Climate Dynamics
https://doi.org/10.1007/s00382-018-04609-9

Sub-seasonal prediction skill for the stratospheric meridional mass circulation variability in CFSv2

Yueyue Yu[1] · Ming Cai[2] · Chunhua Shi[1] · Ruikai Yan[2] · Jian Rao[1]

Received: 26 July 2018 / Accepted: 26 December 2018

Abstract
This study evaluates the prediction skill for the stratospheric mass circulation variability in winter (November–March) in the Climate Forecast System, version 2 (CFSv2), from 2011 to 2018. Three stratospheric mass circulation indices measuring meridional mass transport into the polar stratosphere by the total flow (ST60N), wavenumber-1 (ST60N_W1), and wavenumber-2 waves (ST60N_W2) are considered. The variability of ST60N is mainly contributed by ST60N_W1 and ST60N_W2, and these indices are good indicators of the timing and locations of the continental-scale cold-air outbreaks (CAOs) at mid-latitudes. Thus, their potentially useful prediction skill can be utilized to make sub-seasonal forecasts of CAOs in a dynamical and statistical hybrid paradigm. Systematic forecast bias is found in both the 7-year averaged winter mean and seasonal cycle, which is tied to the overestimation of damping in amplitude and westward tilting variations of total waves and difficulties in forecasting the exact contributions from different spatial scales of waves. The intraseasonal variations of stratospheric mass circulation indices, with the systematic forecast bias corrected, can be modestly predicted at a forecast lead time of about 20 days, in terms of both the anomaly value and timing of negative and positive peak events. The sub-seasonal predictability of the ST60N value mainly comes from the ST60N_W1, while the predictability of the timing of positive and negative peaks comes from the ST60N_W2. The 20-day prediction limit of the stratospheric mass circulation indices is mainly due to the 2-week limit of the CFSv2 model in predicting the variability of anomalous wave tilt angle, whereas the prediction limit of the wave amplitude anomaly can exceed 50 days.

Keywords Subseasonal prediction skill · Stratosphere · Meridional mass circulation · CFSv2

1 Introduction

Subseasonal forecasts issued by most operational forecast centers, such as the Climate Prediction Center (CPC) of the US National Weather Service and the UK Met Office, are made by a combination of empirical and dynamical prediction tools. On the one hand, dynamical tools derive forecasts by integrating global atmospheric models with prescribed ocean surface temperature or coupled ocean–atmosphere global models; while on the other hand, empirical models are derived purely from statistical analysis that either empirically predicts the dominant climate variability modes or relates particular phases of these modes to anomalies via their statistical relationships. Stratospheric variability is an important source of prediction skill to extend the prediction limit of tropospheric forecasts both dynamically and empirically.

As for the role of stratospheric variability as the dynamical tool for tropospheric forecasts, numerous numerical experiments have been conducted to quantify the impact of stratospheric variability on the troposphere in models, as summarized by Tripathi et al. (2015) and references therein. Methods that directly assess the overall impact of the stratosphere on tropospheric forecast skill include degrading the representation of the stratosphere by restricting the stratospheric resolution and raising/lowering the top level of the model (Jung and Leutbecher 2007; Marshall

✉ Yueyue Yu
yyu4@fsu.edu

[1] Key Laboratory of Meteorological Disaster, Ministry of Education (KLME), Joint International Research Laboratory of Climate and Environment Change (ILCEC), Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] Department of Earth, Ocean and Atmospheric Sciences, Florida State University, Tallahassee, FL 32304, USA

Published online: 23 February 2019

and Scaife 2010). Other experimental designs are more suited to assessing the importance for tropospheric forecast skill of the evolution of the stratospheric state in different regions and at different times. Both perturbation techniques (adding additional artificial forcing to the stratosphere) and relaxation techniques (damping the stratospheric state towards observations) have been used to quantify the role of the stratosphere in recent extreme winter seasons in the Northern Hemisphere. All such experiments suggest that improving the stratospheric representation and reducing stratospheric model error should lead to improved model forecasts for tropospheric circulation and weather. Forecasting studies, however, suggest that in the Northern Hemisphere this enhanced predictability in models arises only following certain sudden warmings and is not uniform in time (Mukougawa et al. 2009).

Therefore, we also need the other way to utilize the stratospheric signal for tropospheric forecasts—constructing an empirical model based on the stratosphere-troposphere coupling relationship. The "downward impact" of the low-frequency variability of the stratospheric northern annular mode (NAM) or stratospheric polar vortex intensity on the Arctic oscillation and associated extreme weather has been recognized as a new opportunity for subseasonal climate predictions in winter seasons, since the stratospheric signal provides information at a long lead-time (0–60 days) with respect to anomalous surface weather regimes (Baldwin and Dunkerton 2001; Thompson and Wallace 2001, 2002; Baldwin et al. 2003; Polvani and Waugh 2004). However, such a long-lead-time relationship between stratospheric anomalies and the troposphere is not always found, because of both the large internal variability in the troposphere (Gerber et al. 2009) and specific properties of the stratosphere, including its persistence and the depth of initial warming (Hitchcock and Simpson 2014; Gerber et al. 2009).

Recently, the quasi-simultaneous statistical relationship between continental-scale CAOs and stratospheric circulation anomalies has drawn attention. Kolstad and Charlton-Perez (2010), Mitchell et al. (2013), Kidston et al. (2015), and Lehtonen and Karpechko (2016) presented specific temperature anomaly patterns associated with sudden stratospheric warming (SSW) events of vortex splitting and vortex displacement types. Cai et al. (2016) found a close relationship between individual continental-scale CAO events and the pulse-like signals in the stratospheric mass circulation (denoted as "PULSE" events). The PULSE event was defined as the duration of a rapid increase in the air mass transported into the polar stratosphere (upper layers above 400 K). These PULSE events occur every 2 weeks—almost as frequently as CAO events. There is a higher probability of CAO occurrence over the mid-latitude regions of Eurasia and North America within the 1–2 weeks before and after the peak dates of PULSE events. A follow-up study by Yu et al. (2018b, c) demonstrated that the pattern of cold temperature anomalies is also highly dependent on the spatial scale of the dominant waves driving the stronger stratospheric meridional mass circulation. The cold surface air temperature anomalies often occur over mid-latitude Europe 1–2 weeks before but over North America after peak time of PULSE events dominated by stronger wavenumber-1 waves, while cold anomalies tend to occur over the mid-latitudes of both North America and Eurasia in the 1–2 weeks around the peak time of PULSE events dominated by stronger wavenumber-2 waves. The link between PULSE events and surface cold events can be explained by the global meridional mass circulation theory (e.g., Townsend and Johnson 1985; Johnson 1989; Cai and Shin 2014, and references therein), i.e., that westward-tilted planetary waves with deep vertical structure drive equatorward polar cold-air mass transport at the surface and poleward warm-air mass transport aloft, thereby leading to weather at the surface tending to be accompanied by PULSE events (Yu et al. 2015a, b; Iwasaki and Mochizuki 2012; Iwasaki et al. 2014). These robust "simultaneous" relationships mentioned above yield less information ahead of time but, provided stratospheric signals can be predicted more than 2 weeks in advance or longer, such relationships can still be used to significantly improve subseasonal forecasts.

Due to the significant impacts of the stratosphere on forecasting the troposphere, the predictability of the stratosphere itself is also of considerable interest to operational forecasting centers. With more satellite data in the stratosphere to assimilate as the initial condition, and improved representation of the stratosphere by numerical models with increased vertical resolution and better resolved physical and chemical processes in the stratosphere (e.g., MacDonald 2005), we now have the capability to use numerical weather prediction models as an initial-value problem for stratospheric forecasts in the extratropics with useful skill in the subseasonal range [e.g., Tripathi et al. (2015) for all kinds of extreme polar vortex events; Christiansen (2005), Charlton and Polvani (2007) and Yoden et al. (2014), especially for major SSW events; Marshall et al. (2009) for a rare volcanic eruption event; Stan and Straus (2009), Li and Ding (2011), Zhang et al. (2013) and Noguchi et al. (2014) for zonal mean temperature, geopotential height and wind anomalies; Cai et al. (2016) for PULSE events in the meridional mass transport into the polar stratosphere; and Ngan and Eperon (2011) for global kinetic energy]. Generally speaking, the predictability limit of the circulation anomalies in the stratosphere is indeed longer than that in the troposphere. The predictability limits of the daily zonal-mean geopotential height, temperature and wind fields are more than 1 month in the stratosphere, and the those of SSW events can be 10–20 days on average. For longer forecast lead-times (i.e., longer than 1 month in advance), the prediction skills of stratospheric

circulation variables become unreliable, due to both limitations in forecasting tropospheric planetary waves and stratospheric biases in models (Haynes 2005; Sun et al. 2012).

Notably, the predictability of stratospheric mass circulation intensity as well as PULSE events has yet to receive sufficient attention from society in the same manner that the slowly-varying NAM and rarely-occurring SSW events perhaps have, but nonetheless may be a more important aspect of the stratospheric circulation. Because these signals occur more frequently and link to individual extreme weather events, especially CAOs, rather than period-mean weather conditions (Yu et al. 2018b, c). In addition, the slower variability in the stratosphere, such as the NAM, can be physically attributed to the accumulative results of several rounds of warm air mass gain in the polar stratosphere by PULSE events (Cai et al. 2016; Yu et al. 2018a). The SSW events tend to be a subset of strong and long-lasting PULSE events (Yu et al. 2018c).

Cai et al. (2016) examined the 7-day running-mean time series of meridional mass transport into the polar stratosphere above 400 K (ST60N index) forecasted by the Climate Forecast System, version 2 (CFSv2) for the winter of 2013/14 and found that CFSv2 has the ability to predict the main features of the temporal evolution of the ST60N index up to 1 month in advance. However, due to the limited access to CFSv2 real forecast data with sufficient vertical layers in the stratosphere, Cai et al. (2016) only examined one winter. Despite the overall successful weekly experimental forecasts of continental-scale CAOs in the winter season by utilizing the hybrid paradigm (Cai et al. 2016) made by a forecast team at Florida State University since September 2014 being able to provide supporting evidence for the subseasonal predictability of PULSEs, a direct and quantitative verification has yet to be carried out. Moreover, as argued in Yu et al. (2018b, c), different types of PULSE events—namely, wavenumber-1–dominated and wavenumber-2–dominated events—tend to be associated with different surface temperature patterns. This indicates that the predictability of the meridional mass transport by different wave components is also of importance for subseasonal forecasts of timing and location of CAOs via utilizing PULSE signals. Therefore, it is necessary to document the skill of CFSv2 in predicting the PULSEs in stratospheric mass circulation, as well as its main components, driven by different scales of waves, using datasets over a longer time range. From September 2011 to March 2018, seven winters of multi-layer data from CFSv2 real-time forecasts are available. This allows us to utilize these CFSv2 S2S forecast data to take a closer look at the predictability of meridional mass circulation variability in the winter season.

This paper comprises six sections. Following this introduction, Sect. 2 describes the data and the indices we define to capture the state of stratospheric meridional mass circulation at 60°N. In Sect. 3, we first examine the systematic forecast errors of the stratospheric meridional mass circulation indices; and then, in Sect. 4, evaluate the forecast skill of stratospheric mass circulation indices in terms of correlation scores and event-based probabilistic predictability. Section 5 evaluates the forecast skill of wave properties at 60°N, including both the equivalent amplitude and vertical tilt angle of total waves as well as wavenumber-1 and wavenumber-2 components, which together are the main drivers for the stratospheric mass circulation at 60°N. The final section provides conclusions.

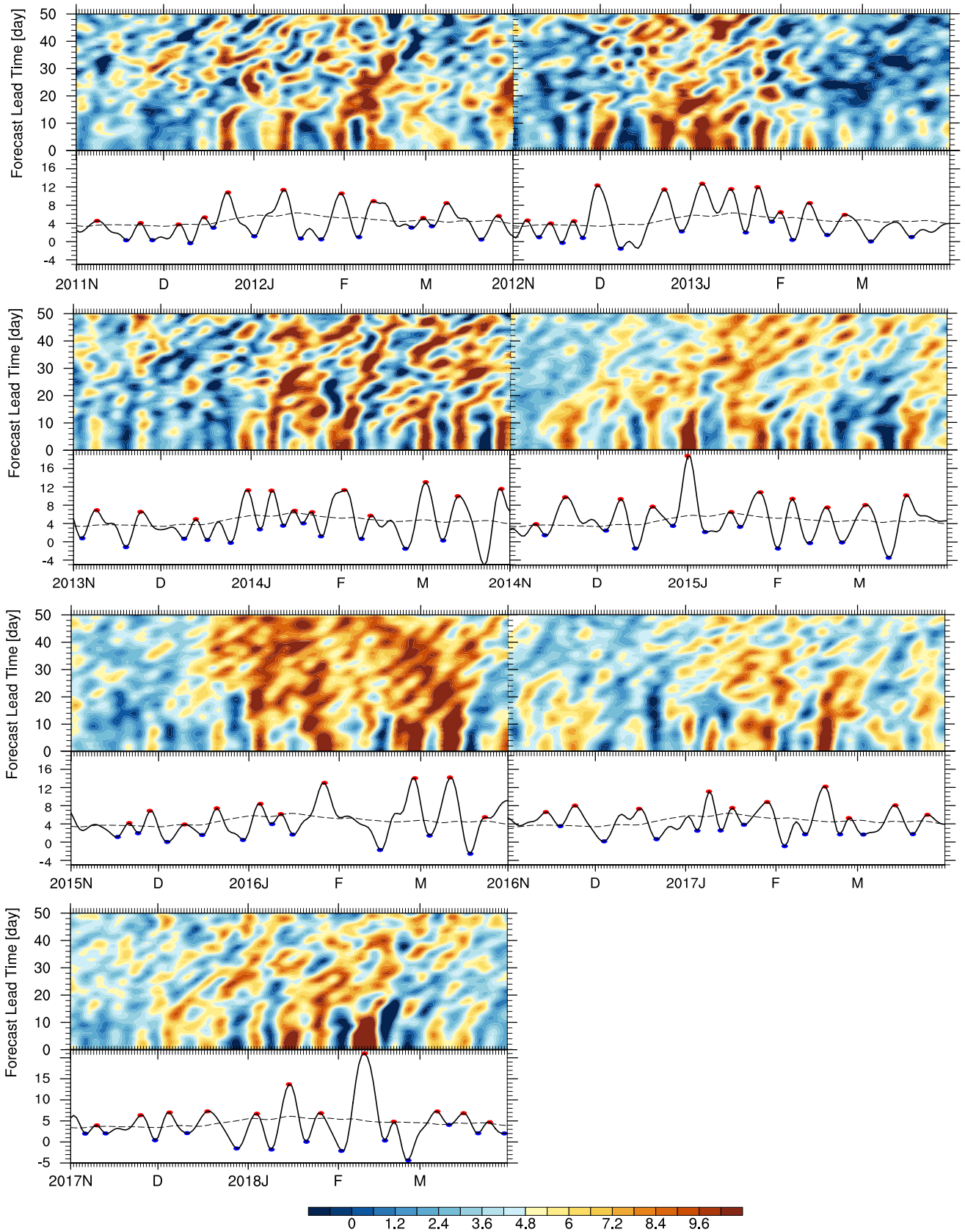
2 Data and methods

2.1 Data

The NCEP's CFSv2 became operational in March 2011 (Saha et al. 2014), as the successor to CFSv1. The CFSv2 model is a global spectral model coupled with ocean, land, and sea ice. It has a horizontal resolution approximately equivalent to 100 km (T126) and a vertical resolution of 64 sigma-pressure hybrid levels. CFSv2 operational forecast products are available from 16 model runs, four time steps (0000, 0600, 1200, and 1800 UTC) per day, and four runs per time step. The 0000 UTC forecast products contain one control run, up to 9 months, and three additional perturbed runs up to one season. While the outputs from the other three time steps also contain a 9-month control run, their three additional perturbed runs extend to 45 days. CFSv2 subseasonal forecast data have been generated for free download via http://nomads.ncep.noaa.gov/pub/data/nccf/com/cfs/prod/cfs/ since 2011, but only forecasts of the control run are stored since then. The three perturbed runs are no longer available for free download beyond 1 week. Thus, we only have forecast data from all four runs for the four winters from 2014 to 2018 (since the FSU forecast team began the experimental real-time CAO subseasonal forecasts) but one run for the 3 winters from 2011 to 2013.

To maximize the sample size of real-time forecast data, the data used in this study include daily data derived from the control run during the three winters (November–March) from November 2011 to March 2014 and the average of the four runs during the four winters from November 2014 to March 2018. However, notably, the four-run average forecasts (results not shown) show better forecast skills than the control run forecast alone. Considering the forecast limit of operational models, forecasts at lead times up to 50 days are to be evaluated; thereby, CFSv2 real-time forecast data from September to October are also needed. The initial field is regarded as the observation for verification purposes.

Variables used in this study include daily 2-m temperature (T_s), surface pressure (P_s), surface meridional wind


Forecast Lead Time [day]
50
40
30
20
10
0
16
12
8
4
0
-4
2011N
D
2012J
F
M
2012N
D
2013J
F
M
Forecast Lead Time [day]
2013N
D
2014J
F
M
2014N
D
2015J
F
M
Forecast Lead Time [day]
2015N
D
2016J
F
M
2016N
D
2017J
F
M
Forecast Lead Time [day]
15
10
5
0
-5
2017N
D
2018J
F
M
0
1.2
2.4
3.6
4.8
6
7.2
8.4
9.6

◂**Fig. 1** Variations of the ST60N index (units: 10^9 kg s^{-1}) as a function of verification time (abscissa) during six winters from November 2011 to March 2017 and forecast lead time (ordinate). Day 0 in the lead time corresponds to the initial conditions of CFSv2 forecasts. Systematic error of the indices in forecasts are adjusted according to the 6-year mean seasonal cycle and standard deviations at each forecast lead time shown in Figs. 2 and 3. Black curves in the lower panels are the time series of ST60N derived from the initial conditions of CFSv2 forecasts; variations with periodicity less than 7 days have been filtered out using Lanczos filtering (Duchon 1979); red and blue dots on the black curves respectively indicate the local positive and negative peaks in the corresponding index derived from observation

(v_s), and three-dimensional air temperature (T) and meridional wind (v) fields. The data fields are on 1° latitude × 1° longitude grids and 37 pressure levels from 1000 to 1 hPa. Three-dimensional and surface potential temperature (θ and θ_s) fields are derived from the daily fields of T, T_s, and P_s. Wavenumber-1 and wavenumber-2 components of meridional wind velocities at pressure levels (v_1 and v_2) and surface level (v_{s1} and v_{s2}) are obtained via decomposing v and v_s using the Fourier transform method.

2.2 Stratospheric mass circulation indices

The daily fields of potential temperature and meridional wind at the surface and pressure levels are first interpolated onto 200 equally spaced sigma (σ) levels. Then, the air mass transported into the polar stratosphere above 400 K per second fulfilled by the total waves, wavenumber-1 and wavenumber-2 waves (denoted as ST60N, ST60N_W1 and ST60N_W2, respectively), as a function of time t and forecast lead time τ (τ = 1 to 50), are derived as

$$ST60N(t,\tau) = \int_0^{2\pi}\int_0^{1} m_\sigma \cdot v(\lambda,\ \phi,\sigma,t,\tau)\cdot H(\theta(\lambda,\ \phi,\sigma,t,\tau)-400K)d\sigma R cos\phi d\phi d\lambda$$

$$ST60N\_W1(t,\tau) = \int_0^{2\pi}\int_0^{1} m_\sigma \cdot v_1(\lambda,\ \phi,\sigma,t,\tau)\cdot H(\theta(\lambda,\ \phi,\sigma,t,\tau)-400K)d\sigma R cos\phi d\phi d\lambda$$

$$ST60N\_W2(t,\tau) = \int_0^{2\pi}\int_0^{1} m_\sigma \cdot v_2(\lambda,\ \phi,\sigma,t,\tau)\cdot H(\theta(\lambda,\ \phi,\sigma,t,\tau)-400K)d\sigma R cos\phi d\phi d\lambda, \quad (1)$$

where $m_\sigma = \frac{\Delta\sigma}{g}P_s$, the air mass between two adjacent sigma surfaces per unit area, in which g is the gravitational constant and $\Delta\sigma$ = 1/200; ϕ is 60°N; λ is longitude; R is the radius of Earth; and $H(x)$ is the Heaviside function, such that $H(x)=1$ for $x\geq 0$, or otherwise $H(x)=0$. The initial condition of CFSv2—namely, $\tau = 0$, is approximately regarded as the observation. The ST60N, ST60N_W1 and ST60N_W2 variabilities with periodicity of less than 7 days are filtered out using Lanczos filtering (Duchon 1979). According to Yu et al. (2018b), the net meridional mass transport into the polar stratosphere is mainly contributed by mass transport driven by wavenumber-1 and wavenumber-2. Thus, focusing on these three indices allows us to test the ability of the CFSv2 model in predicting the main features of the state of stratospheric mass circulation at the polar circle.

2.3 Positive and negative peak events of stratospheric mass circulation

Next, we define the observed stronger events, or positive peak events, of meridional mass transport into the polar stratosphere above 400 K by the total waves (as "PULSE" events named by Cai et al. 2016) as well as by the wavenumber-1 and wavenumber-2 components respectively. Taking positive peak events in the ST60N index for example, we first detect all the local peaks in the time series of the ST60N index with a peak value at least above the seven-winter mean value. As seen from the time series of the ST60N index shown by the curves in Fig. 1, there are peaks close to each other and the valley between them is not deep enough to separate them as two events. In that case, we filter out the comparatively smaller peak if the difference of the peak value with the valley value between is less than 0.5 standard deviation (SD). For instance, a small local peak can be detected on 3 February 2016, but its difference with the local valley on 1 February 2016 is smaller than 0.5 SD; thus, it is not considered as a positive peak event of stratospheric mass circulation. As for negative peak events, we detect all the local minimums in the time series of the ST60N index below the climatological mean. For two neighboring peaks, we again remove the negative peak that has the smaller absolute value if its difference with the peak value between is less than 0.5 SD (e.g., the event around 10 December

2012). The positive and negative peaks in the ST60N_W1 and ST60N_W2 indices are detected in a similar fashion and shown in lower panels in Figs. 2 and 3.

As listed in Table 1, during the six winters from 2011 to 2018, there are a total of 77 positive and 81 negative peaks in ST60N, 39 positive and 38 negative peaks in ST60N_W1, and 57 positive and 56 negative peaks in ST60N_W2. Despite only seven winters of data being available, the sample size of the peak events can be said to be sufficiently large for carrying out the forecast verification of dichotomous forecasts of peak events.

2.4 Wave amplitude and tilt indices

The variations of meridional mass transport are mainly driven by variations of both amplitude and the vertical tilt angle of waves (representing the baroclinic instability) according to Johnson (1989) and Yu et al. (2018a). We strictly follow Zhang et al. (2013) and Yu et al. (2018a) and derived two indices of wave properties: the amplitude (denoted as WA) and vertical westward tilting (denoted as WT) of waves along 60°N. Positive values of the WT index indicate vertically westward-tilted waves (i.e., temperature troughs/ridges are on the west side of geopotential height troughs/ridges), and negative values correspond to eastward-tilted waves. According to Cai et al. (2014) and Yu et al. (2018a), the stratospheric WT index (above 250 hPa) has positive values most of the time (more than 90% of days) in winter, which indicates that the wave field in the extratropical stratosphere is dominated by baroclinic waves with a vertically westward-tilted structure.

2.5 Prediction skill evaluation methods

To assess the skill of the stratospheric mass circulation variability, the present study uses different skill evaluation methods: (i) Visual ("eyeball") verification; (ii) the correlation coefficient (CC); and (iii) the hit rate (HR) and false alarm rate (FAR) in a fuzzy sense. The first two represent the deterministic forecast, whereas the coupling relation between the HR and FAR represents the probabilistic forecast.

2.5.1 Anomaly correlation coefficient

The anomaly correlation coefficient (ACC) is a common verification metric used to evaluate deterministic forecasts and identify similarities in the patterns of departures from the climatological mean field (Wilks 2011). The ACC can be defined as:

$$ACC = \frac{\sum_{i=1}^{n} \frac{1}{n}(f_i')(o_i')}{\sqrt{\sum_{i=1}^{n} \frac{1}{n}(f_i')^2 \sum_{i=1}^{n} \frac{1}{n}(o_i')^2}}, \quad (2)$$

where, n is the number of samples, f_i' is the forecast anomaly from its own climatology (7-year mean for each calendar day), and o_i' is the observed anomaly from its own climatology. To investigate if the correlation skill has any dependency on the amplitude of meridional mass transport into the polar stratosphere, we further split the seven-winter time series of the index into two parts. One is the period when the index is above normal, and the other is the period when the index is below normal. We derive the ACC according to Eqs. (3) and (4) for the above-average potions of the observed indices (denoted as $ACC+$), and the below-average portions ($ACC-$), to yield the prediction skill of stronger stratospheric mass circulation and weaker stratospheric mass circulation, separately:

$$ACC+ = \frac{\sum_{i=1}^{n} \frac{1}{n}\left(f_i'\right)\left(o_i'\right)H\left(o_i'\right)}{\sqrt{\sum_{i=1}^{n} \frac{1}{n}\left(f_i'H\left(o_i'\right)\right)^2 \sum_{i=1}^{n} \frac{1}{n}\left(o_i'H\left(o_i'\right)\right)^2}}, \quad (3)$$

$$ACC- = \frac{\sum_{i=1}^{n} \frac{1}{n}\left(f_i'\right)\left(o_i'\right)H\left(-o_i'\right)}{\sqrt{\sum_{i=1}^{n} \frac{1}{n}\left(f_i'H\left(-o_i'\right)\right)^2 \sum_{i=1}^{n} \frac{1}{n}\left(o_i'H\left(-o_i'\right)\right)^2}}, \quad (4)$$

where $H(x)$ is the Heaviside function, such that $H(x) = 1$ for $x \geq 0$, or otherwise $H(x) = 0$.

2.5.2 Fuzzy HR and FAR

The HR is defined as the proportion of occurrences when the observed PULSE event occurs at the same time as the forecasted PULSE event. Similarly, the FAR is defined as the proportion of nonoccurrences, i.e., when the forecasted PULSE event fails to co-occur with the observed events. Commonly, the HR and the FAR are used to represent the ability of a set of probabilistic forecasts to discriminate a dichotomous event (Swets 1973; Mason 1982; Stanski et al. 1989; Kharin and Zwiers 2003). When HR > FAR, the forecasts are considered to be skillful predictions.

When calculating the HR and FAR, we utilize fuzzy verification methods, also called neighborhood verification methods. Fuzzy verification methods (e.g., Davis and Carr 2000; Ebert 2008) give credit to "close" forecasts via not requiring an exact match between forecasts and observations but instead looking in a space/time/intensity neighborhood around the point of interest. As such, fuzzy verification assumes that it is acceptable for the forecast to be slightly displaced and yet still be useful. Treatment of forecast data within a window includes the mean value (upscaling), occurrence of an event somewhere in the window, the frequency of events in the window, and the probability distribution of values within the window. The appropriate window size

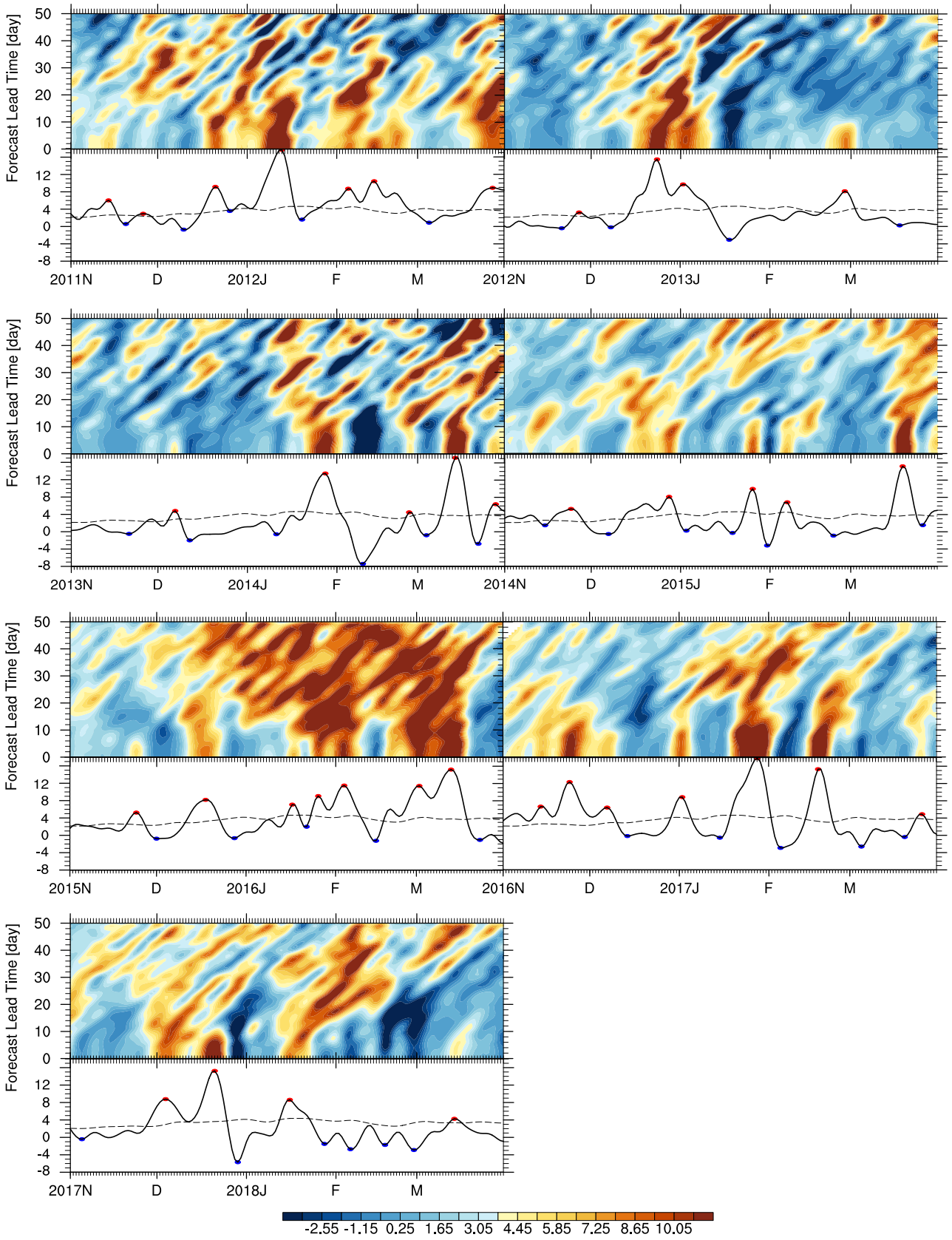


Fig. 2 As in Fig. 1 but for ST60N_W1

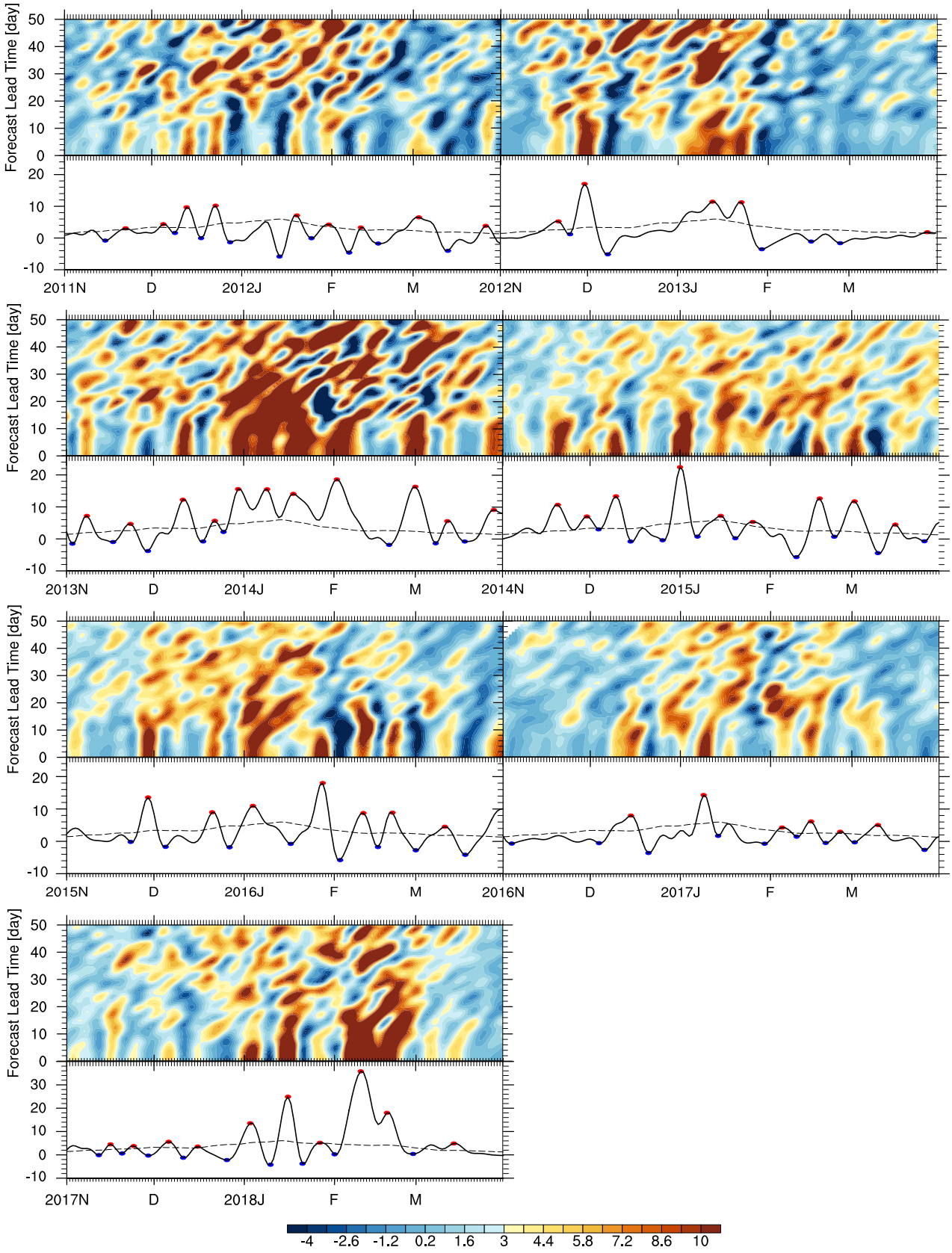


Fig. 3 As in Fig. 1 but for ST60N_W2

Table 1 Number of positive peak events and negative peak events in ST60N, ST60N_W1 and ST60N_W2, derived from the initial conditions of CFSv2 forecasts during the seven winters (November–March) of 2011–2018

	ST60N	ST60N_W1	ST60N_W2
Positive peak			
2011/2012	11	7	9
2012/2013	11	4	5
2013/2014	12	5	11
2014/2015	11	5	9
2015/2016	10	7	7
2016/2017	10	7	6
2017/2018	12	4	10
Seven-winter total	77	39	57
Negative peak			
2011/2012	11	5	9
2012/2013	11	4	5
2013/2014	13	6	8
2014/2015	12	7	9
2015/2016	10	5	8
2016/2017	11	5	9
2017/2018	13	6	9
Seven-winter total	81	38	56

should depend on the temporal resolution and the meteorological situation, so a single value may not be appropriate for all forecasts and domains. Fuzzy verification techniques address this question by allowing the neighborhood size to vary, thereby providing information on forecast quality as a function of scale. According to the total number of events in the seven winters listed in Table 1, the time interval for positive/negative peaks in ST60N, ST60N_W1, ST60N_W2 is about 13, 27 and 18 days, respectively. Forecasts with a time error of less than 3 days could be considered as useful, especially for extended forecast ranges. Therefore, in the present study, we consider three windows for the forecasts of peak events: −1 to 1, −2 to 2, and −3 to 3 days' deviation from the observation. Thus, later in the paper, we examine the fuzzy HR and FAR of stratospheric mass circulation positive/negative peak events allowing time shift errors of 1–3 lead/lag days, separately.

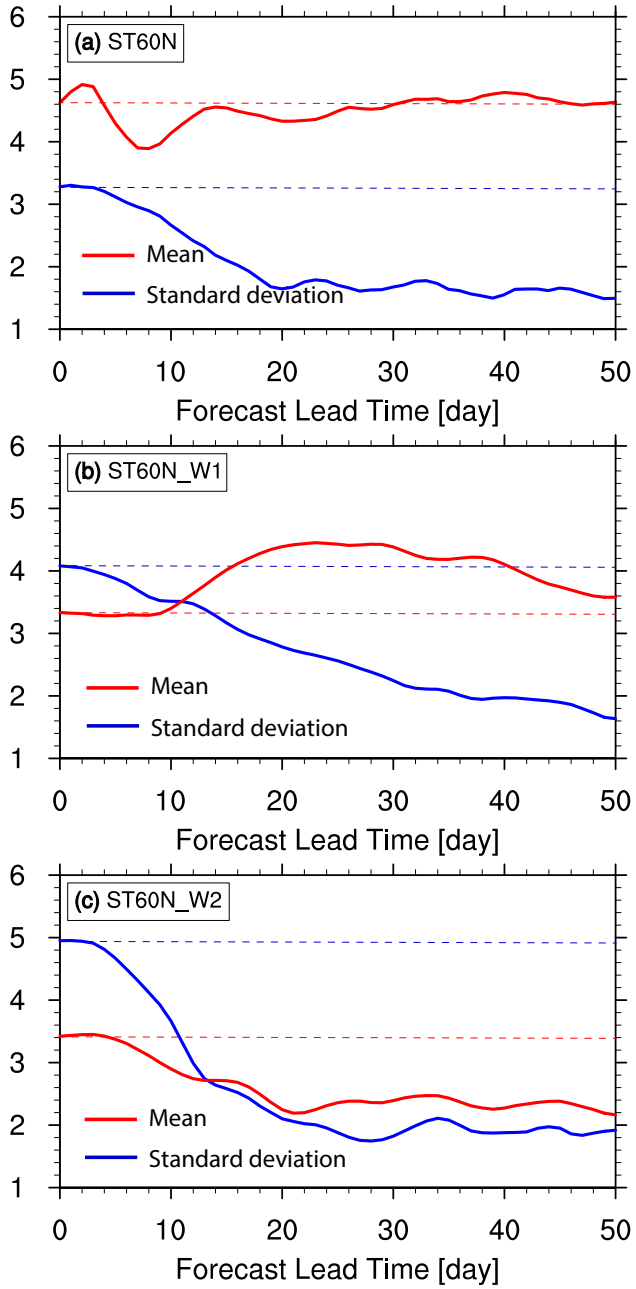


Fig. 4 Seven-winter (November–March) mean (red curve; units: 10^9 kg s^{-1}) and standard deviation (blue curve; units: 10^9 kg s^{-1}) of the **a** ST60N, **b** ST60N_W1 and **c** ST60N_W2 indices, derived from observation (zero forecast lead time) and CFSv2 forecasts at lead times from 1 to 50 days

3 Systematic forecast errors

We first evaluate the forecast error and bias for the climatological mean state of winter stratospheric mass circulation indices, namely ST60N, ST60N_W1 and ST60N_W2. Seen from the red curve of Fig. 4a, the winter mean (from November 2011 to March 2018) of ST60N is slightly overestimated at forecast lead times of 1–2 days compared with the observation (4.6×10^9 kg s^{-1}), then underestimated at 3–14 days, and finally close to observation at longer lead times, indicating that the CFSv2 model can overall capture the winter mean value of ST60N at extended forecast range. The winter mean value of the ST60N_W1 index (red curve in Fig. 4b), however, is overestimated at forecast lead times beyond 1 week, with a maximum forecast error of about 1×10^9 kg s^{-1} above observation (3.5×10^9 kg s^{-1}). In

contrast, the CFSv2 model underestimates the winter mean value of ST60N_W2 (red curve in Fig. 4c) beyond the forecast lead time of 1 week, with a maximum negative deviation of about -0.6×10^9 kg s^{-1} from observation (3.2×10^9 kg s^{-1}). Since the variability of ST60N is mainly contributed by ST60N_W1 and ST60N_W2 (Yu et al. 2018b), the amplification of the mass transport driven by the wavenumber-1 component, and the severe damping of that driven by the wavenumber-2 component, at extended forecast range in the CFSv2 model, explains the overall consistency in the winter mean values of the ST60N index between observations and forecasts made at lead times from 1 to 50 days.

On the other hand, the standard deviations of all three stratospheric mass circulation indices (blue curves in Fig. 4) are significantly damped as the forecast lead increases. The damping is most remarkable for the ST60N_W2 index (blue curve in Fig. 4c), whose standard deviation decreases sharply at forecast lead times within 2 weeks and remains at about 40% the observed standard deviation beyond the 2-week forecast lead. The standard deviation of ST60N_W1 decreases almost linearly with forecast lead time (blue curve in Fig. 4b) and retains more than half the standard deviation derived from observations at the forecast lead time of 1 month. As mainly contributed by ST60N_W1 and S60N_W2, the standard deviation of ST60N (blue curve in Fig. 4a) decreases quickly in the first 2 weeks and then remains almost constant at about 45% the standard deviation derived from observations.

In addition, it is found that the forecast errors in climatological mean and standard deviation of ST60N and ST60N_W1 indices show clear seasonality. We derive the 7-year mean and standard deviation of ST60N at each verification time t (from 1 November to 31 March) and forecast time τ (from 0 to 50 days) as:

$$\overline{ST60N}^{yr}(\tau,t)=\frac{1}{7}\sum_{yr=2011}^{2017}ST60N(\tau,yr,t), \tag{5}$$

$$ST60N\_SD^{yr}(\tau,t)=\frac{1}{7}\sum_{yr=2011}^{2017}\left(ST60N(\tau,yr,t)-\overline{ST60N}^{yr}(\tau,t)\right)^2. \tag{6}$$

A 31-day running mean operator is applied to obtain a smoothly varying annual cycle. The seasonal cycle of the ST60N_W1 and ST60N_W2 indices are calculated in the same way. Shown in the left-hand panels of Fig. 5a–c are the seasonal cycles of ST60N, ST60N_W1 and ST60N_W2 derived from observation and forecasts (i.e., $\overline{ST60N}^{yr}$, $\overline{ST60N\_W1}^{yr}$ and $\overline{ST60N\_W2}^{yr}$). First of all, the CFSv2 model does a good job in capturing the seasonal cycle, supported by the similar seasonal variations between forecasts and observation. However, besides the biases of the winter mean in forecasts as illustrated in Fig. 4, we can clearly see a lead-time-shift error in the wavenumber-1 component. The predicted $\overline{ST60N}^{yr}$, in contrast, tends to lag the observations, by 10–20 days, in forecasts made 1–3 weeks in advance. The CFSv2 model shows little time-shift error for the seasonal variation of the wavenuber-2 component. The seasonal variations of the interannual standard deviation of all three indices (shown in Fig. 5d–f) exhibit a systematic shift of maximum values towards a later time than observed, especially for $\overline{ST60N\_SD}^{yr}$ and $\overline{ST60N\_W1\_SD}^{yr}$.

4 Evaluation of the forecast skill of stratospheric mass circulation indices

To reduce model forecast bias as mentioned above, we conduct a systematic adjustment for forecasts of the ST60N index according to:

$$ST60N_c(yr,\tau,t)=\frac{ST60N(yr,\tau,t)-\overline{ST60N}^{yr}(\tau,t)}{\overline{ST60N}^{yr}_{SD}(\tau,t)}\cdot\overline{ST60N}^{yr}_{SD}(\tau=0,t)+\overline{ST60N}^{yr}(\tau=0,t). \tag{7}$$

Shown in the upper panels of Fig. 1 are the temporal evolutions of systematically adjusted ST60N (i.e., $ST60N_c$) derived from the CFSv2 forecasts at lead times of 1–50 days after systematic error correction and observations (0 forecast lead day) during seven winters from 2011 to 2018. Let us first visually compare the winter time series at different forecast lead times. It is seen that the timing of a large portion of periods of below-normal and above-normal mass transport into the polar stratosphere can be captured in forecasts at lead times up to 30 days. However, we have to admit that the forecasts made within the range of less than 1 month have several flaws: first, an error of a few days exists for the peak time of stronger and weaker ST60N events; in addition, the consistency between the forecasts and observations is not always continuous from 1 to 30 days; the consistency between forecasts and observations looks to possess seasonal variations—namely, slightly weakened in both the early winter month (November) and late winter month (March), compared with December–February. Beyond the forecast lead time of 30 days, the exact timing of stronger and weaker ST60N events looks to be obscured in most cases, but the general intensity of stratospheric mass circulation still shows some consistency with observations at the monthly timescale. For example, the winter of 2015/16 can be easily divided into two periods based on the ST60N, which is below normal during the first half of the winter season (November–December) but above normal during the second half (January–March). The CFSv2 forecasts reproduce this general pattern of the temporal evolution of ST60N, even at forecast lead times of 40–50 days, albeit with a time-shift

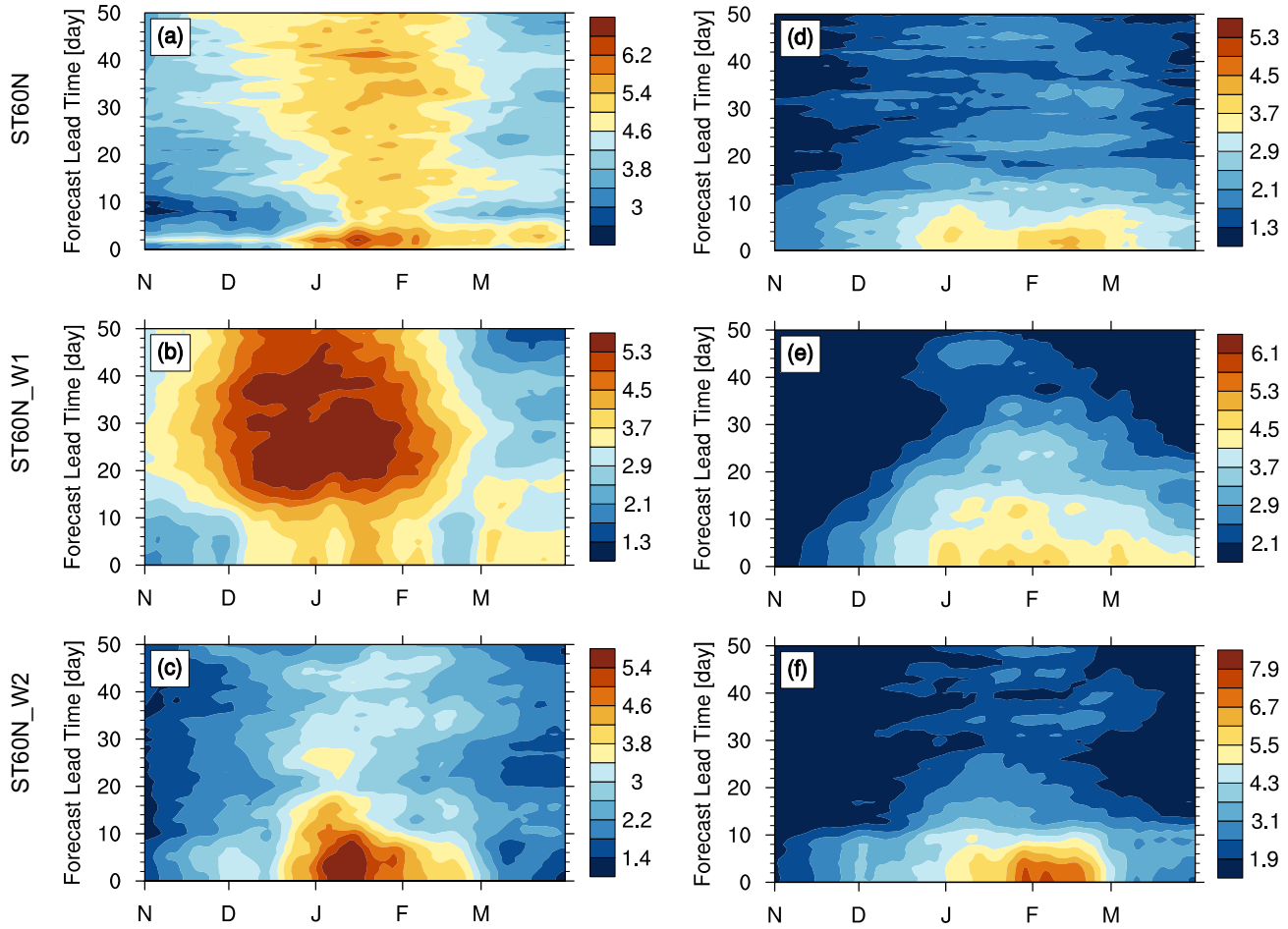


Fig. 5 Winter seasonal cycle of the **a** ST60N, **b** ST60N_W1 and **c** ST60N_W2 indices (units: 10^9 kg s^{-1}) derived from observation (zero forecast lead time) and CFSv2 forecasts at lead times from 1 to 50 days in the seven winters (November–March) from 2011 to 2018. Panels **d–f** are the same as **a–c** but for the 7-year standard deviations

error. The winter of 2012/13 is another demonstration. The systematically adjusted forecast time series of mass transport driven by the wavenumber-1 and wavenumber-2 components (plotted in the upper panels of Figs. 2, 3) shows that forecasts made at less-than-20-day-lead are in good agreement with observations (lower panels of Figs. 2, 3); whereas at longer lead times, the occurrence of some peak events can still be captured by the CFSv2 model but the error in the timing of individual events tends to be larger.

4.1 ACCs

To quantitatively and objectively measure the prediction skill of the stratospheric mass circulation indices, we plot the lead–lag ACCs between observations and forecasts made at lead times from 0 to 50 days in the left-hand panels of Fig. 6. An ACC of 0.3 (marked by the horizontal black line) is considered as the cutoff for "marginally useful" skill in the subseasonal range—a term borrowed from the experience of the predictions of upper-level (i.e., 500-hPa height) charts in the 6–10-day range (e.g., Jones et al. 2000; Hamill et al. 2004; Zhang et al. 2013; Weber and Clifford 2017). It can be seen that the maximum lead–lag correlation of forecasts with observations can be continuously above 0.3 at the forecast lead time of 14 days for ST60N, 23 days for ST60N_W1, and 18 days for ST60N_W2. A returning skill can be seen at the forecast lead time of 20 days for ST60N and 22 days and even 40 days for ST60N_W2. The maximum lead–lag correlation is mainly centered at zero lag but tends to shift towards positive lag days for ST60N_ST60N_W2 and negative lag days for ST60N_W1 at longer lead times, indicating a possible time-shift error in CFSv2 forecasts. We can see more clearly the ACC score of the operational CFSv2 forecasts for the ST60N, ST60N_W1 and ST60N_W2 indices with no time lag from the black curves in the right-hand panels of Fig. 6.

We also observe an improvement of predictability for large amplitude stratospheric mass circulation indices from

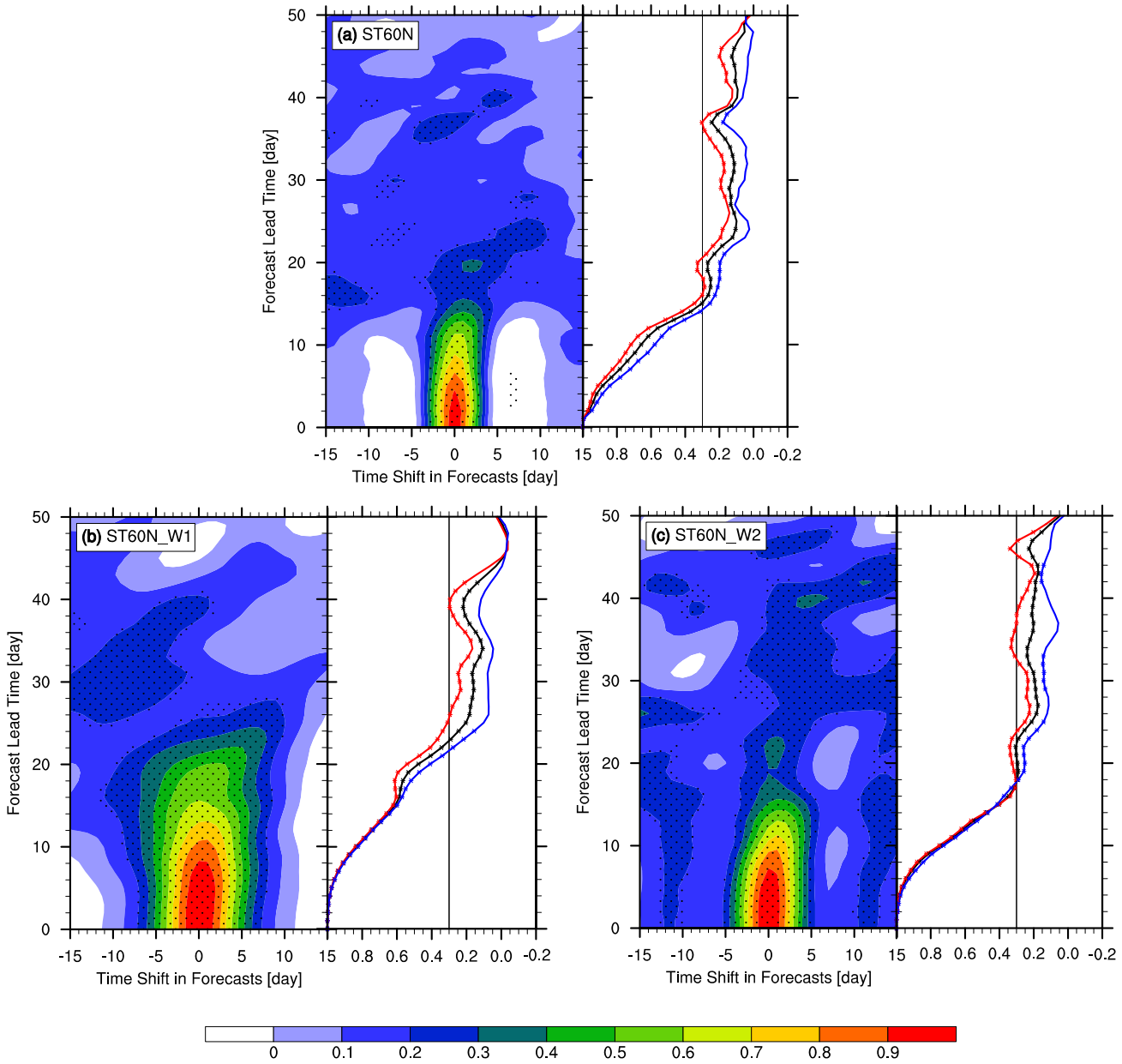


Fig. 6 Lead–lag ACCs (left-hand panels) and zero-lag ACCs (right-hand panels) of the CFSv2 forecasts for the seven winters of 2011–2018 as a function of the forecast lead time (ordinate; units: days): **a** ST60N; **b** ST60N_W1; **c** ST60N_W2. The black curve is for the entire wintertime; the red curve is for the period when the index is above climatology; and the blue curve is for the period when the index is below climatology. The dots indicate statistical significance at the 5% level

the larger ACCs for the forecasts of above-normal values of stratospheric mass circulation indices (i.e., stronger stratospheric mass circulation, ACC+) and below-normal values (i.e., weaker stratospheric mass circulation, ACC−), shown by the red and blue curves overlaid in the right-hand panels of Fig. 7a–c. It can be seen that ACC+ is higher than ACC− at all forecast lead times longer than 2 weeks for all three indices. Accordingly, the ACC+ of ST60N remains above 0.3 till the forecast lead time of 20 days, while its ACC− falls below 0.3 from the forecast lead time of 2 weeks. The ACC+ of ST60N_W1 is continuously above 0.3 till the forecast lead time of 25 days, while its ACC− starts to fall below 0.3 at the forecast lead time of 21 days. The ACC+ of ST60N_W2 is above 0.3 till the forecast lead of 23 days, while its ACC− falls below 0.3 at the forecast lead of 18 days. Of note is that the ACC+ of ST60N and ST60N_W1 approaches 0.3 again at the forecast lead times of 36 and 40 days, respectively. The ACC+ of ST60N_W2 has two periods of useful skill beyond 1 month—one is at forecast lead times of 31–36 days, and the other at around 45 days.

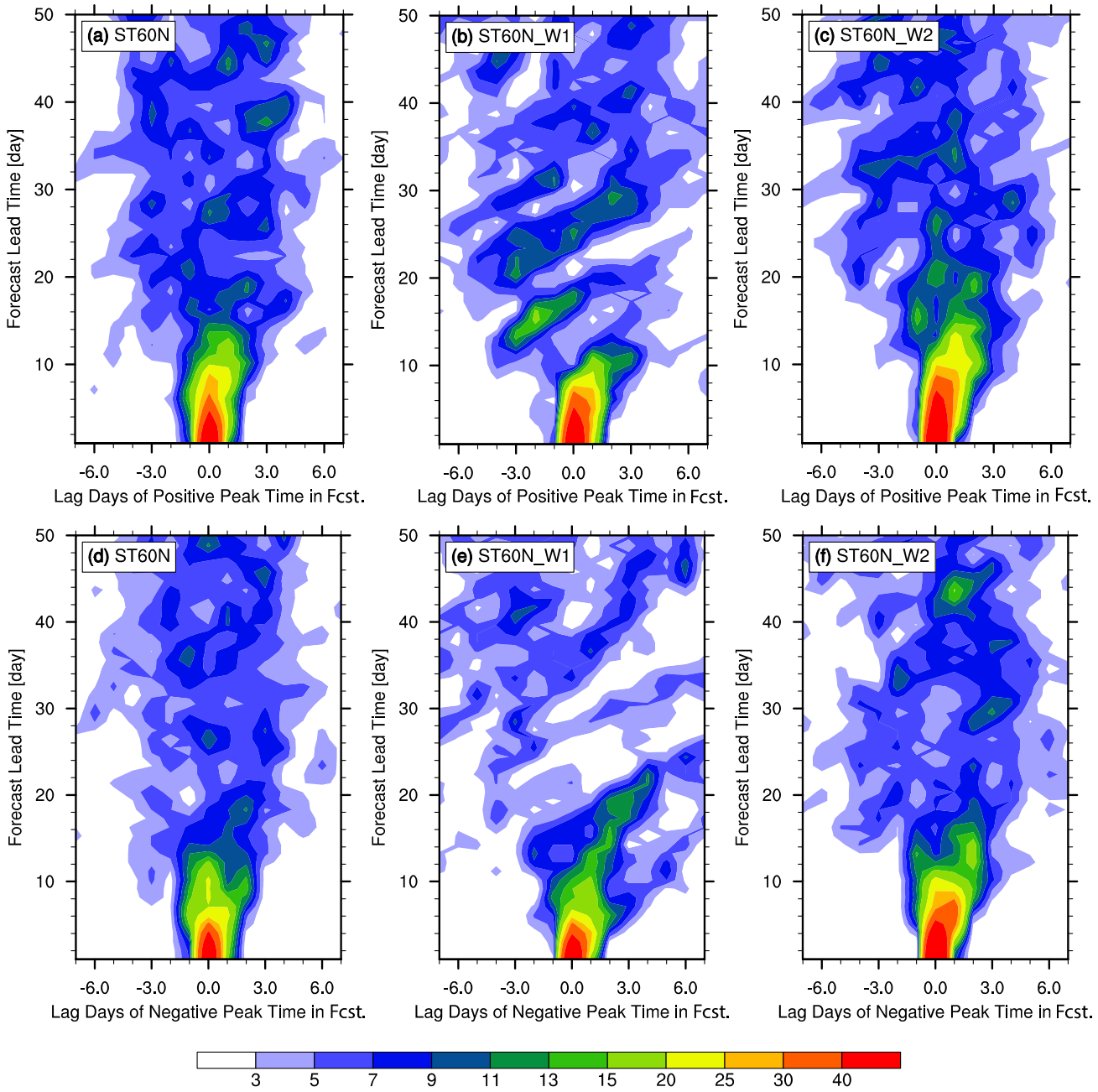


Fig. 7 Probability (units: %) of positive peaks in **a** ST60N, **b** ST60N_W1 and **c** ST60N_W2 in the winter time series forecasted at various lead times to occur at lag times from −7 to 7 days relative to the corresponding positive peaks in observation. Panels **d–f** are the same as **a–c** but for the negative peaks

4.2 Probabilistic prediction skill of peak events

The timing of occurrence of the pulse-like increase or decrease in the stratospheric mass circulation is another important aspect of evaluating the forecast skills of the stratospheric mass circulation. According to Cai et al. (2016) and Yu et al. (2018b, c), it is the timing of the pulse-like signal of ST60N that has an almost one-to-one correspondence with continental-scale CAO events at the surface. Therefore, the forecasts of stratospheric mass circulation variability can also be considered as a dichotomous forecast or yes/no forecast. In this section, we examine the fuzzy FAR and POD of both positive and negative peaks in the ST60N, ST60N_W1 and ST60N_W2 indices. We detect peaks in the time series spanning over the whole winter derived from forecasts made at each of the 1–50 days of lead time and verify the timing of peaks with those derived from observation. The bias (not shown), which is the ratio of the total number of forecast

peak events to the total number of observed events, is close to 1, which indicates that the forecast system does a good job in forecasting the relative frequencies of peak events.

To find the correspondence of the forecasted peaks with the observed peaks as objectively as possible, we make pair of a peak in the observational time series with its counterpart in the forecast time series following three steps. Taking ST60N positive peaks as an example, for the observed positive peak ST60N event #n (n = 1, N, where N is the total number of observed events), we search for the nearest positive peak in the forecast time series at forecast lead time of τ days ($\tau = 1,\ 50$). Thus, for each τ we have N forecast peaks paired with N observed peaks. Considering the dominant timescale of ST60N is about 2 weeks, we discard the forecast peaks that have a time shift of longer than 1 week relative to the paired peaks in the observation. In cases where two or more observed peaks are paired with one matching peak in the forecasts, we only keep the pair bond between this forecast peak and the nearest observed peak, and discard the other pair bond. We denote the time lag of the paired forecast peak with the observed peak as $\mathbf{\Gamma}(\mathrm{n}, \tau)$, and its value range is by design from −7 to 7 or undefined for those observed peaks that fails to pair with any forecast peak. The probabilities of forecasted positive peaks at various forecast lead times, occurring at a specific lag time $\mathbf{\Gamma}_0$ (from −7 to 7 days) relative to the corresponding peaks in observations can be derived as

$$P\left(\mathbf{\Gamma}_0,\ \tau\right) = \frac{\sum_{n=1}^{N} Y(\mathbf{\Gamma}(\mathrm{n}, \tau) - \mathbf{\Gamma}_0)}{N}, \tag{8}$$

where $Y(x) = 1$ for $x = 0$ and otherwise $Y(x) = 0$. $P\left(\mathbf{\Gamma}_0,\ \tau\right)$ indicates the percentage of observed events successfully forecasted by the model despite of time-shift error of $\mathbf{\Gamma}_0$ days. Similar pairing procedure and calculation can be conducted on ST60N_W1 and ST60N_W2 indices.

It can be seen from Fig. 7a–c that the time shift of forecast peaks has a broader spread as the forecast lead time increases. However, most of the forecast peaks occur within a lead–lag time of 3 days with respect to the observations. At forecast lead times less than 2 weeks, the forecast peak time tends to lag the observation by a few days. Beyond the 2-week forecast lead time, positive peaks in each of the three indices show an inconsistent tendency to lead or lag the observed peaks. Similar features can be seen for the forecast and observed negative peaks (shown in Fig. 7d–f), except that the large values of occurrence probability (e.g., > 15%) are confined within a shorter forecast lead time range than the positive peaks. Therefore, it can be stated that the forecast peaks tend to occur within 1 week centered at the peak dates in observations, but little preference for lead–lag error can be found in forecasts made beyond 2-week lead times.

The summation of the probability shown in Fig. 7 within a specific range of $\mathbf{\Gamma}_0$ (the x-axis) represents the POD or HR in a fuzzy sense, as introduced in Sect. 2.5b. If we consider a forecast positive peak of ST60N during the period from 1 day before to 1 day after the corresponding observed positive peak a successful detection, the fuzzy POD of the ST60N positive peak (shown by the solid red curve in Fig. 8a) is obtained by summing up the probability values in Fig. 7a from $\mathbf{\Gamma}_0 = -1$ day to $\mathbf{\Gamma}_0 = 1$ day. In the same fashion, we calculate the fuzzy POD of the ST60N positive peaks allowing 2- and 3-day time error and plot them in Fig. 8b, c. We further derive the fuzzy FAR as the difference between the total number of forecast events and the successfully detected event number (i.e., the fuzzy POD multiplied by the total number of observed events) divided by the total number of forecast events at each lead time τ. The fuzzy FARs for ST60N peak events are shown by the solid blue curves in Fig. 8a–c. The fuzzy POD and FAR (allowing 1-, 2- and 3-day time-shift error) for peaks of ST60N_W1 and ST60N_W2 are also computed and displayed in Fig. 8d–f, g–i, respectively.

Seen from Fig. 8, as the forecast lead time increases, the fuzzy POD decreases but the fuzzy FAR increases, as forecasts of other tropospheric weather events. Following weather forecasters and considering forecasts with a POD exceeding the FAR as useful, we found that the predictable period for ST60N positive peak events is about 12 days for the case allowing a 1-day time shift error, about 15 days for a 2-day time shift error, and 21 days for a 3-day time shift error. The predictable period for ST60N_W2 peaks are almost the same as those for ST60N. The forecast limit for ST60N_W1 positive peaks is shorter, mainly due to the larger timescale of ST60N_W1, which may involve a longer time shift between adjacent peaks. The fuzzy POD and FAR of negative peaks in the stratospheric mass circulation indices exhibit similar variations with forecast lead time, but the forecast lead time when the fuzzy PDO starts to exceed FAR is slightly shorter, especially for the cases of ST60N allowing a 1-day time shift, ST60N_W1 allowing a 1–3-day time shift, and ST60N_W2 allowing a 2-day time shift. This indicates a shorter predictable period for negative peaks than positive peaks, consistent with the relatively shorter predictable period for the smaller amplitude values of indices in terms of ACC shown in Fig. 6.

5 Forecast skills of wave amplitude and tilt angle

Johnson (1989) and a later study by Cai and Shin (2014) pointed out that the main driving forcing of the meridional mass circulation in the extratropics is the baroclinically amplifying waves. According a conceptual model of a baroclinically amplifying wave [see Fig. 10 in Johnson (1989)], the westward tilt of planetary waves with height contributes to the asymmetry between the air mass amount before and behind troughs/ridges in a given isentropic layer, leading

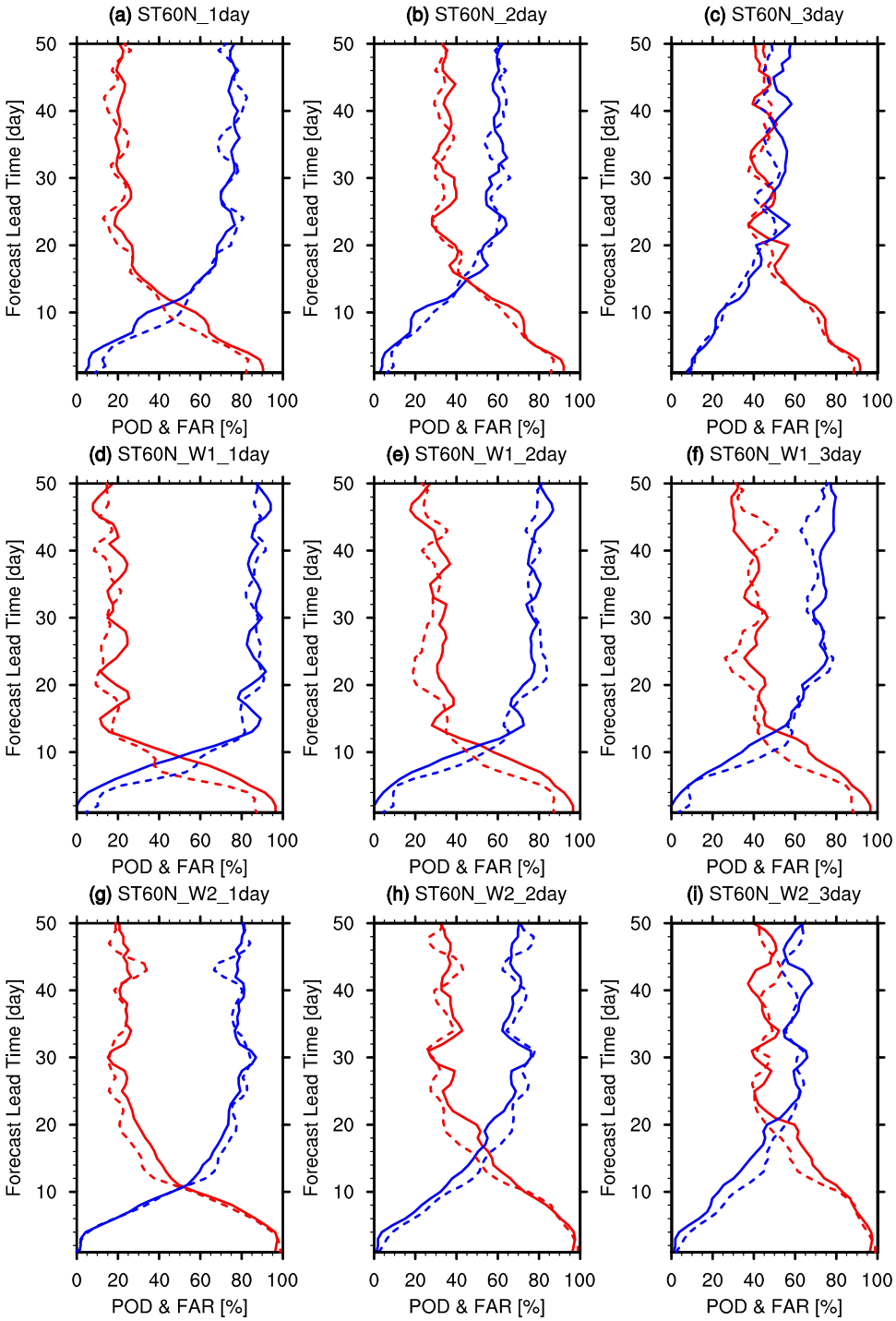


Fig. 8 The fuzzy probability of detection (POD; red solid curve) and false alarm rate (FAR; blue solid curve) of the positive peaks in **a** ST60N, **b** ST60N_W1 and **c** ST60N_W2 in winter at forecast lead times from 1 to 50 days. Red and blue curves indicate cases when a 1-day timing error is allowed—namely, as long as the timing of the forecast peak is within $^{-1}$ to 1 day relative to the observed peak, we regard the peak in observation as successfully detected in the forecasts. Panels **d**–**f** are for the case when a 2-day timing error is allowed, and **g**–**i** are for the case when a 3-day timing error is allowed. Dashed curves are the same at the solid curves but for negative peaks

to net poleward/equatorward mass transport in upper/lower layer. The overall amplitude of waves along a latitudinal band, on the other hand, determines the meridional wind velocity that transports the air mass polewards or equatorwards. Yu et al. (2018a) reported that the maximum correlation of the ST60N index with the wave amplitude at 60°N (WA) is around 70–50 hPa, and that with the wave tilt angle (WT) is around 150 hPa. Therefore, we next investigate the prediction skills of the variability of amplitude and the vertical tilt of waves at stratospheric levels, to gain a better understanding of the prediction skills of the adiabatic mass transport into the polar stratosphere above 400 K.

The forecast error and bias for the seven-winter mean wave indices measuring the equivalent amplitude and vertical westward tilt angle of total waves, the wavenumber-1 component, and the wavenumber-2 component, at various

stratospheric levels, are shown in Fig. 9. It can be seen that the winter mean WA index (Fig. 9a) and WT index (Fig. 9g) for the total waves derived from forecasts are consistent with that derived from observations at levels below 50 hPa. At upper-stratospheric levels above 50 hPa, WA is slightly overestimated around forecast lead times of 2–4 weeks but underestimated at forecast lead times beyond 1 month, and such bias is partly cancelled out by the slight overestimation bias of WT (Fig. 9g). This explains why the winter mean value of the ST60N index remains almost unchanged as the forecast lead time increases (red curve in Fig. 4a). The forecast winter mean WA of wavenumber-1 waves (Fig. 9b) shows similar features with the WA of total waves, while the WT of wavenumber-1 waves (Fig. 9h) exhibits a clear overestimation bias at almost all levels, leading to the overestimation of the winter mean ST60N_W1 (red curve in Fig. 4b). For the

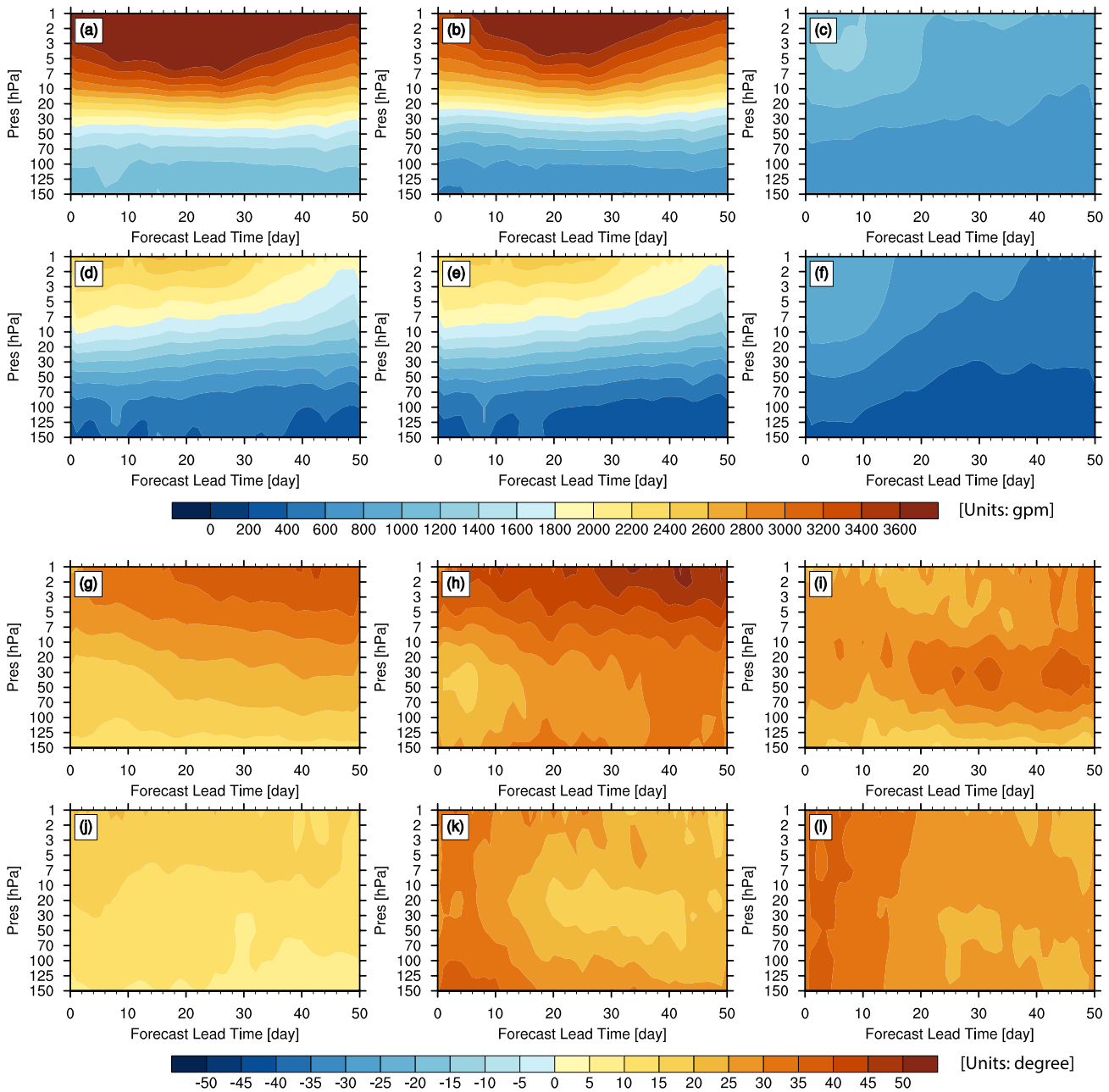


Fig. 9 Seven-winter (November–March) mean of wave amplitude index (WA; units: gpm) of **a** total waves, **b** wavenumber-1 component, and **c** wavenumber-2 component, at stratospheric levels above 150 hPa derived from observation (zero forecast lead time) and CFSv2 forecasts at lead times from 1 to 50 days. Panels (**d**–**f**) are the same as (**a**–**c**) but for the standard deviation of WA. Panels (**g**–**j**) are the same as (**a**–**f**) but for the seven-winter mean and standard deviation of wave westward tilt angle index (WT; units: degrees)

wavenumber-2 waves, both the winter mean WA (Fig. 9c) at all levels aove 150 hPa and WT (Fig. 9i) at lower stratospheric levels (125 and 150 hPa) decrease with an increase in forecast lead time, explaining the underestimation of the winter mean value of ST60N_W2 shown by the red curve in Fig. 4c. From Fig. 9d–f, j–l, the standard deviations of WA and WT in forecasts are all decreasing as the forecast lead increases for total waves and the wavenumber-1 and − 2 components, but the underestimation of the standard deviation of WT is less severe for total waves. By comparing with the blue curves in Fig. 4, it can be stated that the underestimation bias of the standard deviation of both the wave amplitude and vertical tilt gives rise to the forecast bias of the winter standard deviation of stratospheric mass circulation indices.

To investigate CFSv2's capability in capturing the intraseasonal variations of waves, we derive the anomalies of WA and WT of the total waves, wavenumber-1 component, and wavenumber-2 component, at stratospheric levels above 150 hPa, via the removal of the seasonal cycle from the total field, and then calculate the ACCs of the CFSv2-forecast WA and WT. It can be seen from Fig. 10a, d that the rate of decline of the ACC of WA for the total waves is smaller than that for WT. The forecast lead when the ACC falls below 0.3 is about 35–40 days for WA but about 14–25 days for WT, at various stratospheric levels. Specifically, the forecasts of the WA anomaly at 70–20 hPa, which is closely related with ST60N (Yu et al. 2018a), show useful skill beyond 50 days. However, the WT at lower levels (i.e., the level that has maximum positive correlations with ST60N according to Yu et al. 2018a) shows the shortest forecast limit, at 14 days. These results infer that the prediction limit of around 20 days for the stratospheric mass circulation indices and PULSE events is mainly restricted by the prediction limit for the wave tilt angle. Compared with the ACC of WA and WT of total waves, the ACC of WA and WT for wavenumber-1 and wavenumber-2 waves shown in Fig. 10b, c, e, f falls below 0.3 at a shorter forecast lead time, although the prediction of WA of wavenumber-1 component can still be useful at more than 1 month in advance in the middle stratosphere. This implies that the CFSv2 model has greater difficulty in forecasting the spatial scale of waves than forecasting the total waves at longer lead times, resulting in the forecast errors of the variations of the ST60N_W1 and ST60N_W2 indices.

6 Conclusions

Using the subseasonal forecast dataset of CFSv2 for the period 2011–2018, this study evaluates the model's prediction skill for the stratospheric meridional mass circulation variability in seven winters (November–March). We consider three circulation indices, include the meridional mass transport into the polar stratosphere above 400 K by the total flow (ST60N), and the wavenumber-1 (ST60N_W1) and wavenumber-2 waves (ST60N_W2). They are mainly driven by the westward tilt of planetary waves with height and the overall amplitude of waves along a latitudinal band. These three indices represent main features of stratospheric mass circulation around the polar circle, which are coupled with the equatorward cold-air mass transport at lower levels and thus closely related with specific spatiotemporal evolutions of continental-scale CAOs at the surface. The prediction skill is quantified using deterministic skill metrics, the ACC, and probabilistic skill metrics, through a combination of fuzzy POD and FAR. The prediction skills of both the climatological mean seasonal variations and the synoptic-scale anomalous variations of the stratospheric indices, analyzed for lead times up to 50 days, are investigated, and indicate encouraging results in general.

By examining the forecast skills of the climatological mean fields of the stratospheric mass circulation indices, we find that CFSv2 model forecasts well capture the climatological winter mean of ST60N, but overestimate/underestimate the mean of ST60N_W1/ ST60N_W2; forecasts show damping of the wintertime standard deviation of all three indices with an increase in the forecast lead time, especially for the ST60N_W2. These flaws in stratospheric mass circulation indices are tied to the damping of amplitude and westward tilting variations of total waves and the failure to forecast the exact contributions from different spatial scales of waves. The systematic error of the stratospheric mass circulation indices also exhibits some seasonality, especially in the ST60N_W1: the forecasted seasonal variations of ST60N_W1 tend to lead the observations at forecast lead time longer than 2 weeks.

Evaluation of the ability of CFSv2 model to predict the intraseasonal variations of stratospheric mass circulation indices after removing the seasonal cycle yields that: (i) the CFSv2 model can capture not only the overall variability but also the timing of peaks of stratospheric mass circulation beyond a 1- to 2-week prediction limit of the troposphere; (ii) above-normal values and positive peaks of stratospheric mass circulation indices exhibit a longer prediction limit than the below-normal values and negative peaks of stratospheric mass circulation indices; and (iii) the forecast error in the overall temporal variability of ST60N_W1 is smaller than those of the ST60N and ST60N_W2 indices, whereas the forecast error in the timing of peaks in the ST60N and ST60N_W2 indices is smaller than that of ST60N_W1, which infers that the subseasonal predictability of the values of ST60N mainly derives from ST60N_W1, while for the timing of positive and negative peaks it comes from ST60N_W2. We summarize in Fig. 11 the predictable periods of the stratospheric indices and peak events using different methods.

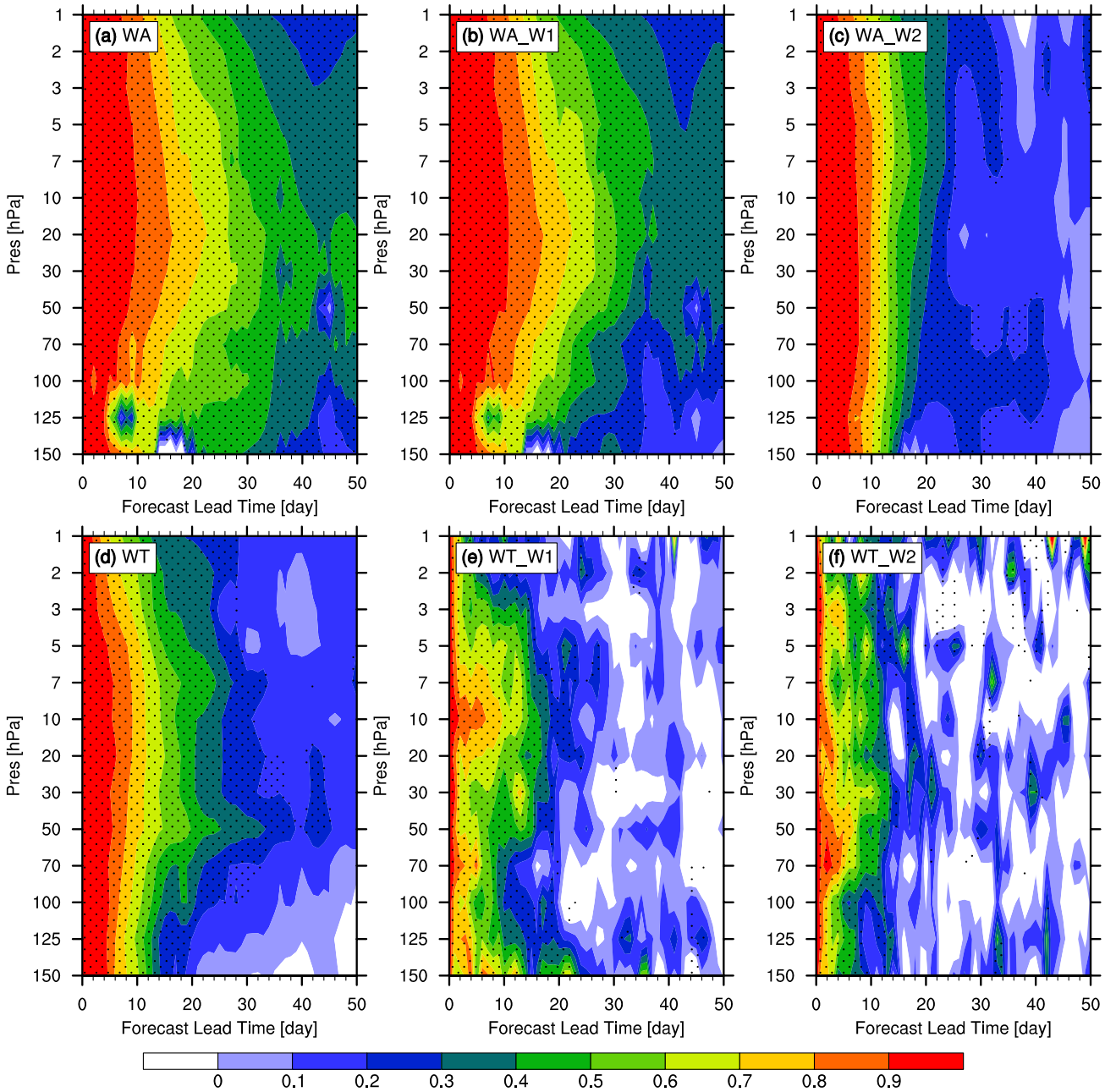


Fig. 10 Correlation scores of the CFSv2 forecasts for the WA of **a** total waves, **b** wavenumber-1 component, and **c** wavenumber-2 component, at stratospheric levels above 150 hPa during the seven winters of 2011–2018 as a function of the forecast lead time (ordinate; units: days). Panels (**d**–**f**) are the same as (**a**–**c**) but for WT. The dots indicate statistical significance at the 5% level

In addition, we examine the prediction skills of the amplitude and vertical tilt angle of waves after removing the seasonal cycle, which drive the anomalous intraseasonal variations of meridional mass circulation. Results show that the CFSv2 model can faithfully capture the wave amplitude more than 50 days in advance, while the predictability of wave tilt angle exhibits a much shorter limit. Thus, the prediction limit of around 20 days for the stratospheric mass circulation indices could mainly be due to the limitation of the CFSv2 model in predicting the variability of wave tilt angle, which represents baroclinic instability. Results also show that the CFSv2 model has greater difficulty in forecasting the spatial scale of waves than the total waves at longer lead times, which results in the larger forecast errors of the synoptic-scale variations of the ST60N_W1 and ST60N_W2 indices than the ST60N index.

This study provides supporting evidence for the feasibility of the prototype hybrid paradigm for sub-seasonal

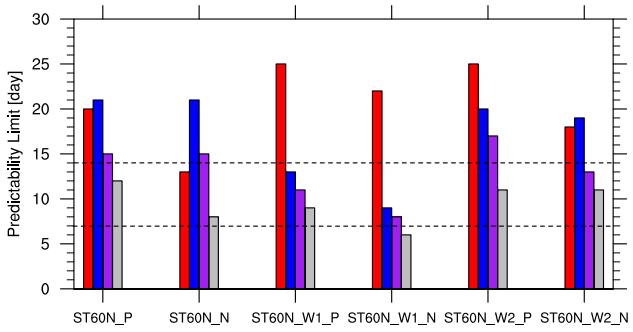


Fig. 11 Predictability limit (units: days) of the ST60N, ST60N_W1, and ST60N_W2 indices in the seven winters (November–March) of 2011–2018. Red bars indicate the maximum forecast lead time with a correlation score exceeding 0.3. Blue, purple and gray bars indicate the maximum forecast lead time with the fuzzy POD exceeding the fuzzy FAR in verification when a 3-, 2- and 1-day time shift is allowed, respectively

forecasts of continental-scale CAOs put forward by Cai et al. (2016). In this hybrid paradigm, the precondition is the faithful forecast skills of stratospheric mass circulation variability (especially for large-amplitude signals). Using the faithful forecast time series of the stratospheric mass circulation indices to detect the timing of PULSE events, and then plugging the information of peak days, duration time, intensity, wave components into the pre-constructed statistical model linking the PULSE events to the continental-scale CAOs, the forecaster can issue forecasts for the temporal and spatial distribution of high probability for cold temperatures of different severity.

Acknowledgements This work was supported by grants from the National Science Foundation of China (41705039, 91537213, 41705024), the Startup Foundation for Introducing Talent of NUIST (2017r068), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The CFSv2 subseasonal forecast data used in this work are available from http://nomads.ncep.noaa.gov/pub/data/nccf/com/cfs/prod/cfs/.

References

Baldwin MP, Dunkerton TJ (2001) Stratospheric harbingers of anomalous weather regimes. Science 294:581–584

Baldwin MP, Stephenson DB, Thompson DWJ, Timothy JD, Andrew JC, Alan O (2003) Stratospheric memory and skill of extended-range weather forecasts. Science 301:636–640

Cai M, Shin CS (2014) A total flow perspective of atmospheric mass and angular momentum circulations: Boreal winter mean state. J Atmos Sci 71:2244–2263

Cai M, Barton C, Shin CS, Chagnon JM (2014) The continuous mutual evolution of equatorial waves and the Quasi-Biennial Oscillation of zonal flow in the equatorial stratosphere. J Atmos Sci 71:2878–2885

Cai M, Yu YY, Deng Y, van den Dool HM, Ren RC, Saha S, Wu XR, Huang J (2016) Feeling the pulse of the stratosphere: an emerging opportunity for predicting continental-scale cold air outbreaks one month in advance. Bull Am Meteorol Soc 97:1475–1489

Charlton AJ, Polvani LM (2007) A new look at stratospheric sudden warmings: Part I: climatology and modelling benchmark. J Clim 20:449–469

Christiansen B (2005) Downward propagation and statistical forecast of the near-surface weather. J Geophys Res 110:D14104

Davis C, Carr F (2000) Summary of the 1998 workshop on mesoscale model verification. Bull Am Meteorol Soc 81:809–819

Duchon CE (1979) Lanczos filtering in one and two dimensions. J Appl Meteorol 18:1016–1022

Ebert (2008) (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. Meteorol Appl 15:51–64

Gerber EP, Orbe C, Polvani LM (2009) Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. Geophys Res Lett 36:L24801

Hamill TM, Whitaker JS, Wei X (2004) Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. Mon Weather Rev 132(6):1434–1447

Haynes P (2005) Stratospheric dynamics. Annu Rev Fluid Mech 37:263–293

Hitchcock P, Simpson IR (2014) The downward influence of stratospheric sudden warmings. J Atmos Sci 71:3856–3876

Iwasaki T, Mochizuki Y (2012) Mass-weighted isentropic zonal mean equatorward flow in the Northern Hemispheric winter. SOLA 8:115–118. https://doi.org/10.2151/sola.2012-029

Iwasaki T, Shoji T, Kanno Y, Sawada M, Takaya K, Ujiie M (2014) Isentropic analysis of polar cold air mass streams in the Northern Hemispheric winter. J Atmos Sci 71:2230–2243. https://doi.org/10.1175/JAS-D-13-058.1

Johnson DR (1989) The forcing and maintenance of global monsoonal circulations: an isentropic analysis. Adv Geophys 31:43–316

Jones C, Waliser DE, Schemm J-KE, Lau WKM (2000) Prediction skill of the Madden and Julian Oscillation in dynamical extended range forecasts. Clim Dyn 16:273–289

Jung T, Leutbecher M (2007) Performance of the ECMWF forecasting system in the Arctic during winter. Quart J R Meteorol Soc 133:1327–1340

Kharin VV, Zwiers FW (2003) On the ROC score of probability forecasts. J Clim 16:4145–4150

Kidston J, Scaife AA, Hardiman SC, Mitchell DM, Butchart N, Baldwin MP, Gray LJ (2015) Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. Nat Geosci 8:433–440

Kolstad EW, Charlton-Perez AJ (2010) Observed and simulated precursors of stratospheric polar vortex anomalies in the Northern Hemisphere. Clim Dyn 37:1443–1456

Lehtonen I, Karpechko AY (2016) Observed and modeled tropospheric cold anomalies associated with sudden stratospheric warmings. J Geophys Res 121:1591–1610

Li JP, Ding RQ (2011) Temporal-spatial distribution of atmospheric predictability limit by local dynamical analogs. Mon Weather Rev 139:3265–3283

MacDonald AE (2005) A global profiling system for improved weather and climate prediction. Bull Am Meteorol Soc 86:1747–1764

Marshall AG, Scaife AA (2010) Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution. J Geophys Res 115:D16114

Marshall AG, Scaife AA, Ineson S (2009) Enhanced seasonal prediction of European winter warming following volcanic eruptions. J Clim 22:6168–6180

Mason I (1982) A model for assessment of weather forecasts. Aust Meteorol Magn 30:291–303

Mitchell DM, Gray LJ, Anstey J, Baldwin MP, Charlton-Perez AJ (2013) The influence of stratospheric vortex displacements and splits on surface climate. J Clim 26:2668–2682

Mukougawa H, Hirooka T, Kuroda Y (2009) Influence of stratospheric circulation on the predictability of the tropospheric Northern Annular Mode. Geophys Res Lett 36:L08814. https://doi.org/10.1029/2008GL037127

Ngan K, Eperon GE (2011) Middle atmosphere predictability in a numerical weather prediction model: revisiting the inverse error cascade. Q J R Meteorol Soc 138:1366–1378

Noguchi S, Mukougawa H, Hirooka T, Taguchi M, Yoden S (2014) Month-to-month predictability variations of the winter-time stratospheric polar vortex in an operational one-month ensemble prediction system. J Meteorolol Soc Jpn 92:543–558. https://doi.org/10.2151/jmsj.2014-603

Polvani LM, Waugh DW (2004) Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. J Clim 17:3548–3554

Saha S et al (2014) The NCEP climate forecast system version 2. J Clim 27: 2185–2208

Stan C, Straus DM (2009) Stratospheric predictability and sudden stratospheric warming events. J Geophys Res 114:1984–2012 (**D12103**)

Stanski HR, Wilson LJ, Burrows WR (1989) Survey of common verification methods in meteorology. WMO World Weather Watch Tech Rep 8 WMO TD 358:114

Sun LT, Robinson WA, Chen G (2012) The predictability of stratospheric warming events: more from the troposphere or the stratosphere? J Atmos Sci 69:768–783

Swets JA (1973) The relative operating characteristic in psychology. Science 182:990–1000

Thompson DWJ, Wallace JM (2001) Regional climate impacts of the northern hemisphere annular mode. Science 293:85–89

Thompson DWJ, Baldwin MP, Wallace JM (2002) Stratospheric connection to Northern Hemisphere wintertime weather: implications for prediction. J Clim 15:1421–1428

Townsend RD, Johnson DR (1985) A diagnostic study of the isentropic zonally averaged mass circulation during the first GARP global experiment. J Atmos Sci 42:1565–1579

Tripathi OP, Baldwin MP, Charlton-Perez AJ et al (2015) The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. Q J R Meteorol Soc 141:987–1003. https://doi.org/10.1002/qj.2432

Weber NJ, Clifford FM (2017) Evaluating CFSv2 subseasonal forecast skill with an emphasis on tropical convection. Mon Weather Rev 145:3795–3815

Wilks DS (2011) Statistical methods in the atmospheric sciences. Elsevier, Amsterdam

Yoden S, Ishioka K, Durran D et al (2014) Theoretical aspects of variability and predictability in weather and climate systems. Bull Am Meteorol Soc 95:1101–1104

Yu YY, Cai M, Ren RC, Van den Dool H (2015a) Relationship between warm air mass transport into upper polar atmosphere and cold air outbreaks in winter. J Atmos Sci 72:349–368

Yu YY, Ren RC, Cai M (2015b) Dynamical linkage between cold air outbreaks and intensity variations of the meridional mass circulation. J Atmos Sci 72:3214–3232

Yu YY, Cai M, Ren RC (2018a) A stochastic model with a low-frequency amplification feedback for the stratospheric northern annular mode. Clim Dyn 9–10:3757–3773. https://doi.org/10.1007/s00382-017-3843-2

Yu YY, Cai M, Ren RC et al (2018b) A closer look at the relationships between meridional mass circulation pulses in the stratosphere and cold air outbreak patterns in northern hemispheric winter. Clim Dyn. https://doi.org/10.1007/s00382-018-4069-7

Yu YY, Cai M, Shi CH, Ren RC (2018c) On the linkage among strong stratospheric mass circulation, stratospheric sudden warming, and cold weather events. Mon Weather Rev. https://doi.org/10.1175/MWR-D-18-0110.1

Zhang Q, Shin CS, van den Dool H, Cai M (2013) CFSv2 prediction skill of stratospheric temperature anomalies. Clim Dyn 41:2231–2249