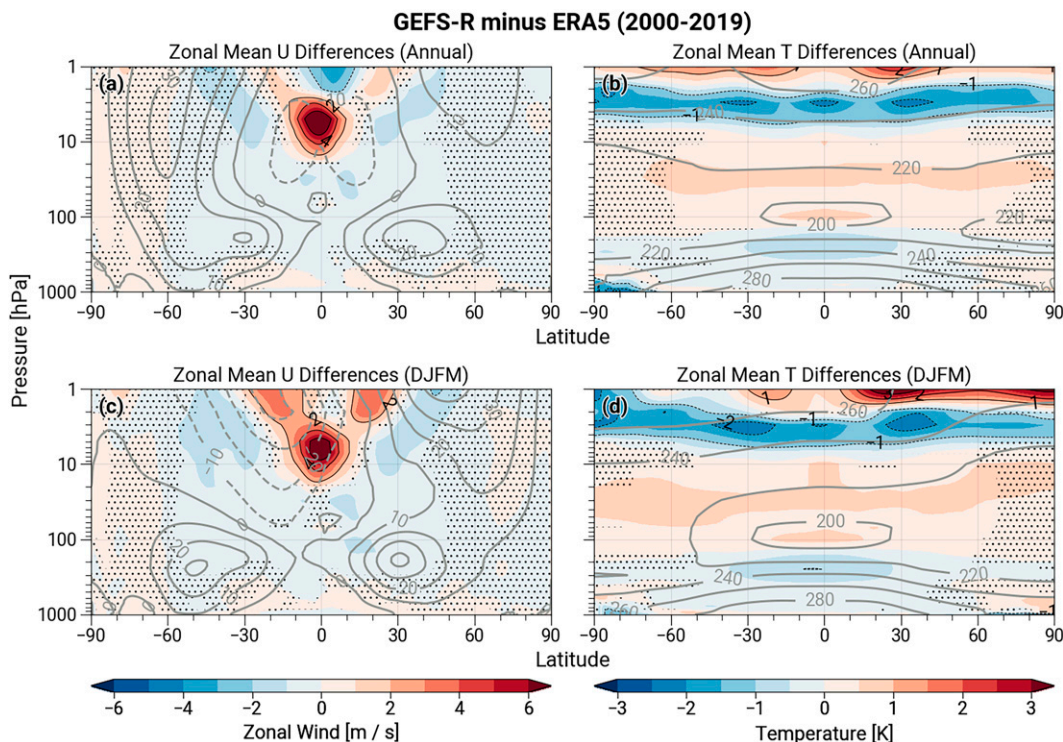


FIG. 1. (a),(b) Annual and (c),(d) December–March differences between GEFS-R and ERA5 (2000–19) for (a),(c) zonal-mean zonal winds (m s^{-1}) and (b),(d) temperatures (K). Additionally, differences between GEFS-R and ERA5 are contoured with thin black/dashed lines every 2 m s^{-1} in (a) and (c), and every 1 K in (b) and (d). The ERA5 climatologies are shown in the thick gray contours, with winds contoured every 10 m s^{-1} , and temperatures every 20 K. Stippling shows where the reanalysis differences are not statistically significant at the 95% level as determined from bootstrap resampling 50 000 times.

real-time multivariate MJO indices, the bivariate anomaly correlation (BCOR; Rashid et al. 2011) is

$$\text{BCOR} = \frac{\sum_{i=1}^N [\bar{x}_1(t_i, \tau) y_1(t_i + \tau) + \bar{x}_2(t_i, \tau) y_2(t_i + \tau)]}{\sqrt{\sum_{i=1}^N [\bar{x}_1^2(t_i, \tau) + \bar{x}_2^2(t_i, \tau)] \sum_{i=1}^N [y_1^2(t_i + \tau) + y_2^2(t_i + \tau)]}} \quad (3)$$

where the subscripts denote the two components of the ensemble mean forecasts \bar{x} and observations y as in Eq. (1).

3. Results

a. Biases in the zonal mean

The stratospheric circulation can largely be explained by the balance of radiatively and dynamically induced temperature variations, the latter of which can be well-understood within the framework of waves interacting with the zonal-mean flow (Andrews et al. 1987). Therefore, model biases in stratospheric zonal-mean temperatures and winds can be directly tied to the model's representation of radiative and dynamical processes. In forecasts, biases may be related to the conditions used to initialize the forecast model (versus the

observations used to verify), and/or they may develop with lead time when model-internal processes cause a systematic drift in modeled fields. Below we assess both of these aspects for GEFSv12 zonal-mean zonal winds and temperatures.

First, we determine how using ERA5 for verifying the GEFSv12 hindcasts may be influenced by differences in the initial stratospheric states from GEFS-R. Figure 1 shows the 2000–19 annual (Figs. 1a,b) and DJFM (Figs. 1c,d) average differences between GEFS-R and ERA5 zonal-mean winds and temperatures. Differences in the zonal-mean zonal winds poleward of about 30° in both hemispheres are generally small, less than 1 m s^{-1} in magnitude throughout the troposphere and stratosphere (Figs. 1a,c). The largest wind differences occur in the tropical middle-to-upper stratosphere, with GEFS-R winds being more westerly than in ERA5 by $2\text{--}6 \text{ m s}^{-1}$ between about 10–3 hPa, and from -10° to 10° latitude in the region of climatological easterlies. These westerly differences extend poleward and upward to about 1 hPa and $\pm 20^\circ$ in latitude, which are stronger in the DJFM-only composite. The temperature differences between GEFS-R and ERA5 do not have a strong seasonal dependence, but they are globally layered, with GEFS-R being warmer between 100 and 10 hPa, colder from 5 to 2 hPa, and warmer above 2 hPa (Figs. 1b,d). The differences are relatively small between 100 and 10 hPa, with GEFS-R warmer by less than 1 K;

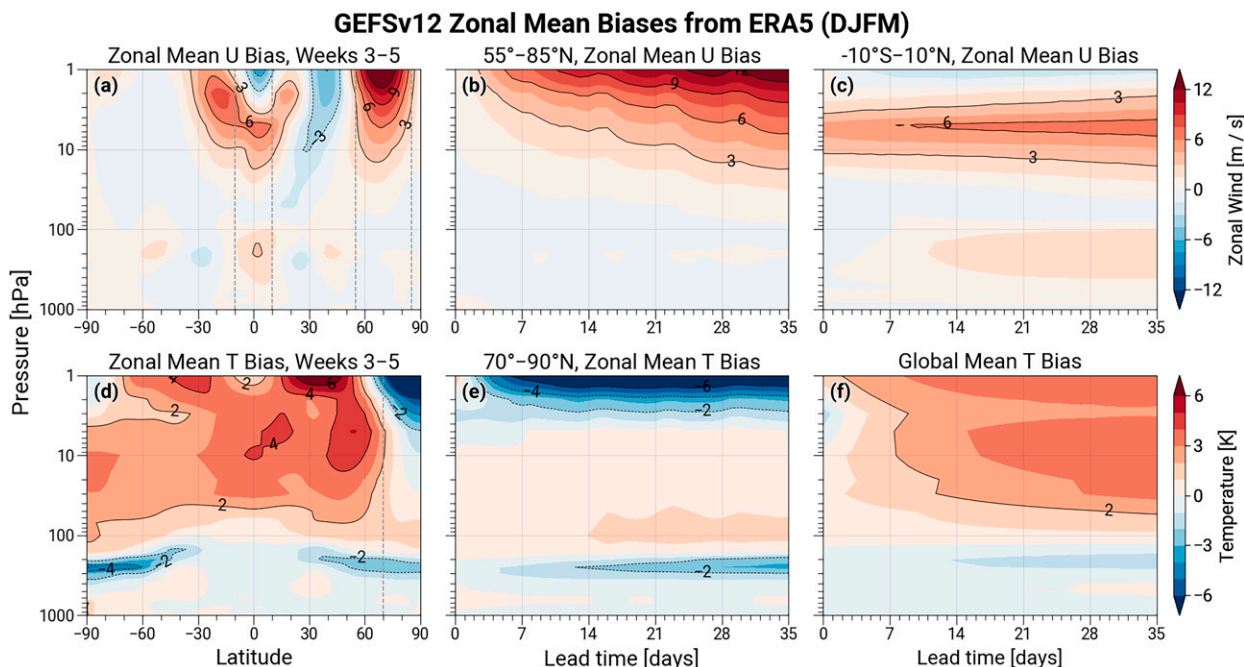


FIG. 2. GEFSv12 biases in zonal-mean (top) zonal winds (m s^{-1}) and (bottom) temperatures (K) relative to ERA5 reanalysis for DJFM valid times. (a),(d) The biases averaged over week 3–5 lead times as a function of pressure and latitude, and (b),(c),(e),(f) biases as a function of pressure and lead time. Panels (b) and (e) highlight zonal wind and temperature biases in the latitudes of the Arctic polar vortex while (c) highlights biases in the tropical zonal winds (10°S – 10°N); (f) shows global mean temperature biases.

however, above this region the temperature differences have higher magnitudes that exceed 1–2 K. These reanalysis differences remain similar for composites of different years, but the temperature differences appear to be more amplified from 2000 to 2009 (not shown), possibly suggesting an effect related to assimilated observations.

The GEFS-R versus ERA5 reanalysis differences are consistent with differences seen between other modern reanalysis datasets. The regions with the largest GEFS-R minus ERA5 differences are in the middle-to-upper stratosphere, which is where reanalysis differences tend to be largest in general due to the relatively sparse observations available for assimilation (Long et al. 2017; Fujiwara et al. 2022). However, the tropical westerly wind differences from ~ 10 - to 3-hPa apparent in Figs. 1a and 1c appear to be unique to GEFS-R, as they also appear when comparing GEFS-R to other modern reanalyses (not shown). ERA5 temperatures have been shown to agree well in the middle-to-upper stratosphere up to 1 hPa relative to Rayleigh lidar measurements (Marlton et al. 2021). These results suggest that some of the GEFS-R differences may be biases inherent to GEFS-R, which will be in the GEFSv12 hindcast initial conditions.

With these reanalysis differences in mind, we now view zonal-mean biases in the GEFSv12 hindcasts to better understand the behavior of the prediction system. Figure 2 shows the DJFM weeks 3–5 biases relative to ERA5 in zonal-mean zonal winds and temperatures (Figs. 2a,d). We show DJFM only since the biases are similar for the Southern Hemisphere during its winter season and in the annual mean (as shown in

Lawrence et al. 2022). Notable features of the biases include: a westerly polar night jet/cold polar cap bias; tropical middle-to-upper stratosphere wind biases; extratropical upper tropospheric/lower stratospheric (UTLS) cold biases in both hemispheres; and a global mean stratospheric warm bias. The westerly polar night jet bias becomes apparent at the highest levels within the first week of the forecasts (Fig. 2b). Due to the competing effect of the global warm bias encroaching into the polar region, the cold pole bias is only apparent down to 2 hPa in the 70° – 90°N average (Fig. 2e). However, the apparent negative equator–pole gradient in the temperature biases in Fig. 2d agrees with the deeper extent of the westerly polar vortex wind biases, which exceed 3 m s^{-1} down to 10 hPa by day 24–25 of the hindcasts. The westerly bias in the tropical middle-to-upper stratospheric winds is an example of a bias present in the initial conditions (Fig. 1c) that is maintained and even amplified throughout the forecasts (Fig. 2c). The extratropical UTLS cold bias is present in both hemispheres and is more amplified in the summer hemisphere (Fig. 2d); however, even in the winter hemisphere this bias becomes apparent within the first week, and amplifies further with cold biases below -2 K becoming predominant by about day 14 (Fig. 2e). In the global mean, the layered nature of the GEFS-R minus ERA5 temperature differences (Figs. 1b,d) is apparent at short lead times, but quickly becomes overwhelmed by a monotonic increase in the temperature biases throughout the stratosphere, with biases exceeding 2 K down to 50 hPa by around day 14 (Fig. 2f).

Many of the biases shown in Fig. 2 were discussed in Lawrence et al. (2022) in the context of other S2S forecast systems.

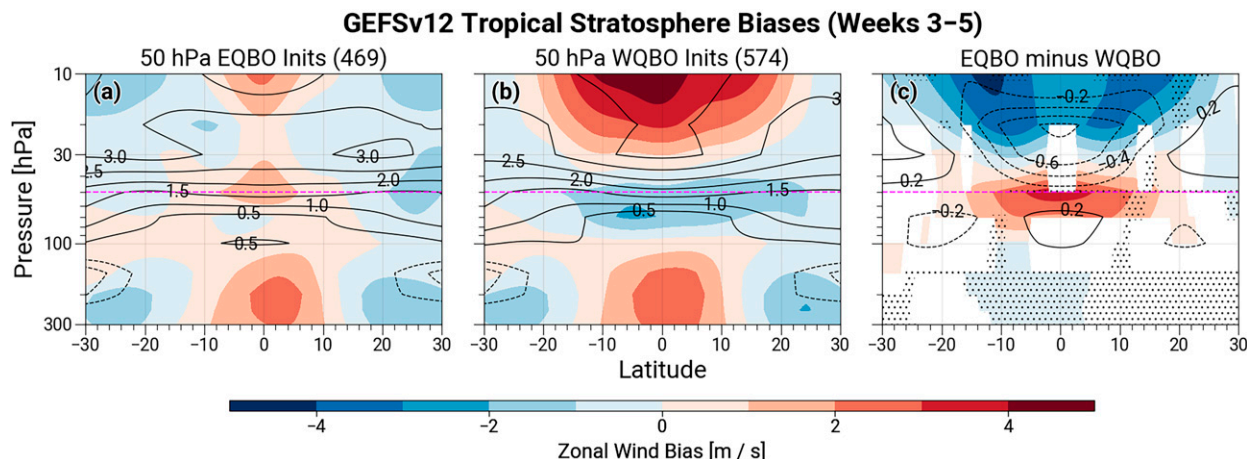


FIG. 3. GEFSv12 tropical zonal-mean zonal wind (m s^{-1}) and temperature biases (K) averaged over weeks 3–5 for all initializations in (a) EQBO, (b) WQBO, and (c) their difference. The phase of the QBO is defined by the 50-hPa zonal-mean zonal winds averaged from 10°S to 10°N . Zonal wind biases are plotted in the colors, whereas temperature biases are plotted in the line contours. In (c), wind bias differences are plotted in color only where differences are significantly different from 0; stippling shows where temperature differences are not significantly different from 0. Statistical significance is assessed at the 95% level via bootstrap resampling 50 000 times. The 50-hPa level is indicated by the magenta dashed line.

However, Fig. 2 additionally shows levels above 10 hPa that are generally not available for other S2S systems, which provides further insight into the structure of GEFSv12 biases and how quickly they develop in time. At the highest levels near 2–1 hPa, the influence of the model sponge layer likely cannot be ignored; the FV3-based dynamical core of GEFSv12 (Lin 2004; Putman and Lin 2007) uses a diffusive sponge layer that applies stronger damping (relative to lower model levels) and three-dimensional Rayleigh drag at the highest model levels (Harris 2021). These numerical characteristics of the model directly influence the winds and temperatures in these layers, and can feedback on the model dynamics (Shepherd et al. 1996).

GEFSv12's biases in the tropical stratosphere from 100 to 10 hPa are best viewed as a function of the alternating phases of westerly and easterly QBO winds (WQBO and EQBO, respectively). Lawrence et al. (2022) showed that the GEFSv12 performed on par with high-top S2S forecast systems in maintaining QBO winds and temperatures, but their analyses were limited to simple time series diagnostics. Figure 3 shows GEFSv12 biases in weeks 3–5 zonal-mean zonal winds and temperatures composited by the sign of the 50-hPa tropical (from -10° to 10° latitude) zonal winds (QBO50) at initialization. Between 100 and 50 hPa, GEFSv12 winds generally decay such that they underestimate the amplitude of the observed QBO. This appears as a westerly bias during initially EQBO50 phases (Fig. 3a), and an easterly bias during initially WQBO50 phases (Fig. 3b). The raw amplitude of QBO50 in observations is asymmetric, with winds typically reaching $\sim 10 \text{ m s}^{-1}$ during WQBO50, but $\sim -15 \text{ m s}^{-1}$ during EQBO50; thus, the easterly GEFSv12 bias for hindcasts initialized in WQBO50 represents a degradation of a larger fraction of the observed signal. From 30 to 10 hPa, GEFSv12 exhibits a general westerly bias (as shown in Fig. 2), but with a much greater amplitude and

latitudinal extent for WQBO50 initializations (Fig. 3b). In observations, the direction of winds around 10 hPa are often opposite to those at 50 hPa (i.e., EQBO50 often corresponds to WQBO10; Wallace et al. 1993), meaning the large 30–10-hPa westerly biases in the WQBO50 composite indicates the winds in the middle stratosphere are not easterly enough.

This dipole pattern between the lower and middle stratosphere is more clear in the difference between the two QBO composites (Fig. 3c), which also highlights the impact on the temperature biases. Under observed conditions, cold anomalies are expected in the tropical tropopause region under EQBO50 (related to negative vertical wind shear), and opposite for WQBO50 (Collimore et al. 2003; Gray et al. 2018). However, Fig. 3c shows a relative warm bias near the equatorial tropopause (and a relative cold bias at levels above) in the hindcasts with initially easterly versus westerly QBO50, which is consistent with an underestimation of the true wind shear. The relative cold biases centered around $\pm 20^{\circ}$ in the lower stratosphere further demonstrate that the mean meridional circulation associated with the QBO becomes too weak in the hindcasts. Under observed EQBO50 conditions, tropical tropopause cold anomalies arise in association with anomalous upwelling near the equator, which tie to anomalous downwelling and warm anomalies in the subtropics (the opposite is true for WQBO50; Baldwin et al. 2001; Collimore et al. 2003). The “checkerboard” pattern of relative temperature biases in Fig. 3c, which are opposite in sign to that expected for raw EQBO50 conditions, points toward these circulations being too weak.

b. Extreme stratospheric polar vortex events

In the NH, extreme polar vortex events have been highlighted as potential “forecast windows of opportunity” for tropospheric prediction (Butler et al. 2019; Domeisen et al. 2020a; Albers and

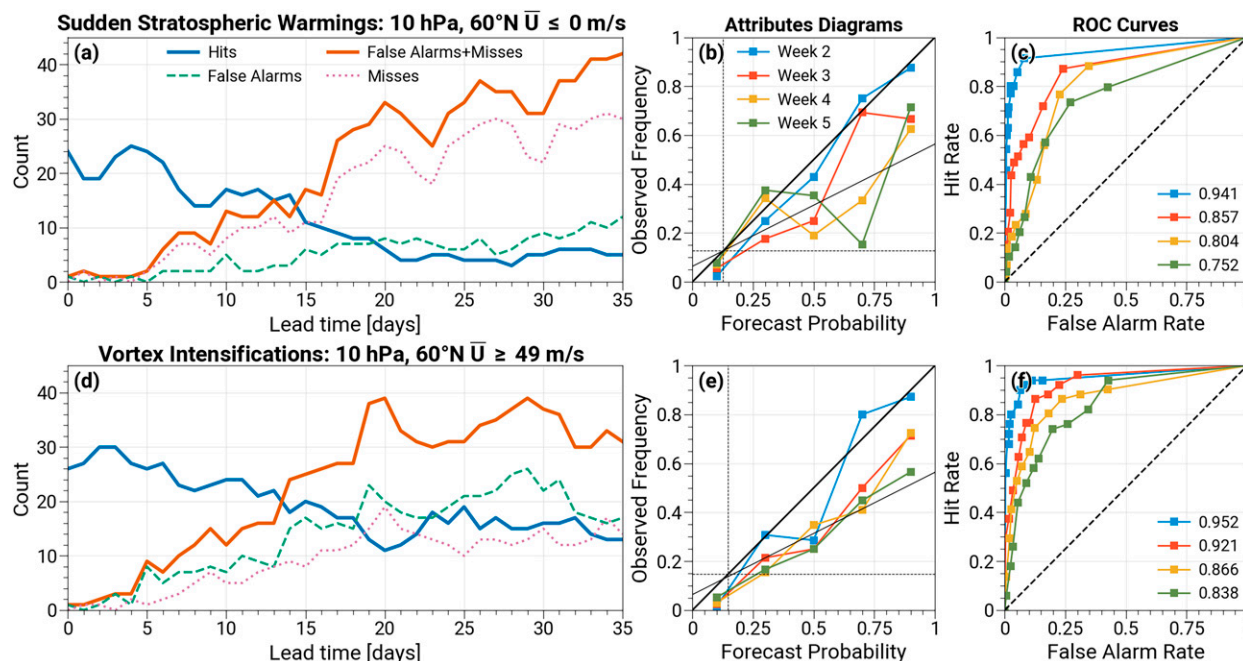


FIG. 4. GEFSv12 dichotomous predictions of (a)–(c) SSWs and (d)–(f) VIs. (left) The rate of GEFSv12 hits, false alarms, and misses as a function of daily lead time for forecasts predicting greater than 50% probability of an event/nonevent. (center) The attributes diagrams and (right) ROC curves for predictions of these events under varying probability thresholds. The attributes diagrams are computed by binning forecast probabilities from 0% to 20%, 20% to 40%, ..., 80% to 100% (plotted at the midpoint of the bins), whereas the ROC curves use probability thresholds corresponding to a change of 1 ensemble member each [from right to left of (c) and (f)], up to 11. The area under the ROC curves is indicated by the numbers in the bottom right of (c) and (f). The sample sizes associated with the attributes diagrams and ROC curves are provided in the supplemental Fig. 1. Results are based on NDJF initializations only.

Newman 2021). These events represent large and sudden changes to the strength of the stratospheric polar vortex such as SSWs (Baldwin et al. 2021) and VIs (Limpasuvan et al. 2005). SSWs and VIs can drive persistent temperature and wind anomalies in the lower stratosphere that may subsequently influence the tropospheric circulation. Skillfully predicting these events may thus extend S2S prediction skill even further (Cai et al. 2016; Scaife et al. 2022). Figure 4 assesses the ability of GEFSv12 to predict these extreme polar vortex events [as defined in section 2e(1)] for forecasts initialized in November–February.

The GEFSv12 hindcasts predict extreme polar vortex events about 2 weeks in advance, but predictions at longer leads exhibit issues possibly related to circulation biases. For SSWs (Fig. 4a), GEFSv12 forecasts with >50% ensemble agreement have a higher fraction of hits out to ~13–14 days compared to false alarms and misses. Across all lead times, the number of missed days with easterly winds exceeds the false alarm rate for this confidence level. The attributes diagram (Hsu and Murphy 1986; Wilks 2019) for SSWs (Fig. 4b) similarly shows that the week 2 predictions are reliable, but the weeks 3–5 predictions exhibit poor resolution (overconfidence) for moderate forecast probabilities. In addition, the areas under the ROC curves (Fig. 4c) drop quickly beyond week 2. Results are similar for forecasts of strong westerlies associated with VIs (Fig. 4d), but here the false alarms exceed misses across the lead times. This is particularly apparent in the

attributes diagram (Fig. 4e), where the weeks 3–5 predictions show consistently high overconfidence near the line of “no skill.” The VI ROC curves (Fig. 4f) show that the GEFSv12 predictions tend to have higher hit rates (and larger areas under the ROC curves) relative to SSWs, but also higher false alarm rates. A caveat of these results are the small sample sizes incorporated in these figures (see Fig. 1 in the online supplemental material), but they are consistent with results above: GEFSv12 exhibits a strong polar vortex bias (Fig. 2), which leads to an overprediction of threshold-based VIs on S2S time scales. On the other hand, SSWs are highly nonlinear events that can depend on both tropospheric and internal stratospheric conditions (Birner and Albers 2017; de la Cámara et al. 2017, 2019; White et al. 2020), making them generally difficult to predict on S2S time scales (possibly exacerbated by model stratospheric biases; Lawrence et al. 2022).

The methods we employ differ slightly from those that have been used in prior work investigating S2S predictability of extreme vortex events. For instance, studies such as Taguchi (2018), Domeisen et al. (2020a), Rao et al. (2021b), and Chwat et al. (2022) define hits as forecasts that successfully predict an observed event onset within a window of time (typically ± 2 –3 days). Using similar methods (see supplemental Fig. 2), we obtain results consistent with the above: The GEFSv12 hindcasts show an average ensemble confidence level of 45%–50% for week 2 predictions of SSWs and VIs in the reanalysis record.

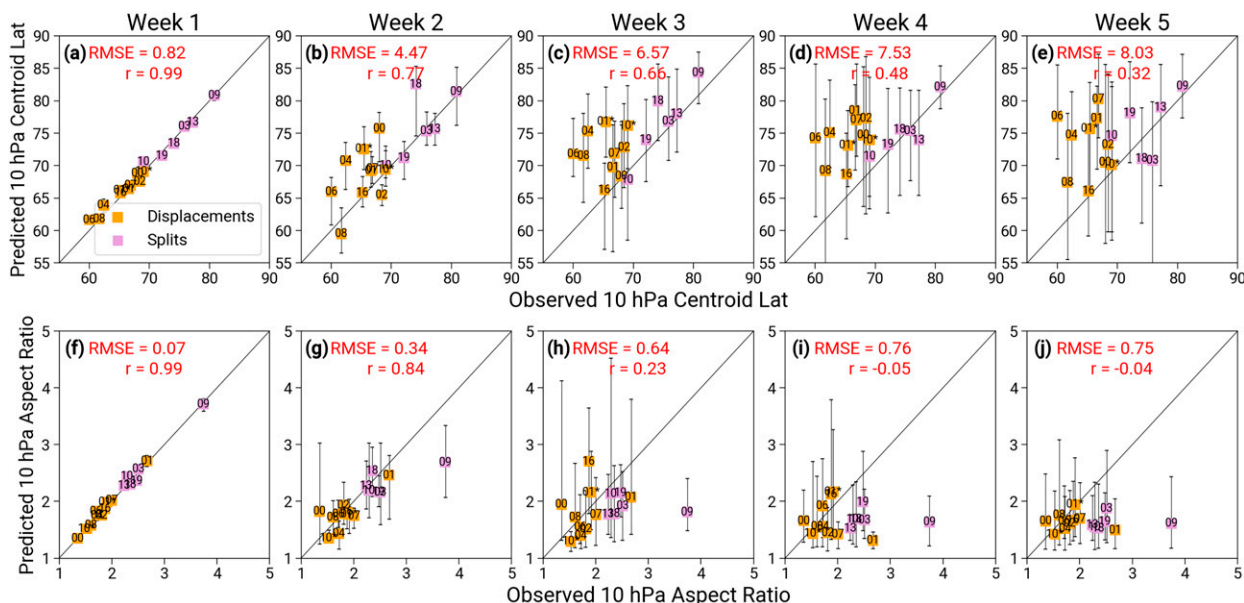


FIG. 5. Scatterplots of weekly mean, ensemble mean 10-hPa polar vortex elliptical diagnostics for hindcasts matched to historical SSWs from 2000 to 2019. (a)–(e) Centroid latitudes and (f)–(j) aspect ratios, with events classified as displacements in orange, and those classified as splits in pink. Each point is labeled with the two-digit year suffix of the event; in cases where there are two events within the same year, the later event is marked with an asterisk (e.g., “10” corresponds to the February 2010 event, whereas “10*” corresponds to the March 2010 event). Each point is plotted with whiskers representing the full ensemble spread. Each panel includes an identity line to show where the ensemble spread contains the observed value; root-mean-square errors and correlations of the composites are shown in red text.

Figure 4 suggests that GEFSv12 can predict SSWs up to 2 weeks in advance, but SSWs can be further classified as displacement or split events depending on the geometrical evolution of the vortex during the SSW. It has been found that vortex-split events are generally less predictable than vortex-displacements (Taguchi 2016, 2018; Rao et al. 2019; Domeisen et al. 2020a). Figure 5 shows scatterplots of GEFSv12 weekly ensemble-mean 10-hPa vortex elliptical diagnostics (see section a of the appendix), including centroid latitudes and aspect ratios. These metrics are compiled for hindcast initializations that place the onset dates of observed SSWs within each weekly bin, including the 13 SSW events from 2000 to 2013 listed in Butler et al. (2017), the dynamically driven “major final warming” from early March 2016 (Manney and Lawrence 2016), and the two SSWs from early 2018 and 2019 (Rao et al. 2020; Butler et al. 2020).

The GEFSv12 predictions of centroid latitude for SSWs show high correlations above 0.6 out to week 3. The correlations for aspect ratio predictions are only above 0.6 out to week 2. However, these predictions must be viewed in light of the event classifications. For instance, some of the high correlation for the weeks 2–3 centroid latitudes is attributable to GEFSv12 predicting vortex-split events to have higher centroid latitudes (Fig. 5c). In contrast, many of the week 3 centroid latitude predictions are too high for vortex-displacement events, and this is associated with relatively large root-mean-square error (RMSE) of 6.6°. The results are similar for the aspect ratio predictions, with GEFSv12 generally predicting lower weeks 2–3 aspect ratios for most of the events that develop as displacements (Figs. 5g,h). The ensemble spread for

these diagnostics can grow to be very large from weeks 3–5, suggesting GEFSv12 shows a large amount of uncertainty, as might be expected. In week 2, GEFSv12 misses some observed events within its ensemble spread, but it is clear that it can distinguish between the event types as demonstrated by the stratification of the centroid latitudes and aspect ratios. Taken together with Fig. 4, the GEFSv12 hindcasts reliably predict the occurrence and geometrical evolution of SSWs up to 2 weeks in advance.

c. Extreme meridional heat flux events

The strength of the polar vortex is determined by a relative balance between radiative and dynamical influences, the latter of which is primarily driven by vertically propagating planetary-scale waves. These upward/downward propagating waves, most commonly from zonal wavenumber 1, can be diagnosed with meridional eddy heat fluxes (Andrews et al. 1987). Extremely positive heat flux events weaken the polar vortex and make it more susceptible to break down from further wave driving, whereas negative heat flux events are associated with the downward reflection of planetary waves, which can dynamically strengthen the polar vortex and affect the tropospheric circulation (Dunn-Sigouin and Shaw 2015). Because these events [as defined in section 2e(1)] tend to occur on shorter time scales relative to SSWs and VIs, we examine the GEFSv12 hit, miss, and false alarm rates as a function of pentad lead time to make the results less sensitive to onset date and event duration.

Extreme heat flux events are only predictable by GEFSv12 up to 6–10 days in advance. Figure 6 shows that by

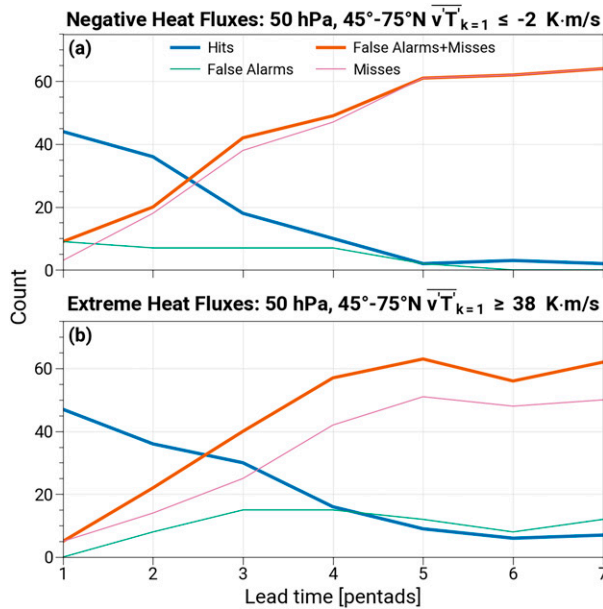


FIG. 6. As in Figs. 4a and 4d, but for (a) negative and (b) extreme positive wave-1 heat flux events. In this case, the results are shown as a function of pentad lead time.

11–15 days, the hit rate of forecasts with greater than 50% ensemble confidence is exceeded by the false alarms and misses. For the extreme positive events (Fig. 6b), GEFSv12 predicts roughly equal numbers of hits and false alarms beyond pentad 3, but misses overall exceed the false alarms. In contrast, for negative heat flux events (Fig. 6a), GEFSv12's hits and false alarms approach 0 by pentad 5, meaning such events are rarely ever predicted to happen with high ensemble confidence. Using the alternative definition for hits and false alarms (e.g., as in Domeisen et al. 2020a) gives similar results, with GEFSv12 only having a hit rate of about 35%–40% for extreme heat flux events in week 2 (supplemental Fig. 2).

GEFSv12's zonal-mean circulation biases may affect its predictions of negative heat flux events. Its polar vortex wind biases that develop in the upper stratosphere (Figs. 2a,b) affect the vertical wind shear, which in turn influences the propagation characteristics for planetary waves by altering the index of refraction (Perlwitz and Harnik 2003). Perlwitz and Harnik (2003, 2004) defined a simple reflection index " \bar{u}_{2-10} ," defined as the 2–10-hPa difference of 58°–74°N zonal-mean zonal winds; negative values of \bar{u}_{2-10} indicate negative vertical wind shear, which correspond to states that support downward wave reflection described by negative heat flux events. Figure 7 shows the GEFSv12 distributions of \bar{u}_{2-10} and the midlatitude 50-hPa wave-1 eddy heat fluxes as a function of pentad lead time compared to ERA5. There is a clear positive shift in the GEFSv12 \bar{u}_{2-10} distributions relative to ERA5 apparent from pentad 1 that amplifies over time (Fig. 7), in agreement with the quick development of zonal wind biases shown in Fig. 2b. Consequently, the frequency of 0 m s^{−1} wind shear events (indicating the threshold for negative shear) starts out around 24% and drops to 19% by pentad 3,

compared to ERA5's approximately constant 28% (Fig. 7b). The results are similar for the 50-hPa eddy heat fluxes; the GEFSv12 distributions show a positive shift with lead time (Fig. 7c), and the frequencies for 0 K m s^{−1} drop from 15% to 10%–11% by pentads 3–5. These results suggest that GEFSv12's strengthening of the polar vortex with lead time (Fig. 2b) often prohibits the development of negative vertical wind shear (Fig. 7b), which may undermine the ability of GEFSv12 to simulate wave reflection events.

d. Extratropical stratosphere–troposphere coupling

On S2S time scales, the NH stratospheric polar vortex can influence the evolution of the tropospheric NAM/NAO. This coupling can be described, in simple terms, as a two-step process: First, anomalous time-integrated wave fluxes from the troposphere drive anomalies in the strength of the polar vortex ("upward coupling"; Newman et al. 2001; Polvani and Waugh 2004). Subsequently, the strength of the polar vortex (represented by the stratospheric NAM) may then influence the tropospheric NAM state down to the surface ("downward coupling"; Kidston et al. 2015), particularly if the anomalous vortex strength is persistent and reaches into the lowest levels of the stratosphere (Karpechko et al. 2017; White et al. 2020). This is a simplified description since the tropospheric teleconnections that drive upward wave fluxes can also "directly" influence the NAM/NAO (i.e., separate from the stratosphere; Jiménez-Esteve and Domeisen 2018; Barnes et al. 2019; Domeisen et al. 2019; Albers and Newman 2021), and extreme polar vortex events can be related to internal stratospheric variability separate from vertical wave fluxes (Birner and Albers 2017; White et al. 2019). Nevertheless, stratosphere–troposphere coupling relationships are clear in reanalysis data, but they are subject to a large amount of interannual variability associated with the occurrence of extreme polar vortex events, and the varying tropospheric responses that can ensue (Gerber and Martineau 2018; Hitchcock 2019). Sampling uncertainty can thus affect statistical relationships when limited to specific periods of years. However, the above processes underpin why many S2S forecast systems show enhanced NAM/NAO predictive skill following extreme weak or strong states of the polar vortex (Sigmond et al. 2013; Tripathi et al. 2015; Domeisen et al. 2020b; Sun et al. 2020).

We first examine the "two-step coupling" relationships in Fig. 8, with polar cap-averaged geopotential height anomalies acting as proxies for the NAM (Baldwin and Thompson 2009). The left and right columns of Fig. 8, respectively, are intended to correspond to the upward and downward coupling processes discussed above, with the 100-hPa anomalous eddy heat flux acting as a proxy for anomalous vertical wave fluxes from the troposphere. GEFSv12 captures the strong relationship between wave fluxes and polar vortex strength with a correlation of 0.71 for NDJF initializations, indicating that weaker (stronger) heat fluxes earlier in the forecasts also correspond to stronger (weaker) vortex states later in the forecasts. This correlation is particularly strong in the full ERA5 record with a 99% confidence interval from 0.73 to 0.8, which excludes the GEFSv12 correlation. However, in the 20-yr

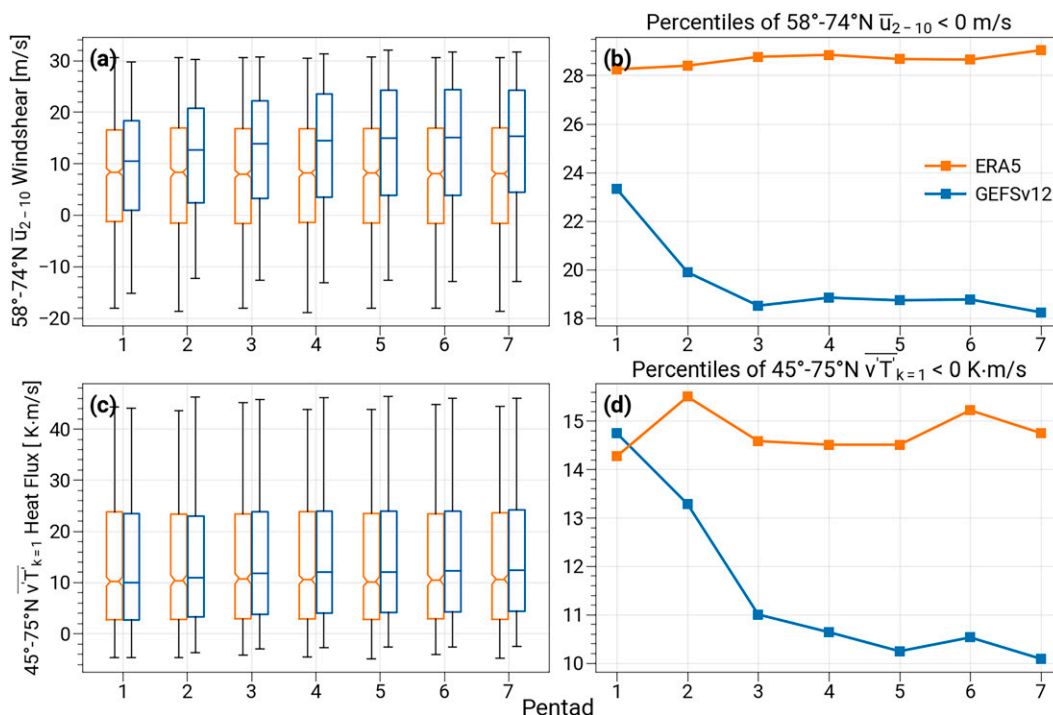


FIG. 7. (left) Boxplots showing the distributions of daily (a) 58°–74°N zonal wind shear between 2 and 10 hPa, and (c) the 45°–75°N wave-1 eddy heat flux at 50 hPa as a function of pentad lead time. The boxplot whiskers show the range between the 5–95 percentiles, with medians plotted in the horizontal lines; the ERA5 boxes in orange have notches representing the 95% bootstrap confidence interval on the medians. (right) The time evolution of the percentiles for (b) negative wind shear and (d) heat flux events (at the threshold of 0 m s⁻¹ and 0 K m s⁻¹, respectively).

GEFSv12 hindcast period, the GEFS-R correlation is 0.72, with a 99% confidence interval from 0.65 to 0.78. The results are similar for “downward coupling” between the 100- and 1000-hPa NAM states, with GEFSv12 showing a correlation of 0.46 for DJFM initializations. This correlation is smaller than those from the reanalyses, but the associated sampling uncertainty for the shorter period of 2000–19 covers the GEFSv12 value in the GEFS-R 99% confidence interval from 0.4 to 0.6. These relationships hold for different nonoverlapping combinations of weeks/pentads, but with correlations that are somewhat weaker (not shown). Generally, these correlations are largest for time averages that exceed the number of days available in the hindcasts (e.g., Polvani and Waugh 2004), which justifies our use of longer overlapping periods. Nevertheless, the GEFSv12 hindcasts show NAM-coupling relationships that match closely with reanalysis.

The persistence of the near-surface NAM is known to be longer during boreal winter, which is partially related to stratospheric variability (Baldwin et al. 2003; Gerber et al. 2010; Simpson et al. 2011; Kidston et al. 2015). Figure 9 shows different lead-lag correlations of the NAM index, comparing the correlation of the NAM at 1000 hPa with the NAM at 100 hPa and with itself (persistence) for different weekly lead times and DJFM initializations. The reanalyses are sampled to match the GEFSv12 hindcasts, which means their patterns of correlations are nearly identical for every lag, by definition. The influence of sampling uncertainty is clear from comparing

the correlations from GEFS-R to those from ERA5; the correlations for GEFS-R are generally smaller for both the stratosphere–troposphere and persistence relationships. For GEFSv12, every “pixel” is based on predicted NAM values. The hindcasts show stratosphere–troposphere and persistence correlations that are nearly constant for every equal lag, similar to the reanalyses (e.g., the week 1 100-hPa NAM correlates with the week 3 1000-hPa NAM to the same extent as week 2 with week 4, and week 3 with week 5). The GEFSv12 correlations are somewhat smaller than those from the reanalyses, but they are not significantly different from GEFS-R. Comparing the stratosphere–troposphere and persistence correlations makes it clear that, on S2S time scales, the surface NAM is more highly correlated with the stratospheric NAM than persistence, with positive differences (Figs. 9g–i) for lags of 2–4 weeks. In GEFSv12 and ERA5, these differences are statistically significant at the 95% level.

To assess the influence of the initialized polar vortex strength on GEFSv12 NAO forecasts, we divide GEFSv12 hindcasts into composites with initially weak, strong, or neutral polar vortex winds following Sun et al. (2020). We use equivalent percentiles and methods to identify strong (80th percentile, 39.3 m s⁻¹), weak (15th percentile, 6.6 m s⁻¹), or neutral initial vortex conditions for NDJFM seasons in the 2000–19 period, which gives 62 weak, 86 strong, and 201 neutral cases (Fig. 10a). Figures 10b and 10c show the weekly correlation skill scores (CSS) and subset correlations (SSC) of

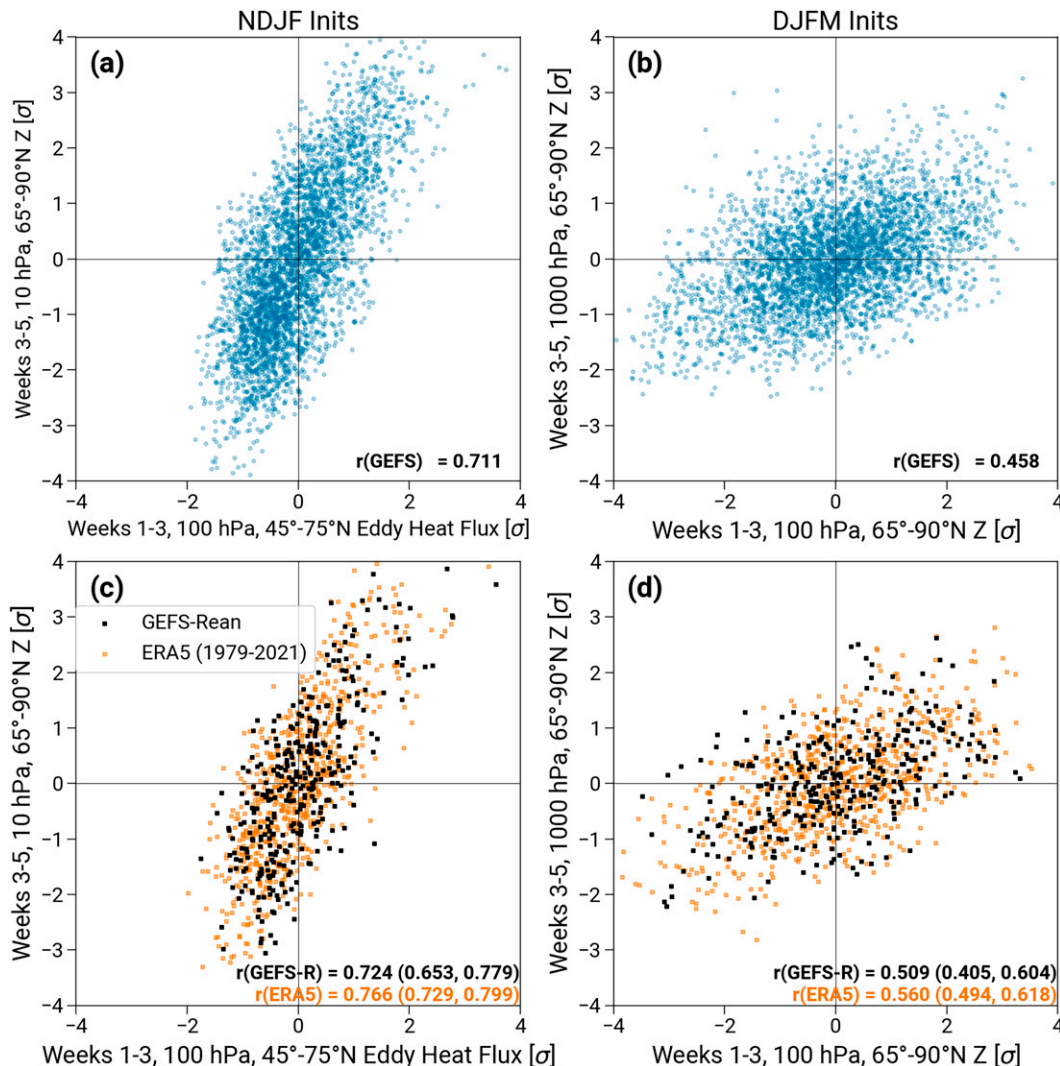


FIG. 8. Scatterplots of (left) weeks 1–3 100-hPa eddy heat flux anomalies against the weeks 3–5 10-hPa polar cap geopotential height anomalies for (a) GEFSv12 NDJF hindcasts and (c) GEFS-R (black squares) and ERA5 (orange squares). (right) Weeks 1–3 polar cap height anomalies at 100 hPa against those for weeks 3–5 at 1000 hPa for (b) GEFSv12 DJFM hindcasts and (d) reanalysis. In (a) and (b), all GEFSv12 ensemble members are plotted for each initialization. In (c) and (d), the reanalyses are sampled to match the GEFSv12 initializations, with ERA5 extended backward to 1979. Correlations are listed in the bottom-right corner of each panel; in (c) and (d), the parentheses indicate the 99% bootstrap confidence interval from resampling 50 000 times.

the ensemble mean NAO forecasts for each of the vortex composites as defined in [Charlton-Perez et al. \(2021\)](#). Similar to prior studies ([Sigmond et al. 2013](#); [Tripathi et al. 2015](#); [Sun et al. 2020](#)), GEFSv12 exhibits enhanced NAO correlation skill at weeks 3–5 for forecasts initialized with weak and strong polar vortex conditions relative to neutral ([Fig. 10b](#)). However, these differences are generally not statistically significant in GEFSv12, except for the strong vortex composite at week 4. The results for SSC ([Fig. 10c](#)) are similar to those for the correlation skill, but they differ in one important way: the SSC values for strong and weak states do not differ as much from the neutral composite. Furthermore, from weeks 3–5, the strong vortex composites have higher SSC values than the

weak vortex composites, especially at week 4. In light of the discussion in [Charlton-Perez et al. \(2021\)](#), we interpret the CSS results as reflecting a shift in the distribution of NAO states within the forecasts (i.e., a shift in the signal resulting in a higher signal-to-noise ratio) related to the polar vortex events. In contrast, the SSC results indicate that forecasts initialized during strong polar vortex states can be significantly more skillful at predicting NAO variations within the shifted background state than those initialized during neutral states.

e. Tropical stratosphere–troposphere coupling

The MJO tends to be more strongly organized and active during EQBO50 relative to WQBO50 ([Yoo and Son 2016](#);

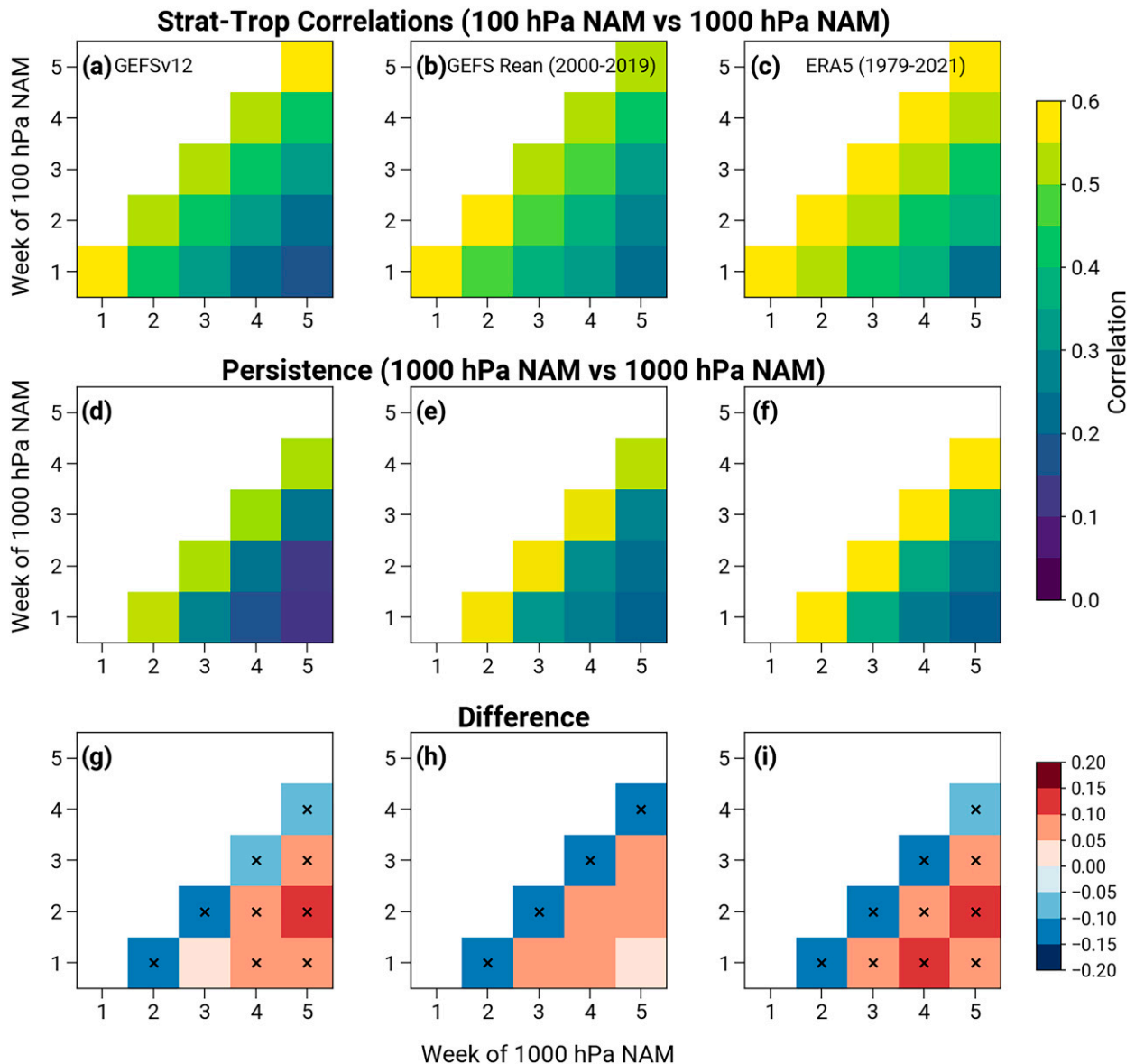


FIG. 9. Lag correlations of the NAM index for (a)–(c) stratosphere–troposphere (100-hPa NAM leading 1000-hPa NAM) and (d)–(f) persistence (1000-hPa NAM leading 1000-hPa NAM) relationships, and their (g)–(i) difference (top row minus middle row). In (d)–(f), the diagonal is omitted, since by definition the correlations are equal to 1. In (g)–(i), statistically significant differences are marked with an x, which are determined at the 95% level from bootstrap resampling 50 000 times.

Son et al. 2017). While studies continue to investigate QBO–MJO relationships (Lim and Son 2022), many unknowns remain. For instance, it is not clear why the QBO–MJO relationship appears to have emerged after 1979 (Klotzbach et al. 2019) or why it is predominantly a boreal winter phenomenon (Son et al. 2017; Sakaeda et al. 2020). Moreover, the QBO–MJO relationship is often not well captured by models (Martin et al. 2021; Kim et al. 2019, 2020), which makes it difficult to study the relationship in the absence of other forcings such as El Niño–Southern Oscillation. Nonetheless, there is a growing understanding that the tropical lower stratospheric temperature and stability perturbations imparted by the QBO

can make for more or less favorable conditions for the MJO. When the tropical lower stratosphere is cooler during EQBO, the MJO typically propagates more slowly and with larger amplitude compared to WQBO (Son et al. 2017; Lim et al. 2019; Densmore et al. 2019). Among S2S forecast models, the MJO is typically predictable at longer lead times during EQBO (Marshall et al. 2017; Abhik and Hendon 2019; Lim et al. 2019; Wang et al. 2019), but the relationship between QBO phase and MJO skill is often not statistically significant (Kim et al. 2019).

We first examine whether GEFSv12 exhibits MJO amplitude differences consistent with the phase of the QBO. Figure 11

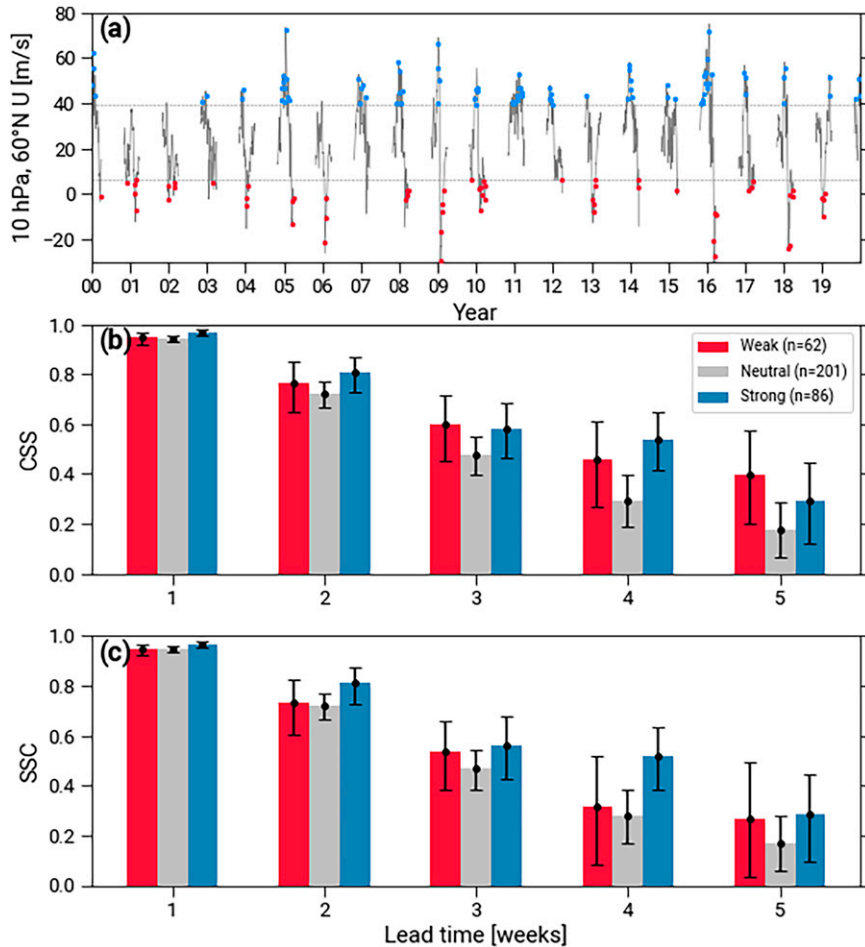


FIG. 10. (a) Time series of NDJFM 10-hPa 60°N zonal-mean zonal winds from 2000 to 2019 in ERA5 reanalysis. (b) Correlation skill scores (CSS) and (c) subset correlations (SSC) for ensemble-mean NAO hindcasts composited by the initialized polar vortex state. The weak, neutral, and strong composites and initializations are shown in red, gray, and blue, respectively. The number of initializations in each composite are shown in the legend in (b). As in [Sun et al. \(2020\)](#), the error bars in (b) and (c) are 90% confidence intervals determined from 50 000 bootstrap resamples.

shows how the MJO amplitude ($\sqrt{\text{RMM1}^2 + \text{RMM2}^2}$) varies according to MJO phase [$\arctan(\text{RMM1}/\text{RMM2})$] when the initialized QBO is defined using 30-hPa (QBO30) versus 50-hPa (QBO50) tropical zonal winds. Unlike other studies ([Kim et al. 2019](#); [Lim et al. 2019](#)), we simply use the sign of the GEFS-R QBO winds at initialization to delineate between QBO phases for the hindcasts. In ERA5, the amplitude of the MJO is generally larger during EQBO than WQBO. For QBO30, the amplitudes are well separated between phases 3–6 ([Fig. 11a](#)). This amplitude separation is weaker but still present for QBO50 ([Fig. 11b](#)). GEFSv12 MJO amplitudes for individual members ([Figs. 11c,d](#)) are biased high relative to those from ERA5 at extended lead times (see supplemental Fig. 3). Note, however, that the GEFSv12 ensemble-mean MJO amplitudes at these same lead times are biased low compared to ERA5 (not shown); this discrepancy arises because the amplitude equation is nonlinear, so the amplitude of the ensemble-mean MJO is

not the same as the average of individual MJO amplitudes ([Kim et al. 2014](#)). Regardless, the GEFSv12 hindcasts show significant week 3–5 MJO amplitude differences under QBO30 ([Fig. 11c](#)) that mostly disappear under QBO50 ([Fig. 11d](#)). These results suggest that the amplitude of the MJO in the GEFSv12 hindcasts, particularly when the MJO passes over the Maritime Continent (phases 4–5) and the tropical western Pacific (phase 6), is sensitive to the level used to define the QBO. Perhaps in agreement with this, [Densmore et al. \(2019\)](#) showed that the amplitude of the MJO increases more after passing over the Maritime Continent during middle stratospheric EQBO than during lower stratospheric EQBO.

The pressure level used to define the QBO also impacts the GEFSv12 MJO prediction skill. [Figure 12](#) shows the bivariate anomaly correlations (BCOR) for MJO RMM1 and RMM2 of the GEFSv12 hindcasts with ERA5 reanalysis for DJFM initializations. We consider only initially strong MJO events

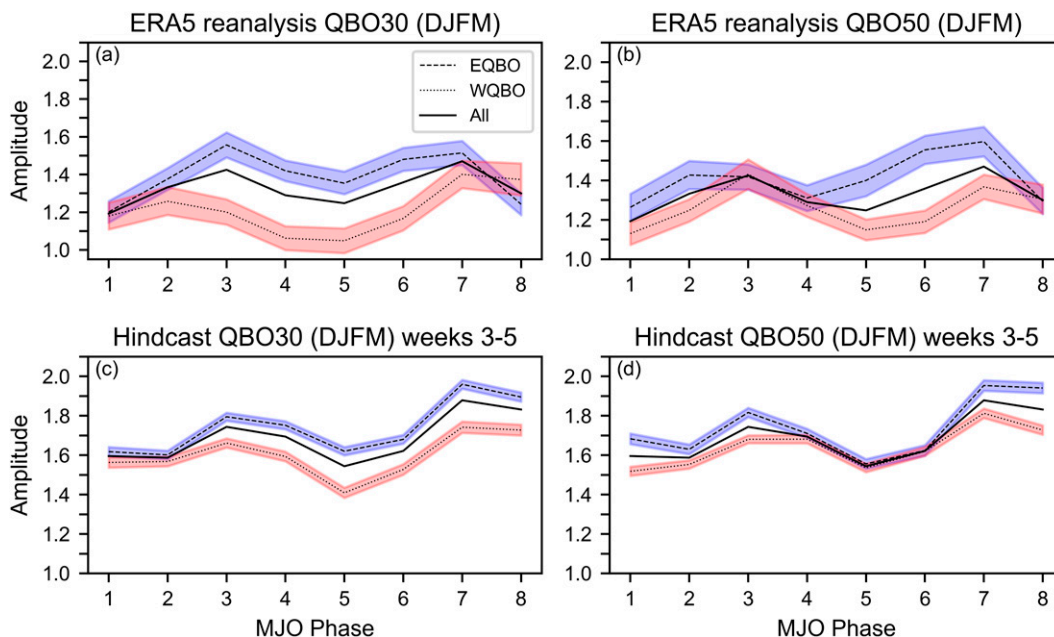


FIG. 11. MJO amplitudes as a function of MJO phase, composited by (a),(b) the initial QBO state for ERA5 and (c),(d) the weeks 3–5 GEFSv12 hindcasts for all individual members. Results for (left) QBO30 and (right) QBO50 are shown. The ERA5 data are sampled equivalently to the GEFSv12 hindcasts. The colored envelopes represent 95% bootstrap confidence intervals around the EQBO and WQBO composites. The black lines show the amplitudes after compositing regardless of QBO phase.

with ERA5 MJO amplitudes greater than 1.5, and use a BCOR threshold of 0.5 to delineate skillful MJO prediction. Across all DJFM initializations, the GEFSv12 BCOR remains above 0.5 out to 26 days when verified against ERA5, which is on par with some of the better performing forecast systems (Lim et al. 2019; Kim et al. 2019). If we composite based on QBO30, the BCOR skill is maintained out to 29 days for EQBO initializations, but only 23 days for WQBO (Fig. 12a). These easterly versus westerly QBO30 skill differences are

statistically significant for three days during week 4. Conversely, QBO50 indexing yields no significant difference in skill between EQBO and WQBO (Fig. 12b). Using different MJO amplitude thresholds (e.g., 0 or 1) yields similar results (not shown). Since prior studies have commonly used QBO50 when evaluating QBO–MJO relationships, it is unclear whether other forecast systems would also exhibit more robust skill separation between EQBO and WQBO if a 30-hPa index were used instead. The stronger relationship between the MJO and the

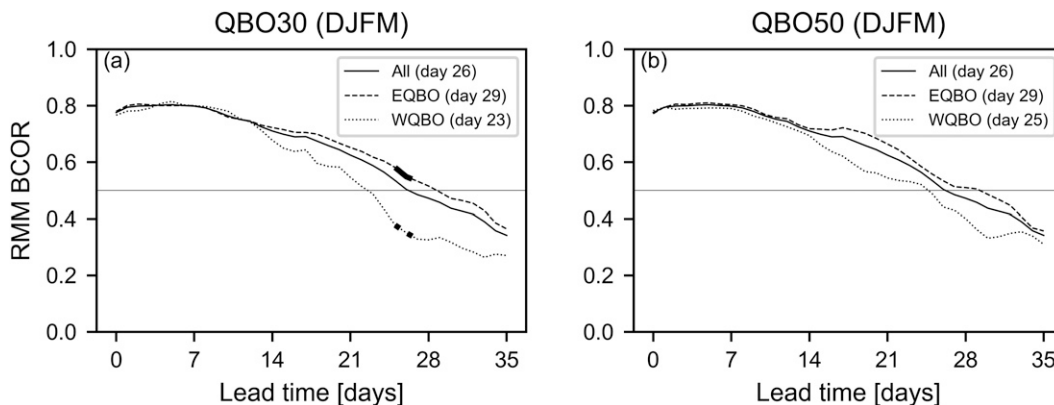


FIG. 12. GEFSv12 MJO bivariate anomaly correlations (BCOR) shown as a function of lead time for DJFM (a) QBO30 and (b) QBO50 initializations. The solid line represents the BCOR for all DJFM initializations, while the dashed and dotted lines are those for EQBO and WQBO composites, respectively. The bold lines show statistically significant differences at the 95% level between the EQBO and WQBO composites as determined from bootstrapping.

phase of the QBO at 30 hPa may arise because middle stratospheric EQBO is associated with stronger decreases in tropopause level stability and temperature than lower-stratospheric EQBO in reanalysis (Densmore et al. 2019).

4. Summary and discussion

We have performed a comprehensive evaluation of GEFSv12 subseasonal hindcasts focused on stratospheric biases, extreme stratospheric events, and sources of extratropical and tropical stratosphere–troposphere coupling. The GEFSv12 reanalysis used to initialize the hindcasts exhibits systematic differences from the more commonly used ERA5 reanalysis in global mean stratospheric temperatures, and tropical stratospheric zonal winds. In the GEFSv12 hindcasts, stratospheric biases develop with lead time, including a global mean warm bias and a too strong wintertime polar vortex (similar to other subseasonal forecast systems, e.g., Lawrence et al. 2022); these biases are generally more pronounced than any modern reanalysis differences. After compositing tropical zonal wind and temperature biases by initial QBO phase, we showed that the QBO-associated winds and meridional circulation become too weak in the GEFSv12 hindcasts. We also assessed GEFSv12's ability to predict SSWs, vortex intensifications, and positive/negative heat flux events; we showed that GEFSv12 can generally predict extreme vortex events about 13 days in advance, and extreme heat flux events 6–10 days (2 pentads) in advance. In terms of stratosphere–troposphere coupling, the GEFSv12 hindcasts reproduce observed statistical relationships between the NH polar vortex and tropospheric NAM, and QBO-dependent differences in MJO amplitudes across different MJO phases. Importantly, GEFSv12 shows enhanced weeks 3–5 correlation skill scores in predicting the NAO when initialized during strong/weak vortex events, and enhanced MJO bivariate anomaly correlations out to 29 days during easterly 30-hPa QBO phases.

Our results have operational implications for forecasts made by GEFSv12. For example, the GEFSv12 system is able to reliably predict SSWs up to 2 weeks in advance (Figs. 4a–c), including their morphology (i.e., split versus displacement; Fig. 5). As a result, week 2 forecasts with a strong signal for a predicted SSW should correspond to a “believable” increased likelihood of a consistent tropospheric NAM/NAO response in the weeks following. To the extent that GEFSv12 predicts consistent stratospheric polar cap height anomalies associated with SSWs within weeks 1–2, it is also more likely to predict a consistent NAM/NAO response for weeks 3–5 (Fig. 8b). By the time GEFSv12 forecasts are initialized with an SSW (or strong vortex), more confidence can be given to its weeks 3–5 NAO predictions (Fig. 10). Similarly, GEFSv12 MJO forecasts are likely to exhibit a dependence on the phase of the QBO. GEFSv12 forecasts initialized with 30-hPa easterly QBO winds are likely to exhibit enhanced MJO amplitudes (Fig. 11) and better weeks 3–4 MJO prediction skill (Fig. 12) than those initialized during westerly phases.

The extratropical stratospheric wind biases exhibited by GEFSv12 have an apparent effect on its stratospheric predictions. While GEFSv12 is able to predict extreme polar vortex

events about 2 weeks in advance (similar to other subseasonal forecast systems; Domeisen et al. 2020a), the evolution toward a strong-biased vortex (Fig. 2) seems to impact its hindcast event statistics (Fig. 4). The attributes and ROC diagrams in Fig. 4 suggest that different ensemble confidence levels could possibly be used to better predict SSW versus VI conditions to partially account for such a bias (fewer members needed for SSWs, but more for VIs). Lawrence et al. (2022) showed that simple linear bias correction could result in modest improvements to subseasonal hindcast probabilities of extreme vortex events. Similarly, Rao et al. (2019) found that correcting the distributions of stratospheric winds in S2S forecasts with reanalysis also improved skill in predicting SSWs. Similar bias correction methods could be used to rule out GEFSv12 false alarm vortex intensifications and potentially increase the predictive skill for SSWs. However, bias correction cannot account for dynamical mechanisms that potentially lead to enhanced tropospheric predictability. For example, the fact that downward wave reflection events are less likely to be forecast by GEFSv12 beyond ~2 weeks is likely tied to the strong vortex bias (Fig. 7), which affects the vertical curvature of the stratospheric winds. These downward wave events are known to be tied to, e.g., positive phases of the NAO (Shaw and Perlwitz 2013; Dunn-Sigouin and Shaw 2015), and North American cold spells (Matthias and Kretschmer 2020; Millin et al. 2022; Messori et al. 2022). GEFSv12 may thus underestimate or miss important stratospheric connections to wintertime tropospheric circulation variability because of its circulation biases. More generally, this complex interplay between model biases and coupling processes could be a factor relevant for other S2S prediction systems.

The patterns of stratospheric temperature and zonal wind biases in GEFSv12 are suggestive that, dynamically, there is a lack of stratospheric wave driving in the model relative to reanalysis. The cold winter pole and strong polar vortex biases (Fig. 2) are indicative of a too weak Brewer–Dobson circulation (Butchart 2014; Kawatani et al. 2019). Similarly, the amplitude of the QBO is known to be sensitive to the amount of parameterized gravity wave flux (Garfinkel et al. 2022), meaning the GEFSv12 decay of tropical stratospheric zonal wind amplitudes (Fig. 3) is also likely related to a lack of wave driving. These processes have been shown to be sensitive to details of the gravity wave parameterizations implemented in models, as well as to horizontal and vertical resolution (Hamilton et al. 1999; Kawatani et al. 2019; Garfinkel et al. 2022). GEFSv12 includes parameterizations for orographically and convectively generated gravity waves (Guan et al. 2022). Its atmospheric model uses 64 model levels up to a top of 0.2 hPa, for which about 25 are at pressures below 100 hPa (Hamill et al. 2022), corresponding to a vertical resolution close to 1 km in the lower stratosphere. Thus, the lack of parameterized gravity wave sources from jets and fronts (e.g., Plougonven and Zhang 2014; de la Cámara and Lott 2015) and the somewhat coarse vertical resolution in the lower stratosphere (Garfinkel et al. 2022) may both play roles in the inability to better maintain initialized QBO conditions.

Several of our results highlight benefits of data availability on an increased number of *output* pressure levels for assessing

model biases and stratosphere–troposphere coupling relationships. Many of the GEFSv12 stratospheric biases are particularly large above 10 hPa (Fig. 2); these include tropical middle-to-upper stratosphere biases apparent in the initial conditions (Fig. 1) that are maintained through the forecasts. Gray et al. (2022) suggest the representation of the upper stratosphere could be important for simulating polar vortex disturbances such as SSWs. An upper limit of 10 hPa as commonly used for SubX and S2S database output (Vitart et al. 2017; Pegion et al. 2019) is therefore insufficient for evaluating potential impacts of the upper stratosphere. We also highlighted that the mean meridional circulation associated with the QBO (QBO-MMC) becomes too weak in GEFSv12 (Fig. 3c). The QBO-MMC is thought to impact the tropospheric jets and teleconnections with the stratospheric polar vortex (Garfinkel and Hartmann 2011; Garfinkel et al. 2012; Gray et al. 2018; Hitchman et al. 2021; Rao et al. 2021a). However, understanding these mechanisms requires usage of, e.g., transformed Eulerian mean diagnostics, which require vertical derivatives of basic state atmospheric fields. The GEFSv12 hindcast archive is likely among the first S2S datasets for which data are publicly available on enough pressure levels to make such computations viable.

Our evaluation of the GEFSv12 subseasonal hindcasts can serve as a benchmark for future versions of the UFS and GEFS. The next version of GEFS/UFS updates will feature model configurations and updates that can have substantial impacts on the representation of the stratospheric circulation. For instance, the current version 16 of NOAA's Global Forecast System included updates that increased the height of the model top from ~55 to ~80 km, increase the number of model vertical levels from 64 to 127, and new parameterizations for nonstationary gravity wave drag (Farrar 2021). In-development subseasonal UFS "prototypes" also include similar updates (Xue et al. 2021). Our results therefore provide a baseline from which the representation of the stratosphere and stratosphere–troposphere coupling in future versions of GEFS/UFS can be measured.

Acknowledgments. This material is based upon work supported by the NOAA/NWS Office of Science and Technology Integration under Grant NA20NWS4680051. ZDL, DE, and ER were supported in part by NOAA Cooperative Agreements NA17OAR4320101 and NA22OAR4320151. We thank the three anonymous reviewers whose comments helped improve our paper.

Data availability statement. The GEFSv12 hindcast archive is maintained at <https://registry.opendata.aws/noaa-gefs-reforecast/>, while the GEFSv12 reanalysis data are available via FTP from ftp://ftp.emc.ncep.noaa.gov/GEFSv12/reanalysis/FV3_reanalysis/. The ERA5 reanalysis data for atmospheric fields used herein are available from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels>. Our post-processed zonal-mean datasets are available at <https://downloads.psl.noaa.gov/Projects/GEFSv12/zmdiags/>, while those for the NAO and MJO indices are available at <https://doi.org/10.25921/be7v-1s13> and <https://doi.org/10.25921/pqq7-3n58>.

The Python package "pyzome" (<https://github.com/zdlawrence/pyzome>) was used to compile the zonal-mean datasets, while the code for computing the polar vortex moment diagnostics was based on <https://github.com/zdlawrence/grid2d-moments-utils>.

APPENDIX

Computation of Weather and Climate Indices

a. Elliptical diagnostics of the polar vortex

To quantify geometrical characteristics of the stratospheric polar vortex, we calculate elliptical diagnostics (Vaugh 1997; Matthewman et al. 2009) using geopotential heights, similar to Seviour et al. (2013). We define the Northern Hemisphere (NH) polar vortex edge at 10 hPa as the climatological 60°N zonal-mean geopotential height for November–March; these values are calculated separately for ERA5 and GEFS-R, with the GEFS-R value being used for the GEFSv12 hindcasts. The elliptical diagnostics are then calculated "geometrically" – that is, without weighing the calculations by the heights inside the vortex edge. These calculations describe the geometry of the vortex, including its centroid location and stretch (aspect ratio), which can be useful for understanding polar vortex disturbances such as SSWs (Hannachi et al. 2011; Taguchi 2016; Hall et al. 2021). In cases where the vortex is split, these diagnostics are computed for the bulk region (not for individual daughter vortices), as in Seviour et al. (2013).

b. North Atlantic Oscillation and Northern Annular Mode indices

The predominant NAO pattern describing a north–south dipole of pressure anomalies between high- and midlatitudes over the North Atlantic is known to be seasonally dependent (Barnston and Livezey 1987; Hurrell and Van Loon 1997). To compute NAO indices we project daily 500-hPa geopotential height (Z500) anomalies in the 20°–90°N, 90°W–40°E region onto monthly varying EOF1 loading patterns. These loading patterns are determined from monthly reanalysis Z500 anomalies. We create the monthly varying loading patterns by pooling together anomalies from the 3 months centered on the month in question and taking the first EOF. For example, the January EOF1 loading pattern is computed using all historical December–February months. The daily GEFSv12 hindcast Z500 anomalies used to compute the hindcast NAO indices are calculated by removing lead-dependent climatologies as described by (Pegion et al. 2019). For reference, the daily NAO indices calculated from the GEFS-R and ERA5 datasets are highly correlated with $r = 0.96$ (0.97 for NDJFM days), suggesting the loading patterns are not highly sensitive to the period of years considered (1979–2019 for ERA5 versus 2000–19 for GEFS-R).

The NAM pattern characterizes vertically resolved hemispheric-scale geopotential height anomalies of opposing sign between the pole and midlatitudes. We use standardized 65°–90°N geopotential height anomalies as a

proxy for measuring the NAM index on pressure levels (Baldwin and Thompson 2009).

c. Real-time multivariate MJO indices

We compute reanalysis and hindcast RMM indices according to the methods described in Wheeler and Hendon (2004) and Gottschalck et al. (2010), respectively. We obtain the RMM indices from a combined EOF analysis of equatorially averaged (15°S–15°N) anomalies in daily mean outgoing longwave radiation (OLR) and zonal winds at 200 and 850 hPa (U200, U850). We calculate EOFs from ERA5 for the verification RMMs, but use GEFS-R to compute the GEFSv12 RMMs. Prior to the EOF analysis, the time mean and first three harmonics of the annual cycle are removed from the reanalysis fields. For GEFSv12, lead-dependent climatologies are removed following Pegion et al. (2019). To isolate intraseasonal variability, for the reanalysis, we subtract the mean of the preceding 120 days of data from each time step; for the hindcasts, we subtract the mean of the preceding 120 days of combined GEFS-R/GEFSv12 data. The ERA5 reanalysis and GEFSv12 hindcast RMMs are retrieved by projecting the OLR/U200/U850 anomalies onto the ERA5 EOFs and GEFS-R EOFs, respectively, then normalizing the resulting projection coefficients by their respective observed standard deviations.

REFERENCES

- Abhik, S., and H. H. Hendon, 2019: Influence of the QBO on the MJO during coupled model multiweek forecasts. *Geophys. Res. Lett.*, **46**, 9213–9221, <https://doi.org/10.1029/2019GL083152>.
- Albers, J. R., and M. Newman, 2021: Subseasonal predictability of the North Atlantic Oscillation. *Environ. Res. Lett.*, **16**, 044024, <https://doi.org/10.1088/1748-9326/abe781>.
- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.
- Baldwin, M. P., and D. W. J. Thompson, 2009: A critical comparison of stratosphere–troposphere coupling indices. *Quart. J. Roy. Meteor. Soc.*, **135**, 1661–1672, <https://doi.org/10.1002/qj.479>.
- , and Coauthors, 2001: The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179–229, <https://doi.org/10.1029/1999RG000073>.
- , D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301**, 636–640, <https://doi.org/10.1126/science.1087143>.
- , and Coauthors, 2021: Sudden stratospheric warmings. *Rev. Geophys.*, **59**, e2020RG000708, <https://doi.org/10.1029/2020RG000708>.
- Barnes, E. A., S. M. Samarasinghe, I. Ebert-Uphoff, and J. C. Furtado, 2019: Tropospheric and stratospheric causal pathways between the MJO and NAO. *J. Geophys. Res. Atmos.*, **124**, 9356–9371, <https://doi.org/10.1029/2019JD031024>.
- Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2).
- Birner, T., and J. R. Albers, 2017: Sudden stratospheric warmings and anomalous upward wave activity flux. *SOLA*, **13A**, 8–12, <https://doi.org/10.2151/sola.13A-002>.
- Bland, J., S. Gray, J. Methven, and R. Forbes, 2021: Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts. *Quart. J. Roy. Meteor. Soc.*, **147**, 3878–3898, <https://doi.org/10.1002/qj.4150>.
- Butchart, N., 2014: The Brewer–Dobson circulation. *Rev. Geophys.*, **52**, 157–184, <https://doi.org/10.1002/2013RG000448>.
- Butler, A. H., J. P. Sjöberg, D. J. Seidel, and K. H. Rosenlof, 2017: A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, **9**, 63–76, <https://doi.org/10.5194/essd-9-63-2017>.
- , and Coauthors, 2019: Sub-seasonal predictability and the stratosphere. *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*, A. W. Robertson and F. Vitart, Eds., Elsevier, 223–241, <https://doi.org/10.1016/B978-0-12-811714-9.00011-5>.
- , Z. D. Lawrence, S. H. Lee, S. P. Lillo, and C. S. Long, 2020: Differences between the 2018 and 2019 stratospheric polar vortex split events. *Quart. J. Roy. Meteor. Soc.*, **146**, 3503–3521, <https://doi.org/10.1002/qj.3858>.
- Cai, M., Y. Yu, Y. Deng, H. M. van den Dool, R. Ren, S. Saha, X. Wu, and J. Huang, 2016: Feeling the pulse of the stratosphere: An emerging opportunity for predicting continental-scale cold-air outbreaks 1 month in advance. *Bull. Amer. Meteor. Soc.*, **97**, 1475–1489, <https://doi.org/10.1175/BAMS-D-14-00287.1>.
- Charlton-Perez, A. J., and Coauthors, 2013: On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *J. Geophys. Res. Atmos.*, **118**, 2494–2505, <https://doi.org/10.1002/jgrd.50125>.
- , J. Bröcker, A. Y. Karpechko, S. H. Lee, M. Sigmond, and I. R. Simpson, 2021: A minimal model to diagnose the contribution of the stratosphere to tropospheric forecast skill. *J. Geophys. Res. Atmos.*, **126**, e2021JD035504, <https://doi.org/10.1029/2021JD035504>.
- Chwat, D., C. I. Garfinkel, W. Chen, and J. Rao, 2022: Which sudden stratospheric warming events are most predictable? *J. Geophys. Res. Atmos.*, **127**, e2022JD037521, <https://doi.org/10.1029/2022JD037521>.
- Cikanek, H., and Coauthors, 2019: A vision paper for the Earth Prediction Innovation Center (EPIC): Version 5.0. NOAA Institutional Repository, 13 pp., <https://doi.org/10.25923/TW4Q-5998>.
- Collimore, C. C., D. W. Martin, M. H. Hitchman, A. Huesmann, and D. E. Waliser, 2003: On the relationship between the QBO and tropical deep convection. *J. Climate*, **16**, 2552–2568, [https://doi.org/10.1175/1520-0442\(2003\)016<2552:OTRBTQ>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<2552:OTRBTQ>2.0.CO;2).
- de la Cámara, A., and F. Lott, 2015: A parameterization of gravity waves emitted by fronts and jets. *Geophys. Res. Lett.*, **42**, 2071–2078, <https://doi.org/10.1002/2015GL063298>.
- , J. R. Albers, T. Birner, R. R. Garcia, P. Hitchcock, D. E. Kinnison, and A. K. Smith, 2017: Sensitivity of sudden stratospheric warmings to previous stratospheric conditions. *J. Atmos. Sci.*, **74**, 2857–2877, <https://doi.org/10.1175/JAS-D-17-0136.1>.
- , T. Birner, and J. R. Albers, 2019: Are sudden stratospheric warmings preceded by anomalous tropospheric wave activity? *J. Climate*, **32**, 7173–7189, <https://doi.org/10.1175/JCLI-D-19-0269.1>.
- Densmore, C. R., E. R. Sanabia, and B. S. Barrett, 2019: QBO influence on MJO amplitude over the Maritime Continent: Physical mechanisms and seasonality. *Mon. Wea. Rev.*, **147**, 389–406, <https://doi.org/10.1175/MWR-D-18-0158.1>.
- Domeisen, D. I. V., and A. H. Butler, 2020: Stratospheric drivers of extreme events at the Earth's surface. *Commun. Earth Environ.*, **1**, 59, <https://doi.org/10.1038/s43247-020-00060-z>.

- , C. I. Garfinkel, and A. H. Butler, 2019: The teleconnection of El Niño–Southern Oscillation to the stratosphere. *Rev. Geophys.*, **57**, 5–47, <https://doi.org/10.1029/2018RG000596>.
- , and Coauthors, 2020a: The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.*, **125**, e2019JD030920, <https://doi.org/10.1029/2019JD030920>.
- , and Coauthors, 2020b: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere–troposphere coupling. *J. Geophys. Res. Atmos.*, **125**, e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Dunn-Sigouin, E., and T. A. Shaw, 2015: Comparing and contrasting extreme stratospheric events, including their coupling to the tropospheric circulation. *J. Geophys. Res. Atmos.*, **120**, 1374–1390, <https://doi.org/10.1002/2014JD022116>.
- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, <https://doi.org/10.1002/met.25>.
- Farrar, M., 2021: Service change notice: Upgrade NCEP Global Forecast Systems (GFS) to v16. National Centers for Environmental Prediction, 13 pp., https://www.weather.gov/media/notification/pdf2/scn21-20_gfsv16.0_aac.pdf.
- Fujiwara, M., G. L. Manney, L. J. Gray, and J. S. Wright, 2022: SPARC Reanalysis Intercomparison Project (S-RIP) final report. SPARC Rep. 10, WCRP-6/2021, 612 pp., <https://doi.org/10.17874/800dee57d13>.
- Garfinkel, C. I., and D. L. Hartmann, 2011: The influence of the quasi-biennial oscillation on the troposphere in winter in a hierarchy of models. Part II: Perpetual winter WACCM runs. *J. Atmos. Sci.*, **68**, 2026–2041, <https://doi.org/10.1175/2011JAS3702.1>.
- , T. A. Shaw, D. L. Hartmann, and D. W. Waugh, 2012: Does the Holton–Tan mechanism explain how the quasi-biennial oscillation modulates the Arctic polar vortex? *J. Atmos. Sci.*, **69**, 1713–1733, <https://doi.org/10.1175/JAS-D-11-0209.1>.
- , E. P. Gerber, O. Shamir, J. Rao, M. Jucker, I. White, and N. Paldor, 2022: A QBO cookbook: Sensitivity of the quasi-biennial oscillation to resolution, resolved waves, and parameterized gravity waves. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002568, <https://doi.org/10.1029/2021MS002568>.
- Gerber, E. P., and P. Martineau, 2018: Quantifying the variability of the annular modes: Reanalysis uncertainty vs. sampling uncertainty. *Atmos. Chem. Phys.*, **18**, 17 099–17 117, <https://doi.org/10.5194/acp-18-17099-2018>.
- , and Coauthors, 2010: Stratosphere–troposphere coupling and annular mode variability in chemistry–climate models. *J. Geophys. Res.*, **115**, D00M06, <https://doi.org/10.1029/2009JD013770>.
- Gottschalck, J., and Coauthors, 2010: A framework for assessing operational Madden–Julian Oscillation forecasts: A CLIVAR MJO working group project. *Bull. Amer. Meteor. Soc.*, **91**, 1247–1258, <https://doi.org/10.1175/2010BAMS2816.1>.
- Gray, L. J., J. A. Anstey, Y. Kawatani, H. Lu, S. Osprey, and V. Schenzinger, 2018: Surface impacts of the quasi biennial oscillation. *Atmos. Chem. Phys.*, **18**, 8227–8247, <https://doi.org/10.5194/acp-18-8227-2018>.
- , H. Lu, M. J. Brown, J. R. Knight, and M. B. Andrews, 2022: Mechanisms of influence of the semi-annual oscillation on stratospheric sudden warmings. *Quart. J. Roy. Meteor. Soc.*, **148**, 1223–1241, <https://doi.org/10.1002/qj.4256>.
- Guan, H., and Coauthors, 2022: GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Mon. Wea. Rev.*, **150**, 647–665, <https://doi.org/10.1175/MWR-D-21-0245.1>.
- Hall, R. J., D. M. Mitchell, W. J. M. Seviour, and C. J. Wright, 2021: Tracking the stratosphere-to-surface impact of sudden stratospheric warmings. *J. Geophys. Res. Atmos.*, **126**, e2020JD033881, <https://doi.org/10.1029/2020JD033881>.
- Hamill, T. M., and Coauthors, 2022: The reanalysis for the global ensemble forecast system, version 12. *Mon. Wea. Rev.*, **150**, 59–79, <https://doi.org/10.1175/MWR-D-21-0023.1>.
- Hamilton, K., R. J. Wilson, and R. S. Hemler, 1999: Middle atmosphere simulated with high vertical and horizontal resolution versions of a GCM: Improvements in the cold pole bias and generation of a QBO-like oscillation in the tropics. *J. Atmos. Sci.*, **56**, 3829–3846, [https://doi.org/10.1175/1520-0469\(1999\)056<3829:MASWHV>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<3829:MASWHV>2.0.CO;2).
- Hannachi, A., D. Mitchell, L. Gray, and A. Charlton-Perez, 2011: On the use of geometric moments to examine the continuum of sudden stratospheric warmings. *J. Atmos. Sci.*, **68**, 657–674, <https://doi.org/10.1175/2010JAS3585.1>.
- Harris, L., X. Chen, W. Putman, L. Zhou, and J.-H. Chen, 2021: A scientific description of the GFDL finite-volume cubed-sphere dynamical core. Tech. Memo. GFDL2021001, NOAA, 109 pp., <https://doi.org/10.25923/6NHS-5897>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hitchcock, P., 2019: On the value of reanalyses prior to 1979 for dynamical studies of stratosphere–troposphere coupling. *Atmos. Chem. Phys.*, **19**, 2749–2764, <https://doi.org/10.5194/acp-19-2749-2019>.
- Hitchman, M., S. Yoden, P. H. Haynes, V. Kumar, and S. Tegtmeier, 2021: An observational history of the direct influence of the stratospheric quasi-biennial oscillation on the tropical and subtropical upper troposphere and lower stratosphere. *J. Meteor. Soc. Japan*, **99**, 239–267, <https://doi.org/10.2151/jmsj.2021-012>.
- Hogan, R., and Coauthors, 2017: Radiation in numerical weather prediction. ECMWF Tech. Memo. 816, 51 pp., <https://doi.org/10.21957/2bd5dkj8x>.
- Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram a geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting*, **2**, 285–293, [https://doi.org/10.1016/0169-2070\(86\)90048-8](https://doi.org/10.1016/0169-2070(86)90048-8).
- Hurrell, J. W., and H. Van Loon, 1997: Decadal variations in climate associated with the North Atlantic Oscillation. *Climatic Change*, **36**, 301–326, <https://doi.org/10.1023/A:1005314315270>.
- , Y. Kushnir, and M. Visbeck, 2001: The North Atlantic Oscillation. *Science*, **291**, 603–605, <https://doi.org/10.1126/science.1058761>.
- Jacobs, N. A., 2021: Open innovation and the case for community model development. *Bull. Amer. Meteor. Soc.*, **102**, E2002–E2011, <https://doi.org/10.1175/BAMS-D-21-0030.1>.
- Jiménez-Esteve, B., and D. I. V. Domeisen, 2018: The tropospheric pathway of the ENSO–North Atlantic teleconnection. *J. Climate*, **31**, 4563–4584, <https://doi.org/10.1175/JCLI-D-17-0716.1>.
- Karpechko, A. Y., P. Hitchcock, D. H. W. Peters, and A. Schneidereit, 2017: Predictability of downward propagation of major sudden stratospheric warmings. *Quart. J. Roy. Meteor. Soc.*, **143**, 1459–1470, <https://doi.org/10.1002/qj.3017>.
- Kawatani, Y., K. Hamilton, L. J. Gray, S. M. Osprey, S. Watanabe, and Y. Yamashita, 2019: The effects of a well-resolved stratosphere on the simulated boreal winter circulation in a climate model. *J. Atmos. Sci.*, **76**, 1203–1226, <https://doi.org/10.1175/JAS-D-18-0206.1>.

- Kidston, J., A. A. Scaife, S. C. Hardiman, D. M. Mitchell, N. Butchart, M. P. Baldwin, and L. J. Gray, 2015: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nat. Geosci.*, **8**, 433–440, <https://doi.org/10.1038/ngeo2424>.
- Kim, H., P. J. Webster, V. E. Toma, and D. Kim, 2014: Predictability and prediction skill of the MJO in two operational forecasting systems. *J. Climate*, **27**, 5364–5378, <https://doi.org/10.1175/JCLI-D-13-00480.1>.
- , J. H. Richter, and Z. Martin, 2019: Insignificant QBO-MJO prediction skill relationship in the SubX and S2S subseasonal reforecasts. *J. Geophys. Res. Atmos.*, **124**, 12655–12666, <https://doi.org/10.1029/2019JD031416>.
- , J. M. Caron, J. H. Richter, and I. R. Simpson, 2020: The lack of QBO-MJO connection in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2020GL087295, <https://doi.org/10.1029/2020GL087295>.
- Klotzbach, P., S. Abhik, H. H. Hendon, M. Bell, C. Lucas, A. G. Marshall, and E. C. J. Oliver, 2019: On the emerging relationship between the stratospheric quasi-biennial oscillation and the Madden-Julian Oscillation. *Sci. Rep.*, **9**, 2981, <https://doi.org/10.1038/s41598-019-40034-6>.
- Lawrence, Z. D., and Coauthors, 2022: Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems. *Wea. Climate Dyn.*, **3**, 977–1001, <https://doi.org/10.5194/wcd-3-977-20.2>.
- Lim, Y., and S.-W. Son, 2022: QBO wind influence on MJO-induced temperature anomalies in the upper troposphere and lower stratosphere in an idealized model. *J. Atmos. Sci.*, **79**, 2219–2228, <https://doi.org/10.1175/JAS-D-21-0296.1>.
- , —, A. G. Marshall, H. H. Hendon, and K.-H. Seo, 2019: Influence of the QBO on MJO prediction skill in the subseasonal-to-seasonal prediction models. *Climate Dyn.*, **53**, 1681–1695, <https://doi.org/10.1007/s00382-019-04719-y>.
- Limpasuvan, V., D. L. Hartmann, D. W. J. Thompson, K. Jeev, and Y. L. Yung, 2005: Stratosphere-troposphere evolution during polar vortex intensification. *J. Geophys. Res.*, **110**, D24101, <https://doi.org/10.1029/2005JD006302>.
- Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDG>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDG>2.0.CO;2).
- Long, C. S., M. Fujiwara, S. Davis, D. M. Mitchell, and C. J. Wright, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14 593–14 629, <https://doi.org/10.5194/acp-17-14593-2017>.
- Manney, G. L., and Z. D. Lawrence, 2016: The major stratospheric final warming in 2016: Dispersal of vortex air and termination of Arctic chemical ozone loss. *Atmos. Chem. Phys.*, **16**, 15 371–15 396, <https://doi.org/10.5194/acp-16-15371-2016>.
- Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101**, E608–E625, <https://doi.org/10.1175/BAMS-D-18-0326.1>.
- Marlton, G., and Coauthors, 2021: Using a network of temperature lidars to identify temperature biases in the upper stratosphere in ECMWF reanalyses. *Atmos. Chem. Phys.*, **21**, 6079–6092, <https://doi.org/10.5194/acp-21-6079-2021>.
- Marshall, A. G., H. H. Hendon, S.-W. Son, and Y. Lim, 2017: Impact of the quasi-biennial oscillation on predictability of the Madden-Julian oscillation. *Climate Dyn.*, **49**, 1365–1377, <https://doi.org/10.1007/s00382-016-3392-0>.
- Martin, Z., C. Orbe, S. Wang, and A. Sobel, 2021: The MJO–QBO relationship in a GCM with stratospheric nudging. *J. Climate*, **34**, 4603–4624, <https://doi.org/10.1175/JCLI-D-20-0636.1>.
- Matthewman, N. J., J. G. Esler, A. J. Charlton-Perez, and L. M. Polvani, 2009: A new look at stratospheric sudden warmings. Part III: Polar vortex evolution and vertical structure. *J. Climate*, **22**, 1566–1585, <https://doi.org/10.1175/2008JCLI2365.1>.
- Matthias, V., and M. Kretschmer, 2020: The influence of stratospheric wave reflection on North American cold spells. *Mon. Wea. Rev.*, **148**, 1675–1690, <https://doi.org/10.1175/MWR-D-19-0339.1>.
- McLandress, C., T. G. Shepherd, S. Polavarapu, and S. R. Beagley, 2012: Is missing orographic gravity wave drag near 60°S the cause of the stratospheric zonal wind biases in chemistry–climate models? *J. Atmos. Sci.*, **69**, 802–818, <https://doi.org/10.1175/JAS-D-11-0159.1>.
- Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, E869–E896, <https://doi.org/10.1175/BAMS-D-19-0037.1>.
- Messori, G., M. Kretschmer, S. H. Lee, and V. Wendt, 2022: Stratospheric wave reflection events modulate North American weather regimes and cold spells. *Wea. Climate Dyn.*, **3**, 1215–2022, <https://doi.org/10.5194/wcd-2022-18>.
- Miller, D. E., and Z. Wang, 2019: Assessing seasonal predictability sources and windows of high predictability in the Climate Forecast System, version 2. *J. Climate*, **32**, 1307–1326, <https://doi.org/10.1175/JCLI-D-18-0389.1>.
- Millin, O. T., J. C. Furtado, and J. B. Basara, 2022: Characteristics, evolution, and formation of cold air outbreaks in the Great Plains of the United States. *J. Climate*, **35**, 4585–4602, <https://doi.org/10.1175/JCLI-D-21-0772.1>.
- Newman, P. A., E. R. Nash, and J. E. Rosenfield, 2001: What controls the temperature of the Arctic stratosphere during the spring? *J. Geophys. Res.*, **106**, 19 999–20 010, <https://doi.org/10.1029/2000JD000061>.
- Pegion, K., and Coauthors, 2019: The subseasonal experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, <https://doi.org/10.1175/BAMS-D-18-0270.1>.
- Perlwitz, J., and N. Harnik, 2003: Observational evidence of a stratospheric influence on the troposphere by planetary wave reflection. *J. Climate*, **16**, 3011–3026, [https://doi.org/10.1175/1520-0442\(2003\)016<3011:OEAOASI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3011:OEAOASI>2.0.CO;2).
- , and —, 2004: Downward coupling between the stratosphere and troposphere: The relative roles of wave and zonal mean processes. *J. Climate*, **17**, 4902–4909, <https://doi.org/10.1175/JCLI-3247.1>.
- Plougonven, R., and F. Zhang, 2014: Internal gravity waves from atmospheric jets and fronts. *Rev. Geophys.*, **52**, 33–76, <https://doi.org/10.1002/2012RG000419>.
- Polichtchouk, I., T. Stockdale, P. Bechtold, M. Diamantakis, S. Malardel, I. Sandu, F. Vána, and N. Wedi, 2019: Control on stratospheric temperature in IFS: Resolution and vertical advection. ECMWF Tech. Memo. 847, ECMWF, 38 pp., <https://doi.org/10.21957/cz3t12t7e>.
- , and Coauthors, 2021: Stratospheric modelling and assimilation. ECMWF Tech. Memo. 877, ECMWF, 63 pp., <https://doi.org/10.21957/25hegfoq>.

- Polvani, L. M., and D. W. Waugh, 2004: Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *J. Climate*, **17**, 3548–3554, [https://doi.org/10.1175/1520-0442\(2004\)017<3548:UWAFAA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2).
- Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Rao, J., and C. I. Garfinkel, 2020: Arctic ozone loss in March 2020 and its seasonal prediction in CFSv2: A comparative study with the 1997 and 2011 cases. *J. Geophys. Res. Atmos.*, **125**, e2020JD033524, <https://doi.org/10.1029/2020JD033524>.
- , R. Ren, H. Chen, X. Liu, Y. Yu, J. Hu, and Y. Zhou, 2019: Predictability of stratospheric sudden warmings in the Beijing climate center forecast system with statistical error corrections. *J. Geophys. Res. Atmos.*, **124**, 8385–8400, <https://doi.org/10.1029/2019JD030900>.
- , C. I. Garfinkel, and I. P. White, 2020: Predicting the downward and surface influence of the February 2018 and January 2019 sudden stratospheric warming events in subseasonal to seasonal (S2S) models. *J. Geophys. Res. Atmos.*, **125**, e2019JD031919, <https://doi.org/10.1029/2019JD031919>.
- , —, and —, 2021a: Development of the extratropical response to the stratospheric quasi-biennial oscillation. *J. Climate*, **34**, 7239–7255, <https://doi.org/10.1175/JCLI-D-20-0960.1>.
- , —, T. Wu, Y. Lu, Q. Lu, and Z. Liang, 2021b: The January 2021 sudden stratospheric warming and its prediction in subseasonal to seasonal models. *J. Geophys. Res. Atmos.*, **126**, e2021JD035057, <https://doi.org/10.1029/2021JD035057>.
- , —, —, and M. Chu, 2022: Mean state of the Northern Hemisphere stratospheric polar vortex in three generations of CMIP models. *J. Climate*, **35**, 4603–4625, <https://doi.org/10.1175/JCLI-D-21-0694.1>.
- , T. Wu, C. I. Garfinkel, J. Luo, Y. Lu, M. Chu, and J. Hu, 2023: Impact of the initial stratospheric polar vortex state on East Asian spring rainfall prediction in seasonal forecast models. *Climate Dyn.*, **60**, 4111–4131, <https://doi.org/10.1007/s00382-022-06551-3>.
- Rashid, H. A., H. H. Hendon, M. C. Wheeler, and O. Alves, 2011: Prediction of the Madden-Julian oscillation with the POAMA dynamical prediction system. *Climate Dyn.*, **36**, 649–661, <https://doi.org/10.1007/s00382-010-0754-x>.
- Richter, J. H., A. Solomon, and J. T. Bacmeister, 2014: Effects of vertical resolution and nonorographic gravity wave drag on the simulated climate in the Community Atmosphere Model, version 5. *J. Adv. Model. Earth Syst.*, **6**, 357–383, <https://doi.org/10.1002/2013MS000303>.
- , J. A. Anstey, N. Butchart, Y. Kawatani, G. A. Meehl, S. Osprey, and I. R. Simpson, 2020: Progress in simulating the quasi-biennial oscillation in CMIP models. *J. Geophys. Res. Atmos.*, **125**, e2019JD032362, <https://doi.org/10.1029/2019JD032362>.
- Riddle, E. E., A. H. Butler, J. C. Furtado, J. L. Cohen, and A. Kumar, 2013: CFSv2 ensemble prediction of the wintertime Arctic Oscillation. *Climate Dyn.*, **41**, 1099–1116, <https://doi.org/10.1007/s00382-013-1850-5>.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Sakaeda, N., J. Dias, and G. N. Kiladis, 2020: The unique characteristics and potential mechanisms of the MJO-QBO relationship. *J. Geophys. Res. Atmos.*, **125**, e2020JD033196, <https://doi.org/10.1029/2020JD033196>.
- Scaife, A. A., and Coauthors, 2022: Long-range prediction and the stratosphere. *Atmos. Chem. Phys.*, **22**, 2601–2623, <https://doi.org/10.5194/acp-22-2601-2022>.
- Schwartz, C., C. I. Garfinkel, P. Yadav, W. Chen, and D. I. V. Domeisen, 2022: Stationary wave biases and their effect on upward troposphere–stratosphere coupling in sub-seasonal prediction models. *Wea. Climate Dyn.*, **3**, 679–692, <https://doi.org/10.5194/wcd-3-679-2022>.
- Serva, F., and Coauthors, 2022: The impact of the QBO on the region of the tropical tropopause in QBOi models: Present-day simulations. *Quart. J. Roy. Meteor. Soc.*, **148**, 1945–1964, <https://doi.org/10.1002/qj.4287>.
- Seviour, W. J. M., D. M. Mitchell, and L. J. Gray, 2013: A practical method to identify displaced and split stratospheric polar vortex events. *Geophys. Res. Lett.*, **40**, 5268–5273, <https://doi.org/10.1002/grl.50927>.
- Shaw, T. A., and J. Perlwitz, 2013: The life cycle of Northern Hemisphere downward wave coupling between the stratosphere and troposphere. *J. Climate*, **26**, 1745–1763, <https://doi.org/10.1175/JCLI-D-12-00251.1>.
- , —, and O. Weiner, 2014: Troposphere-stratosphere coupling: Links to North Atlantic weather and climate, including their representation in CMIP5 models. *J. Geophys. Res. Atmos.*, **119**, 5864–5880, <https://doi.org/10.1002/2013JD021191>.
- Shepherd, T., K. Semeniuk, and J. N. Koshyk, 1996: Sponge layer feedbacks in middle-atmosphere models. *J. Geophys. Res.*, **101**, 23 447–23 464, <https://doi.org/10.1029/96JD01994>.
- , I. Polichtchouk, R. Hogan, and A. J. Simmons, 2018: Report on stratosphere task force. ECMWF Tech. Memo. 824, ECMWF, 34 pp., <https://doi.org/10.21957/0vkp0t1xx>.
- Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 98–102, <https://doi.org/10.1038/ngeo1698>.
- Simmons, A., and Coauthors, 2020: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. ECMWF Tech. Memo. 859, ECMWF, 40 pp., <https://doi.org/10.21957/rcxqfmg0>.
- Simpson, I. R., P. Hitchcock, T. G. Shepherd, and J. F. Scinocca, 2011: Stratospheric variability and tropospheric annular-mode timescales. *Geophys. Res. Lett.*, **38**, L20806, <https://doi.org/10.1029/2011GL049304>.
- Son, S.-W., Y. Lim, C. Yoo, H. H. Hendon, and J. Kim, 2017: Stratospheric control of the Madden-Julian Oscillation. *J. Climate*, **30**, 1909–1922, <https://doi.org/10.1175/JCLI-D-16-0620.1>.
- , H. Kim, K. Song, S.-W. Kim, P. Martineau, Y.-K. Hyun, and Y. Kim, 2020: Extratropical prediction skill of the sub-seasonal-to-seasonal (S2S) prediction models. *J. Geophys. Res. Atmos.*, **125**, e2019JD031273, <https://doi.org/10.1029/2019JD031273>.
- Sun, L., J. Perlwitz, J. H. Richter, M. P. Hoerling, and J. W. Hurrell, 2020: Attribution of NAO predictive skill beyond 2 weeks in boreal winter. *Geophys. Res. Lett.*, **47**, e2020GL090451, <https://doi.org/10.1029/2020GL090451>.
- Taguchi, M., 2016: Connection of predictability of major stratospheric sudden warmings to polar vortex geometry. *Atmos. Sci. Lett.*, **17**, 33–38, <https://doi.org/10.1002/asl.595>.
- , 2018: Comparison of subseasonal-to-seasonal model forecasts for major stratospheric sudden warmings. *J. Geophys. Res. Atmos.*, **123**, 10 231–10 247, <https://doi.org/10.1029/2018JD028755>.
- Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J.*

- Climate*, **13**, 1000–1016, [https://doi.org/10.1175/1520-0442\(2000\)013<1000:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2).
- Tripathi, O. P., A. Charlton-Perez, M. Sigmond, and F. Vitart, 2015: Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environ. Res. Lett.*, **10**, 104007, <https://doi.org/10.1088/1748-9326/10/10/104007>.
- Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Wallace, J. M., R. L. Panetta, and J. Estberg, 1993: Representation of the equatorial stratospheric quasi-biennial oscillation in EOF phase space. *J. Atmos. Sci.*, **50**, 1751–1762, [https://doi.org/10.1175/1520-0469\(1993\)050<1751:ROTESQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<1751:ROTESQ>2.0.CO;2).
- Wang, S., M. K. Tippett, A. H. Sobel, Z. K. Martin, and F. Vitart, 2019: Impact of the QBO on prediction and predictability of the MJO convection. *J. Geophys. Res. Atmos.*, **124**, 11 766–11 782, <https://doi.org/10.1029/2019JD030575>.
- Waugh, D. W., 1997: Elliptical diagnostics of stratospheric polar vortices. *Quart. J. Roy. Meteor. Soc.*, **123**, 1725–1748, <https://doi.org/10.1002/qj.49712354213>.
- , and L. M. Polvani, 2010: Stratospheric polar vortices. *The Stratosphere: Dynamics, Transport, and Chemistry*, L. M. Polvani, A. H. Sobel, and D. W. Waugh, Eds., Amer. Geophys. Union, 43–57.
- , A. H. Sobel, and L. M. Polvani, 2017: What is the polar vortex and how does it influence weather? *Bull. Amer. Meteor. Soc.*, **98**, 37–44, <https://doi.org/10.1175/BAMS-D-15-00212.1>.
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- White, I. P., C. I. Garfinkel, E. P. Gerber, M. Jucker, V. Aquila, and L. D. Oman, 2019: The downward influence of sudden stratospheric warmings: Association with tropospheric precursors. *J. Climate*, **32**, 85–108, <https://doi.org/10.1175/JCLI-D-18-0053.1>.
- , —, —, —, P. Hitchcock, and J. Rao, 2020: The generic nature of the tropospheric response to sudden stratospheric warmings. *J. Climate*, **33**, 5589–5610, <https://doi.org/10.1175/JCLI-D-19-0697.1>.
- Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Sciences*. 4th ed. Elsevier, 840 pp.
- Xue, Y., and Coauthors, 2021: Development of a coupled subseasonal-to-seasonal prediction model using community-based unified forecast system for NCEP operations. *EGU 2021 Fall Meeting*, online, European Geosciences Union, EGU Abstract EGU21-5722, <https://doi.org/10.5194/egusphere-egu21-5722>.
- Yoo, C., and S.-W. Son, 2016: Modulation of the boreal winter-time Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophys. Res. Lett.*, **43**, 1392–1398, <https://doi.org/10.1002/2016GL067762>.
- Yu, Y., M. Cai, C. Shi, R. Yan, and J. Rao, 2019: Sub-seasonal prediction skill for the stratospheric meridional mass circulation variability in CFSv2. *Climate Dyn.*, **53**, 631–650, <https://doi.org/10.1007/s00382-018-04609-9>.
- Zhang, Q., C.-S. Shin, H. van den Dool, and M. Cai, 2013: CFSv2 prediction skill of stratospheric temperature anomalies. *Climate Dyn.*, **41**, 2231–2249, <https://doi.org/10.1007/s00382-013-1907-5>.
- Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System version 12. *Wea. Forecasting*, **37**, 1069–1084, <https://doi.org/10.1175/WAF-D-21-0112.1>.