Variations in the Frequency of Stratospheric Sudden Warmings in CMIP5 and CMIP6 and Possible Causes

ZHENG WU[a] AND THOMAS REICHLER

Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah

(Manuscript received 15 February 2020, in final form 3 September 2020)

ABSTRACT: The climatological frequency of stratospheric sudden warming events (SSWs) is an important dynamical characteristic of the extratropical stratosphere. However, modern climate models have difficulties in simulating this frequency, with many models either considerably under- or overestimating the observational estimates. Past research has found that models with a higher upper lid tend to simulate a higher and more realistic number of SSWs. The present study revisits this issue and investigates causes for biases in the simulated SSW frequency from the CMIP5 and CMIP6 models. It is found that variations in the frequency are closely related to 1) the strength of the polar vortex and 2) the upward-propagating wave activity in the stratosphere. While it is difficult to explain the variations in the polar vortex strength from the available model output, the stratospheric wave activity is influenced by different aspects of the climatological mean state of the atmosphere in the lower stratosphere. We further find that models with a finer vertical resolution in the stratosphere are overall more realistic: vertical resolution is associated with a smaller cold bias above the extratropical tropopause, more upward-propagating wave activity in the lower stratosphere, and a higher frequency of SSWs. We conclude that not only a high model lid but also a fine vertical resolution in the stratosphere is important for simulating the dynamical variability of the stratosphere.

KEYWORDS: Atmospheric circulation; Stratosphere-troposphere coupling; Stratosphere; Climate models; Model evaluation/performance; Model output statistics

1. Introduction

Stratospheric sudden warming events (SSWs) represent extreme perturbations of the wintertime Arctic stratosphere and contribute significantly to the predictability of tropospheric weather (Sigmond et al. 2013; Tripathi et al. 2015; Karpechko et al. 2017). Observations show that SSWs occur on average in 6 out of 10 years, a frequency that is an essential characteristic of the coupled stratosphere–troposphere system (Butler et al. 2019). Given this fundamental nature, it is surprising that even the most sophisticated models used for climate assessments have major difficulties in reproducing this frequency. This is demonstrated in Fig. 1, comparing the SSW frequencies of models from phases 5 and 6 of the Coupled Model Intercomparison Projects (CMIP5 and CMIP6). The simulated frequencies vary widely, with some models producing no or almost no SSWs, and others simulating more than one SSW per year. Similar variations have been reported from other model intercomparison studies (Charlton et al. 2007; SPARC 2010; Butchart et al. 2011; Charlton-Perez et al. 2013; Ayarzagüena et al. 2018, 2020). With this in mind, one may ask why models have such difficulties in reproducing the observed number of SSWs, and which aspect of a model needs to be improved to correct the situation. Answering this question and reducing the SSW biases of models may lead to better climate predictions, as there is currently neither robust evidence nor a clear consensus on how the number of SSWs will respond to climate change (Ayarzagüena et al. 2018).

One difficulty in reproducing the observed SSW frequency is the relatively rare nature of the events, creating large interannual variability and necessitating relatively long time series to achieve statistically representative results. However, there is much more to simulating a realistic SSW frequency. For example, SSWs depend on the upward propagation of planetary waves into the polar vortex region, and the interaction of the waves with the mean vortex winds. The wave propagation is sensitive to subtle variations in the stratospheric base state, as expressed by the index of refraction for planetary waves (Matsuno 1970). The index contains higher derivatives of winds and temperatures, which are difficult to simulate correctly. The model generated spectrum of planetary waves, model resolution, and uncertainties about the parameterization of gravity wave drag only add to the list of complicating factors.

Charlton-Perez et al. (2013) and other studies have shown that models with a low upper lid tend to underestimate the SSW frequency. Lee and Black (2015) indicated that a low model top underestimates the variability of the stratospheric planetary wave activity and of the polar vortex strength. Shaw and Perlwitz (2010) demonstrated that a low-top model creates excessive wave damping near the upper boundary, reduced extreme heat flux events, and too few SSWs. However, a high model top alone does not guarantee a realistic SSW frequency (e.g., the two GFDL models in Fig. 1), and some models with a low lid simulate reasonable frequencies (e.g., CNRM-CM5 and BCC-ESM1 in Fig. 1).

[a] Current affiliation: Institute of Atmospheric and Climate Science IAC ETH, The Swiss Federal Institute of Technology (ETHZ), Zurich, Switzerland.

Corresponding author: Zheng Wu, zheng.winnie.wu@utah.edu

DOI: 10.1175/JCLI-D-20-0104.1

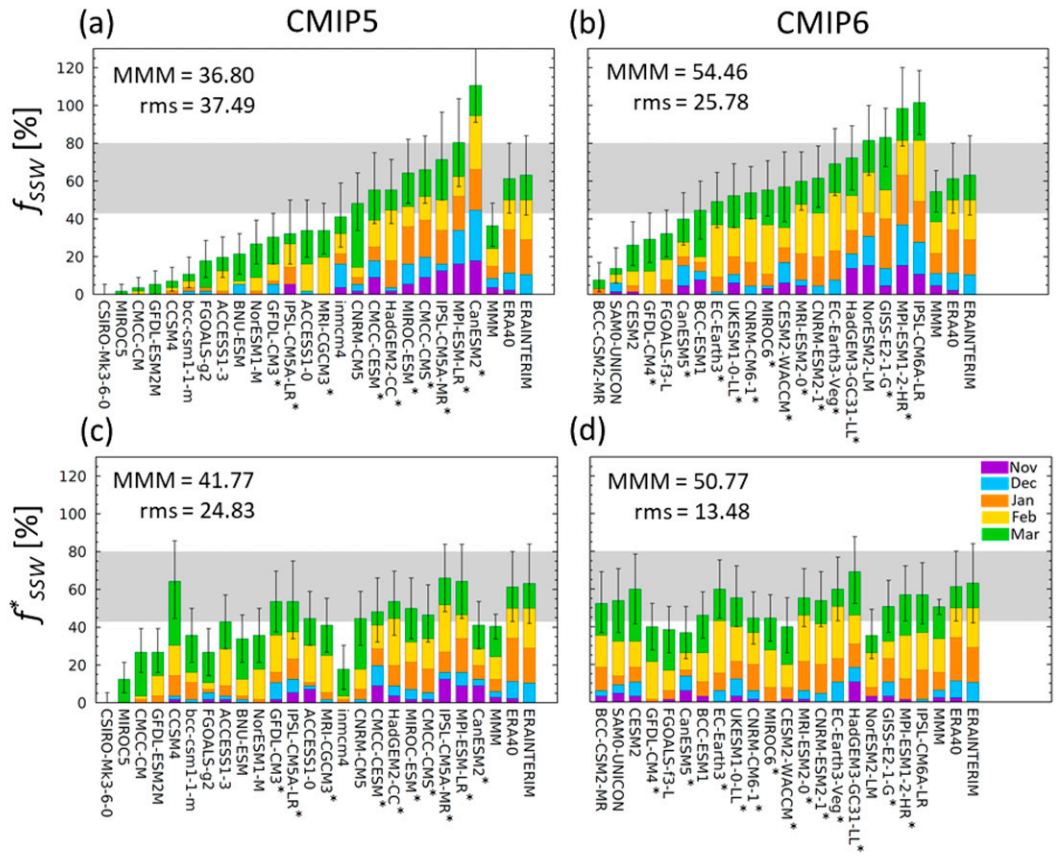

FIG. 1. Seasonal and monthly distribution of the SSW frequency (f_{SSW}) of CMIP5 and CMIP6 models, where f_{SSW} is calculated (a),(b) from the raw model output and (c),(d) after replacing each model's daily zonal-mean zonal wind climatology at 10 hPa and 60°N (U_{1060}) with that of the ERA-40. The last three columns in each panel show the multimodel mean (MMM), the ERA-40 (1958–2001), and the ERA-Interim (1979–2016), respectively. Whiskers denote 95% confidence intervals. The gray bar highlights the 95% confidence interval of the ERA-40. Colors show monthly breakdowns. Model names on the x axes followed by an asterisk are "high-top" models with a lid above 1 hPa.

It has also been suggested that the simulated SSW frequency is linked to climatological characteristics of a model. Examples for such characteristics include the meridional heat flux in the lower stratosphere (Charlton et al. 2007) and the strength of the polar vortex (Charlton-Perez et al. 2013). Horan and Reichler (2017) demonstrated that the seasonal variations in the SSW frequency can be accurately explained from the seasonal cycle statistics of the polar vortex strength (i.e., its mean, standard deviation, and skewness). Similarly, Taguchi (2017) related the intermodel spread of the SSW frequency in CMIP5 to the polar vortex strength and its variability.

The region just above the extratropical tropopause, the so-called upper troposphere–lower stratosphere (UTLS) region, where the vertical profiles of temperature, wind, and chemical constituents change dramatically (Gerber and Manzini 2016), is also relevant for SSWs. For example, Chen and Robinson (1992) highlighted the important role of the tropopause in regulating the upward propagation of planetary waves into the stratosphere. Various other studies came to similar conclusions (Shaw et al. 2014; Jucker 2016; de la Cámara et al. 2017; Martineau et al. 2018). Sjoberg and Birner (2014) suggested that the tropopause inversion layer could potentially be a source of wave activity in the lower stratosphere, and Birner and Albers (2017) found that the tropopause layer is crucial for a nonlinear wave–mean flow feedback that can cause SSWs. There is also consensus that sufficient vertical resolution is one of the essential criteria to correctly locate the tropopause and represent the dynamical processes in its vicinity (Birner 2006; Hegglin et al. 2010; Gettelman et al. 2011; Gerber and Manzini 2016). In a recent study, Wang et al. (2019) showed how increased vertical resolution improves the representation of the tropical tropopause in the WACCM model.

The goal of the present study is to revisit the SSW frequency problem and identify some of the processes that are involved in the simulation of the correct frequency. This is accomplished by examining the CMIP5 and CMIP6 ensembles and relating their intermodel spread in SSW frequency to variations in other basic quantities. As we will show, the strength of the polar vortex is the most important indicator for the number of SSWs, but the simulation of the climatological mean state in the UTLS region is also important. We further find that a finer vertical model resolution in the stratosphere tends to improve the simulation of the climatological mean state and the upward wave activity flux. Last, this study also sheds some new light on the performance of the two latest generations of climate models in the stratosphere.

The paper is organized as follows. Section 2 describes our data and diagnostic methods. The main results are described in section 3, identifying how the model simulated SSW frequency is related to other climatological characteristics, and explaining the underlying dynamical mechanisms. The paper ends with section 4, providing a summary, a discussion, and conclusions.

2. Models and analysis

a. Models

The data for this study consist of daily and monthly variables from the historical experiments of the CMIP5 (1950–2005) and CMIP6 (1950–2014) projects. As shown in Fig. 1, we investigate 23 CMIP5 models and 20 CMIP6 models, based on the availability of daily model output. We use only one ensemble member for each model, which is in most cases r1i1p1 (CMIP5) or r1i1p1f1 (CMIP6). Our analysis is carried out for November–March (NDJFM) since this is the time when SSWs are defined. The model climatologies are validated against ERA-40 (Uppala et al. 2005), which covers a similar time period (1958–2001) as the model simulations. We also use the more modern reanalysis ERA-Interim (1979–2016; Dee et al. 2011) to repeat some of our analysis, leading to very similar results; we therefore present our results only for ERA-40. To define SSWs and calculate the Eliassen–Palm (EP) fluxes (Eliassen and Palm 1961), we employ daily output of three-dimensional (longitude, latitude, level) zonal wind, meridional wind, and temperature. The daily output from the CMIP models, however, is only archived at 8 vertical pressure levels (1000, 850, 700, 500, 250, 100, 50, 10 hPa). This somewhat limits our ability to calculate reliable EP fluxes in the stratosphere, because doing so involves taking vertical derivatives. The monthly model output, however, is available for at least 17 pressure levels, with ~9 levels (250, 200, 150, 100, 70, 50, 30, 20, 10 hPa) in the stratosphere. We therefore calculate quantities that do not require daily data from monthly data.

b. Analysis

Our SSW definition follows the criterion of Charlton and Polvani (2007), based on the reversal of the daily zonal-mean zonal wind at 10 hPa and 60°N (U_{1060}), and a return to westerlies afterward for at least 10 consecutive days. Two wind reversals in the same season are treated as distinct SSWs if they are separated by at least 20 days. The SSW frequency (f_{SSW}) is the number of SSWs per 100 years. Based on ERA-40 over the period 1958–2001, the observation-based climatological SSW frequency is 61% (Butler et al. 2017), or in other words, the probability of an SSW in any given year is $\hat{p} = 0.61$. In calculating a confidence interval for $\hat{p}$, we apply a nonparametric bootstrapping method similar to Ayarzagüena et al. (2020). We perform the bootstrapping on individual years with replacement, using the same number of years as the actual number of years. We repeat 50 000 times, compute each time f_{SSW}, and derive the 2.5th–97.5th-percentile range of these samples. The 95% confidence interval of f_{SSW} in ERA-40 is 43%–80%. The intervals for the models are calculated in the same way.

SSWs are defined using the zero-wind threshold, and one reason for biases in crossing this threshold is the simulated climatological strength of the vortex. To estimate the contribution from this, we follow Scaife et al. (2010) and correct a model's climatological mean vortex strength by the reanalysis climatological U_{1060} but retain the variability of the vortex. We then use the corrected U_{1060} time series to detect SSWs and derive a new SSW frequency.

We employ EP flux diagnostics to analyze the wave–mean flow interaction. More specifically, we calculate the horizontal and vertical components of the quasigeostrophic version of the EP flux vector $\mathbf{F}$ (Edmon et al. 1981) in pressure coordinates using daily data. A complicating issue arises when calculating the EP fluxes from vertically coarsely resolved model output. To assess the impact of coarse vertical resolution, we use daily ERA-40 data to calculate the EP fluxes from using the same pressure levels that are available for the daily model data. The result is then compared to the EP fluxes from using the original data with all pressure levels. We find that the vertical component of the EP flux (FZ) from the vertically reduced data consistently underestimates the fluxes compared to using data from all levels. To further examine the influence of coarse vertical resolution on the model-to-model variation in CMIP5, we calculate the EP fluxes from monthly data twice, first, by using all levels, and second, by using only a subset of levels. The results suggest that the variations of FZ across the 23 models calculated from using 1) sublevels and 2) all levels are very similar. To calculate the upward wave activity, we could have also replaced the vertical EP fluxes with the poleward eddy heat fluxes, which do not necessitate calculating vertical derivatives. We repeated our analysis using heat fluxes and found very similar results. However, Taguchi (2017) suggested that FZ is more appropriate than heat fluxes in a multimodel comparison because variations in static stability are also important for the propagation of the waves. We therefore use FZ to diagnose the waves and their forcing.

We use the 40°–70°N latitude-weighted climatological FZ at 100 hPa (FZ_{100}) for the strength of the stratospheric wave driving, and the zonal-mean zonal wind at 10 hPa and 60°N (U_{1060}) for the strength of the stratospheric polar vortex. The propagation of planetary waves depends on the atmospheric refractive properties, expressed in terms of an unitless quasi-geostrophic index of refraction squared (RI^2) (Matsuno 1970). Assuming the planetary waves are stationary, RI^2 is given by

$$\mathrm{RI}^2 = \frac{a_e^2 \overline{q}_y}{U} - \frac{k^2}{\cos^2\varphi} - \left(\frac{a_e f}{2\mathrm{NH}}\right)^2, \tag{1}$$

where

$$\overline{q}_y = \frac{1}{a_e}\frac{\partial \overline{q}}{\partial \varphi} = \beta - \frac{1}{a_e^2}\frac{\partial}{\partial \varphi}\left[\frac{1}{\cos\varphi}\frac{\partial (U\cos\varphi)}{\partial \varphi}\right] + \frac{f^2}{HN^2}\frac{\partial U}{\partial z} - \frac{f^2}{N^2}\frac{\partial^2 U}{\partial z^2} - f^2\frac{\partial U}{\partial z}\frac{\partial}{\partial z}\left(\frac{1}{N^2}\right) \tag{2}$$

TABLE 1. CMIP5 and CMIP6 models and their mean vertical resolution ($r_{vertical}$) in the stratosphere between 100 and 1 hPa (or the model top, whichever is lower).

CMIP5	$r_{vertical}$ (m)	CMIP6	$r_{vertical}$ (m)
CMCC-CMS	733	MPI-ESM1–2-HR	768
MIROC-ESM	750	MIROC6	871
HadGEM2-CC	1900	MRI-ESM2–0	1075
MPI-ESM-LR	2015	EC-Earth3	1112
CMCC-CESM	2015	EC-Earth3-Veg	1112
MIROC5	2065	CNRM-CM6–1	1112
GFDL CM3	2149	CNRM-ESM2–1	1112
ACCESS1.0	2238	UKESM1–0-LL	1151
ACCESS1.3	2238	HadGEM3-GC31-LL	1151
IPSL-CM5A-LR	2303	IPSL-CM6A-LR	1273
IPSL-CM5A-MR	2303	BCC-CSM2-MR	1345
MRI-CGCM3	2303	CESM2-WACCM	1600
INM-CM4	2686	CanESM5	1639
BCC-CSM1.1-m	3093	NorESM2-LM	2455
BNU-ESM	3093	GISS-E2–1-G	2931
CMCC-CM	3224	FGOALS-f3-L	2983
CNRM-CM5	3224	SAM0-UNICON	3000
CCSM4	3342	BCC-ESM1	3093
NorESM1-M	3342	CESM2	3300
FGOALS-g2	3342	GFDL-CM4	3582
CanESM2	3582		
CSIRO-Mk3.6.0	7236		
GFDL-ESM2M	7724		

is the meridional gradient of potential vorticity, a_e is the radius of Earth, $k = 1$ is the zonal wavenumber of the wave, φ is latitude, f is the Coriolis parameter, H is the scale height, and N is the buoyancy frequency. The thermally defined tropopause pressure is calculated from three-dimensional monthly temperature fields, following the algorithm of Reichler et al. (2003).

We follow Charlton-Perez et al. (2013) and use the pressure of the highest model level to classify the models into high-top (above 1 hPa) and low-top (below 1 hPa) models. An asterisk after the model name in Fig. 1 indicates models with a high top. With this definition, 10 out of 23 CMIP5 models and 13 out of 20 CMIP6 models are high-top models. The mean vertical resolution in the stratosphere ($r_{vertical}$) of each model (Table 1) is derived by dividing the vertical distance between 100 and 1 hPa (or the model top, whichever is lower) by the number of native model levels in this layer. We also calculate the vertical resolution only in the 300–100-hPa layer (not shown) and find that in both ensembles it is strongly correlated with $r_{vertical}$ (r = 0.7 in CMIP5 and 0.8 in CMIP6).

Throughout our study, we use linear regression to explain variations in f_{SSW} from the intermodel spread in some other climatological quantities. To be more specific, we first use a one-predictor (unary) linear regression to link intermodel variations in f_{SSW} to one quantity. Then, we use bilinear regression to estimate f_{SSW} from the original quantity and one additional quantity. From this, we calculate the increase in explained variance. We use bootstrapping to determine the statistical significance of the explained variance and its increase by choosing 50 000 random model combinations with replacement, with each combination containing the same number of models as the original sample. The regressions reported in this study are all statistically significant at the 95% limit, unless otherwise specified.

3. Results

Figure 1 shows the monthly and seasonal distribution of the SSW frequency of the CMIP5 and CMIP6 models, their multimodel means (MMM), ERA-40, and ERA-Interim. Overall, there is a large spread in f_{SSW} among both groups of models, ranging from 0% to more than 100%. The mean f_{SSW} is 37% in CMIP5 and 54% in CMIP6, with a standard deviation of 29% and 25%, respectively. This can be compared to f_{SSW} = 61% (±18%) in ERA-40 (1958–2001) and f_{SSW} = 63% (±21%) in ERA-Interim (1979–2016). The models' root-mean-square (rms) error with respect to the ERA-40 is 37% (CMIP5) and 26% (CMIP6). Just considering the high-top models from each ensemble reduces the rms error to 24% in CMIP5 and 17% in CMIP6. If consistency with ERA-40 is measured by the overlap of a model's confidence interval with that of ERA-40 (gray bar), then 48% of the CMIP5 and 75% of the CMIP6 models fall into this category. Therefore, CMIP6 is closer to ERA-40, with fewer extreme outliers compared to CMIP5. In particular, more than two-thirds of the CMIP5 models have a lower frequency than ERA-40. Models with a high overall f_{SSW} tend to produce more early events, in November (purple) and December (blue). In ERA-40, most SSWs occur in January (orange), while the models produce most events later in winter, consistent with Horan and Reichler (2017). For example, in the MMM, March events (green) comprise 33% of all SSWs in CMIP5 and 29% in CMIP6, whereas in ERA-40 March events amount only to 19%.

As stated earlier, errors in the simulated f_{SSW} are associated with biases in the climatological mean strength and variability (on daily to interannual time scales) of the polar vortex. To assess to what extent the mean vortex strength is responsible, we recalculate f_{SSW} after replacing the U_{1060} daily climatology of each model with that of ERA-40 (Figs. 1c,d), and the remaining differences in f_{SSW} must be due to biases in vortex variability alone. Overall, f_{SSW} after the correction is in better agreement with ERA-40. In CMIP6, f_{SSW} improves in most cases, indicating that the models' vortex variability is realistic and that biases in vortex strength are the main reason for their f_{SSW} errors. In contrast, even after the correction, many of the CMIP5 models still produce a too low f_{SSW}. In fact, in around 57% of the CMIP5 models, f_{SSW} is still outside the observed range, and this must be caused by biases in vortex variability. Figure 1 also indicates that correcting the vortex strength improves the total f_{SSW}, but this does not guarantee a realistic seasonal distribution of SSW. The correction leads to a substantial increase in late events for models with too-low overall frequencies (e.g., CCSM4, CESM2). On the other hand, after the correction, models with too high SSW frequencies (e.g., CanESM2, MPI-ESM1–2-HR) undergo a considerable reduction in December SSWs.

To summarize, CMIP5 and CMIP6 have somewhat different reasons for their biases in f_{SSW}, motivating us to examine the two groups of models separately from each other. Doing so may also shed some new light on the recent progress in climate model development, as Fig. 1 already indicates that in terms of f_{SSW} CMIP6 performs better than CMIP5.

a. *Variations in polar vortex strength and wave activity*

As discussed in the introduction, we follow earlier work (Horan and Reichler 2017; Taguchi 2017; Martineau et al. 2018) and relate the climatological SSW frequency to the variability and strength of the polar vortex, and the vortex variability to the mean upward wave flux (Wu and Reichler 2019). In fact, Wu and Reichler (2019) demonstrated that the SSW frequency can be well estimated from the climatological strength of the polar vortex and upward wave activity fluxes. In the following analysis, we will pursue similar ideas by examining the vortex strength and wave activity to explain variations in SSW frequency.

Figures 2a and 2b show how variations in f_{SSW} relate to U_{1060}. In this and the following figures, we consistently use color to represent individual models, with reddish for models with high f_{SSW} and blueish for models with low f_{SSW}. As expected, the correlations between f_{SSW} and vortex strength are negative, with a tighter relationship for CMIP6 ($r = -0.87$) than for CMIP5 ($r = -0.62$). In CMIP6, U_{1060} alone is a very good indicator for f_{SSW}. Many of its models are clustered around ERA-40 (Fig. 2b, black marker), indicating a reasonable performance. Also, the marker for ERA-40 is located at the regression line, another indicator for the consistency of CMIP6 with the reanalysis. In CMIP5, the linear fit between f_{SSW} and U_{1060} underestimates f_{SSW} of the "high-frequency" models (reddish markers) and also of the ERA-40. In the MMM (gray markers) and compared to ERA-40, the vortex of both ensembles is too strong and f_{SSW} is too low. The tendency of climate models to have overly strong polar vortices has been the subject of previous studies (e.g., Charlton et al. 2007), and Shaw et al. (2014) suspected that this could be due to a misrepresentation of the stratospheric gravity wave drag.

One may argue that variations in f_{SSW} impact the climatological polar vortex strength, so that the vortex strength cannot be considered as an independent estimator for f_{SSW}. To address this concern, we repeat our analysis by excluding years with SSWs. In this case, the correlations are reduced but still sizeable ($r = -0.44$ in CMIP5, $r = -0.77$ in CMIP6) and significant at the 95% limit, suggesting that U_{1060} contains indeed distinct and independent information about the likelihood of SSWs. Therefore, we include in our subsequent analysis data from all years.

Figures 2c and 2d show, for each model, vertical profiles of the 40°–70°N averaged upward wave fluxes (FZ) relative to the ERA-40. In general, the intermodel variations increase with altitude, are larger in CMIP5 than in CMIP6, and are vertically quite uniform in CMIP6. The MMM of FZ (dashed black) is larger than ERA-40 (solid black) in the troposphere and smaller in the stratosphere, indicating that processes in the lower stratosphere limit the amount of wave activity that enters from below into the stratosphere. Of note is that stratospheric FZ is particularly underestimated in CMIP5. When comparing all panels, it becomes clear that models with low f_{SSW} (blueish and purplish colors) not only tend to be associated with a stronger U_{1060} but also with a reduced stratospheric wave activity compared to ERA-40.

b. *Estimating the SSW frequency*

In this section, we explore the usefulness of other climatological quantities beside U_{1060} as "predictors" for the intermodel variations in SSW frequency. As shown in Fig. 2b, in CMIP6, U_{1060} alone is already an excellent indicator for f_{SSW}, making it difficult to make further improvements by including additional predictors in the linear regression.

Figure 3 shows for CMIP5 the actual versus the estimated f_{SSW} from linear regression, using the predictors indicated at the top of each panel. U_{1060} alone explains 38% of the total variance (Fig. 3a), with large errors at the two extremes of the distribution. The stratospheric wave driving (FZ_{100}) alone explains only 28% (Fig. 3b). We note that using FZ at levels higher than 100 hPa improves the explained variance (r^2) (not shown) because of a closer relationship with the vortex at 10 hPa. However, we prefer FZ_{100} as predictor because traditionally the 100-hPa level is considered as gateway for planetary waves entering the stratosphere, and the wave activity flux at 10 hPa is highly correlated with FZ_{100}.

Based on the ideas of Wu and Reichler (2019), we next combine the effects of U_{1060} and FZ_{100} by employing a bilinear regression to estimate f_{SSW} (Fig. 3c). The explained variance is now 70%, much improved over the single predictor regressions, and the regressions now also correctly predict f_{SSW} of the ERA-40. As expected, the regression coefficient of U_{1060} is negative and of FZ_{100} is positive (not shown), consistent with previous results (Jucker et al. 2014; Wu and Reichler 2019) that a weaker polar vortex and stronger stratospheric wave driving favor more SSWs.

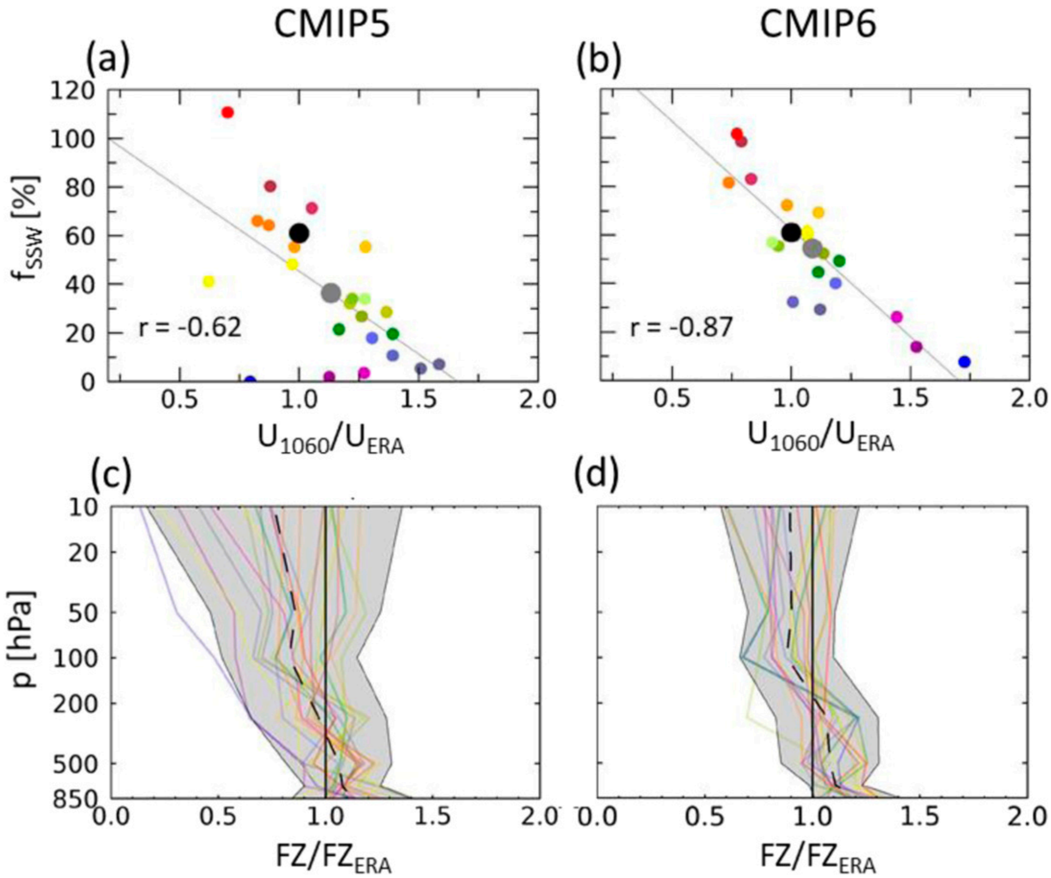

FIG. 2. (a),(b) Model variations in polar vortex strength and vertical EP fluxes, showing individual models in terms of their f_{SSW} and normalized U_{1060}. Black dots represent ERA-40, gray dots are the MMM, and the gray line is the regression of f_{SSW} on U_{1060}, with the correlation coefficient r indicated at the bottom left. (c),(d) Vertical profiles of the normalized vertical EP flux (FZ) for individual models. Gray areas highlight the ± 2 standard deviation among the models; solid black line is ERA-40 and dashed black line is the MMM. In all panels, the normalization factors for U_{1060} and FZ are the respective climatologies from ERA-40. Colors represent specific models, ordered by ascending f_{SSW}, as in Fig. 1; reddish corresponds to higher and blueish to lower f_{SSW}.

The bar plots in Fig. 4a summarize our results by showing the explained variances (r^2) from the linear regressions for CMIP5 and CMIP6. In a unary linear regression, the predictive power of both U_{1060} (blue) and FZ_{100} (orange) is stronger in CMIP6, and U_{1060} is the single best predictor for f_{SSW} in both ensembles. Combining U_{1060} and FZ_{100} (gray) leads to a large improvement in r^2 in CMIP5. The improvement in CMIP6 is more subtle, likely because r^2 from U_{1060} alone is already high. It is also of note that the intermodel variations U_{1060} and FZ_{100} in CMIP6 are negatively correlated ($r = -0.47$), while in CMIP5 this is not the case ($r = 0.06$). In other words, in CMIP6, U_{1060} and FZ_{100} are not independent, helping us to understand the only small improvements in explained variance when both predictors are used in combination. A negative relationship between U_{1060} and FZ_{100} is also revealed from their interannual variations in ERA-40 (not shown) and idealized models (Jucker et al. 2014; Wu and Reichler 2019). In this respect, the negative correlation between U_{1060} and FZ_{100} in CMIP6 is more plausible and realistic than the near-zero correlation in CMIP5. Repeating this analysis for only the high-top CMIP5 models or the CMIP5 models with a relatively fine vertical resolution does not change much the small correlations between U_{1060} and FZ_{100}. This indicates that in CMIP5, reasons more complicated than the vertical structure of the model grid are responsible for the unphysical relationship between U_{1060} and FZ_{100}.

We use bootstrapping to test the robustness of our regressions. Figure 4b shows the distributions of the regression coefficients β, normalized by their own means, and derived from 50 000 regressions, in which we randomly determine the combination of models. Note that for the two bilinear regressions, we only present the coefficients of FZ_{100} (gray) and T_{TROPO} (green) (for definition see the caption of Fig. 3) since the ones of U_{1060} are in both cases highly significant. The regression coefficients are considered robust if their distributions are narrow, and they are significant if their 5th percentile is positive (i.e., of the same sign as the mean). As shown in Fig. 4b, this is the case for all regression models.

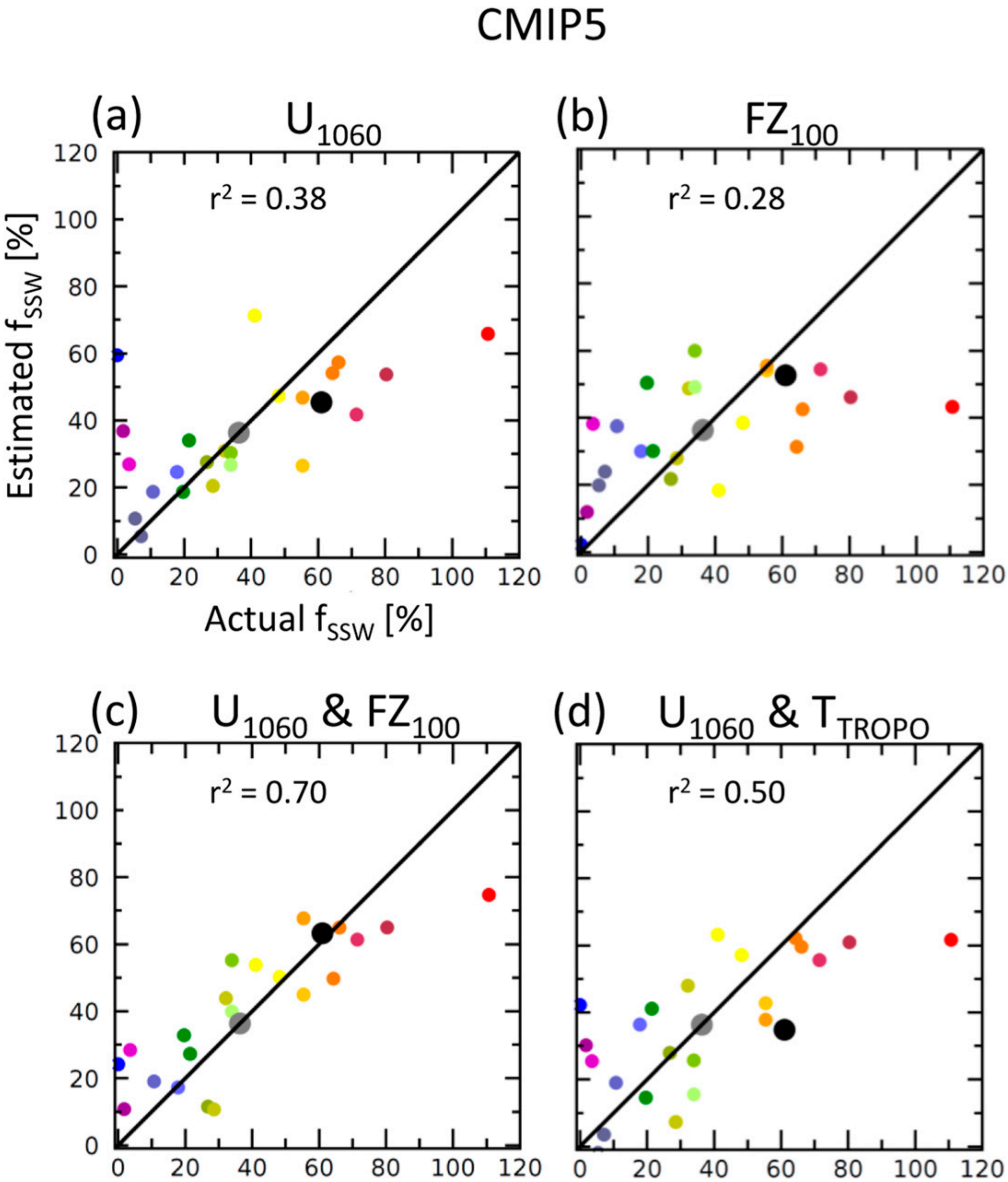

FIG. 3. Estimated vs actual f_{SSW} for CMIP5 from various regression models. The predictors are (a) U_{1060}, (b) upward EP flux at 100 hPa (FZ_{100}), (c) U_{1060} and FZ_{100}, and (d) U_{1060} and temperatures averaged along the tropopause north of 50°N (T_{TROPO}). Colors mark the same models as in Figs. 2a and 2c.

However, the fact that a bilinear regression coefficient passes the significance test does not imply that the second predictor adds real information. Therefore, we employ stepwise regression (Draper and Smith 1998, 307–312) to test whether the improvements of r^2 are significant at the 5% level when the second predictor is added. We find that for FZ_{100}, this is the case in 95% (86%) of all 50 000 bootstrapping combinations of CMIP5 (CMIP6).

Given that U_{1060} is the single best f_{SSW} predictor, we next ask whether, besides FZ_{100}, there are other more basic quantities that can be used as second predictors for the f_{SSW} intermodel spread. Figure 5 shows latitude–pressure cross sections of the r^2 differences (Δr^2) between a bilinear regression using U_{1060} and the quantity indicated at the top of each panel and a unary regression using U_{1060} alone. Note that the actual Δr^2 is always positive, and that the color shading of Δr^2 in the plots indicates the sign of the second regression coefficient. In other words, reddish colors indicate a positive correlation and blueish colors indicate a negative correlation between f_{SSW} and the second predictor. From Fig. 5, one can readily identify the locations where, in combination with U_{1060}, a second predictor improves explaining f_{SSW}.

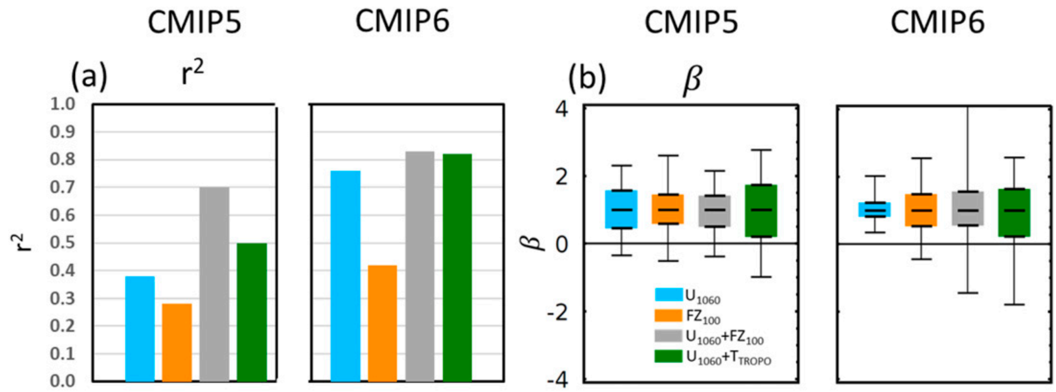

FIG. 4. Outcome of different regression models to explain f_{SSW}. (a) Explained variance r^2. (b) Distribution of the normalized regression coefficients β, obtained from bootstrapping (see text for details); the coefficients are normalized by their own mean; box and whiskers indicate the 5th–95th-percentile range, minimum, maximum, and mean of the regression coefficients. For the two bilinear regressions, only the β distributions for FZ_{100} (gray) and T_{TROPO} (green) are shown.

Figure 5a shows Δr^2 for CMIP5 from using climatological zonal-mean zonal wind (U) as second predictor. Stronger U in the midlatitudes favors the occurrence of SSWs, with the largest contributions in the UTLS region. Figure 5c indicates that in CMIP5, colder zonal-mean temperatures (T) in the UTLS region over the high latitudes also favor more SSWs. CMIP6 (right panels) shows similar relationships, even though the magnitude of Δr^2 is generally weaker. For example, colder T along the high-latitude tropopause (Fig. 5d), stronger U in the extratropics (Fig. 5b), and more upward wave fluxes (not shown) are all associated with higher f_{SSW}. Finally, Figs. 5e and 5f show the predictive capabilities of the index of refraction (RI^2), which conveniently explains the propagation of planetary waves and its dependence on the atmospheric background state. Planetary waves tend to propagate from regions of low RI^2 to regions of high RI^2 (Karoly and Hoskins 1982; Simpson et al. 2009). In both CMIP5 and CMIP6, there is a region centered at 40°N and 100 hPa where RI^2 is positively related to f_{SSW}.

c. Role of the index of refraction

To aid our subsequent analysis, we define from Fig. 5 three indices, each representing area averages of U, T, or RI^2 over specific regions: U_{TROPO} is 35°–55°N and 200–70 hPa, T_{TROPO} is 50°–90°N and 300–200 hPa, and RI^2_{TROPO} is 35°–45°N and 200–70 hPa. These three indices are all located in the lower extratropical stratosphere, consistent with various studies that the conditions in this region are important for the upward propagation of planetary waves from the troposphere. For example, Gerber and Polvani (2009) suggested that strong winds and potential vorticity gradients in the lower stratosphere are essential for bursts of planetary waves into the vortex region, and thus for the creation of SSWs.

We examine the spatial structure of RI^2, focusing on stationary wave one, the dominant wave component of the stratosphere. Figures 6a and 6b show that the MMM of RI^2 in both ensembles is fairly similar, and it is also close to that of the ERA-40 (not shown). There is a region of maximum RI^2 in the middle to upper troposphere, which acts to trap the wave activity and corresponds to large EP flux convergences (Chen and Robinson 1992). More importantly, there is a band of small RI^2 just above the extratropical tropopause, which is a potential barrier to upward wave propagation and is associated with wave evanescence (Sigmond and Scinocca 2010). As shown by the black rectangles, the minima in RI^2 in both ensembles coincide more or less with our definition for RI^2_{TROPO}. This region also contains some negative values for CMIP5. The MMM of RI^2_{TROPO} in CMIP6 is therefore somewhat larger than in CMIP5 (also see Figs. 7a,b). The cross sections of the intermodel spread in RI^2 (Figs. 6c,d) indicate that in both ensembles, the region of RI^2_{TROPO} is characterized by large modeling uncertainties. Given that the expression for RI^2 [Eq. (2)] is dominated by various derivatives of the zonal-mean zonal wind, the large spread in RI^2 is indicative for difficulties in simulating the wind structure in this region. This is confirmed by Figs. 6e and 6f, showing the local correlations between intermodel variations in U and RI^2. The correlations between RI^2_{TROPO} and U_{TROPO} are 0.8 in CMIP5 and 0.7 in CMIP6 (Table 2). In other words, over the RI^2_{TROPO} region, RI^2 is very sensitive to uncertainties in the simulated wind, and from additional analysis (not shown) we find that it is mostly the u_{zz} term in (2) that explains the influence of U on RI^2 in this region.

We next assess the role of RI^2_{TROPO} in influencing the frequency of SSWs. Figures 7a and 7b show RI^2_{TROPO} of the individual models plotted against their f_{SSW}. As expected, there is a positive relationship between RI^2_{TROPO} and f_{SSW}, which is best explained from the increase in the upward-propagating wave activity with increasing RI^2_{TROPO}. Indeed, Table 2 shows that there is a positive correlation between RI^2_{TROPO}, FZ_{100}, and FZ_{10}. In terms of T_{TROPO} (Figs. 7c,d), the MMM of both ensembles is ~3 K colder than ERA-40, but in terms of RI^2_{TROPO} and U_{TROPO} (not shown) both ensembles are about right. Also, U_{TROPO} and T_{TROPO} are only correlated at $r \sim 0.6$

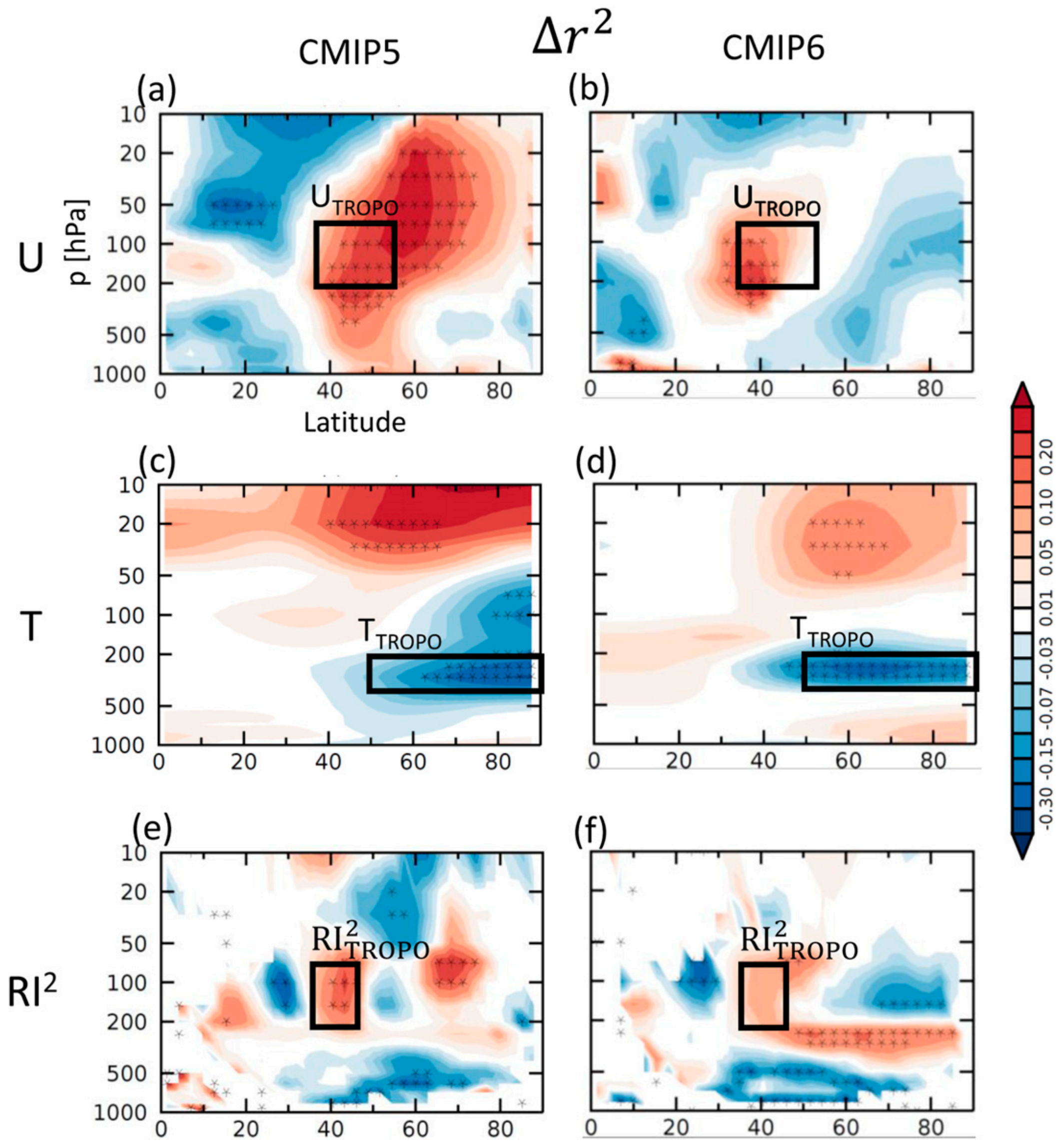

FIG. 5. Improvements in explained variance (Δr^2) over the unary U_{1060} regression from bilinear regression using U_{1060} in combination with other climatological mean quantities as second predictors. The quantities are (a),(b) zonal-mean zonal wind, (c),(d) zonal-mean temperature, and (e),(f) the index of refraction squared (RI^2). The r^2 from unary regression using U_{1060} is 0.38 in CMIP5 and 0.76 in CMIP6. For better visualization, the correlations are Fisher-z-transformed before differences are calculated. Since Δr^2 is always positive, color is used to indicate the sign of the underlying regression coefficient. Stippling indicates the improvements are significant at the 95% confidence level. The rectangular areas show the definitions of the U_{TROPO} (zonal-mean zonal wind for 35°–55°N and 200–70 hPa), T_{TROPO} (T for 50°–90°N and 300–200 hPa), and $\mathrm{RI}^2_{\mathrm{TROPO}}$ (RI^2 for 35°–45°N and 200–70 hPa) indices.

(Table 2), indicating a certain degree of independence between the two quantities. We note that a cold lowermost stratospheric temperature bias is not exclusive to CMIP5 and CMIP6. For example, Polichtchouk et al. (2019) pointed out that all operational forecast systems at ECMWF suffer from such a bias over the tropical lower stratosphere, which depends on the vertical resolution and is associated with an inadequate representation of gravity waves. Figures 7c and 7d, and also Table 2 further show that $\mathrm{RI}^2_{\mathrm{TROPO}}$ is negatively correlated with T_{TROPO}, a connection that may help explain why T_{TROPO} has an explanatory power for f_{SSW} (Figs. 5c,d). This issue will be explored next.

d. Relationship between tropopause temperature and SSW frequency

We begin by examining the predictive capabilities of T_{TROPO} for f_{SSW}. Getting back to Fig. 3, using T_{TROPO} on top of U_{1060} as a second predictor (Fig. 3d) improves the estimation of f_{SSW} in CMIP5, with r^2 increasing from 0.38 to 0.50. However, the

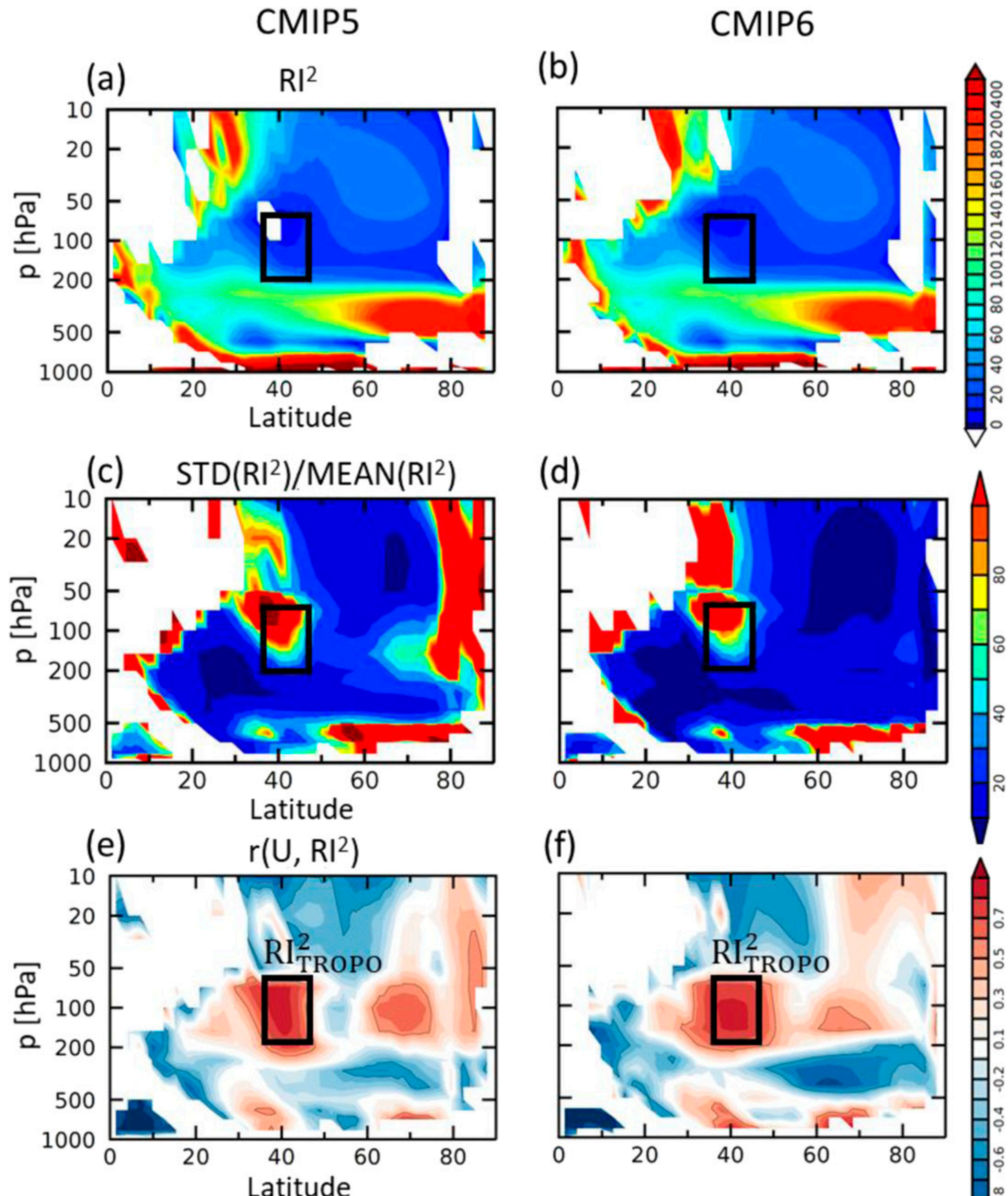

FIG. 6. Latitude–pressure cross sections of RI^2. (a),(b) Multimodel mean RI^2 (unitless) for stationary wave 1. (c),(d) Intermodel spread in RI^2 (%), given by the ratio of the intermodel standard deviation of RI^2 and the mean RI^2. (e),(f) Local correlations between U and RI^2; the shading interval is 0.1; values of ± 0.5 are contoured in black. Black rectangles indicate RI^2_{TROPO}.

increase is smaller than from using FZ_{100} and U_{1060} (Figs. 3c and 4a). In CMIP6, using T_{TROPO} as a second predictor also increases r^2, and this increase is similar to that from using FZ_{100} (Fig. 4a). The distributions of the regression coefficient β for T_{TROPO} are broader than for FZ_{100} (Fig. 4b), and in 61% (79%) of the bootstrapping cases T_{TROPO} contains additional information about the f_{SSW} estimates in CMIP5 (CMIP6). Therefore, FZ_{100} is a more robust second predictor than T_{TROPO} in both ensembles. Nevertheless, we think that the connection between T_{TROPO} and f_{SSW} deserves further explanation.

We believe that T_{TROPO} is connected to f_{SSW} because a colder high-latitude tropopause is associated with a background state that favors the upward propagation of planetary waves into the polar vortex, leading to an increase in f_{SSW}. This

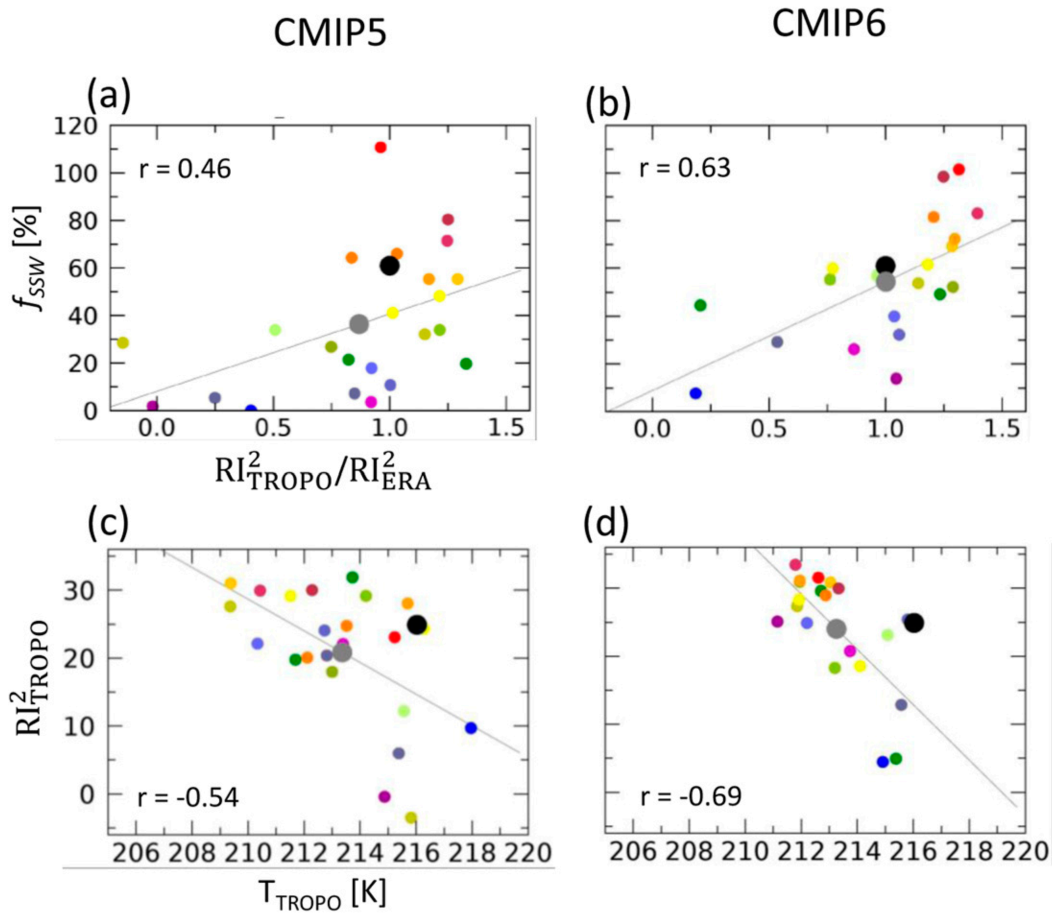

FIG. 7. Intermodel variations. (a),(b) RI^2_{TROPO} vs f_{SSW}; RI^2_{TROPO} is normalized by the value from ERA-40. (c),(d) T_{TROPO} vs RI^2_{TROPO}. Colors show individual models, as in Fig. 2; black represents the ERA-40 and gray the MMM. Regression lines are also shown, with correlation coefficients given in the plots.

is consistent with previous studies, which also indicate that a colder lower stratosphere is linked to stronger U and enhanced upward directed wave activity (Gerber and Polvani 2009; Schimanke et al. 2013; Shaw et al. 2014; Martineau et al. 2018). From the spatial correlations between T_{TROPO} and U we find a band of strong negative correlations that tilts upward and northward from the tropopause at 40°N into the stratosphere (not shown), as one would expect from the thermal wind relationship. This band is similar in structure to the improvements in r^2 from using U as a second predictor for f_{SSW} (Figs. 5a,b), suggesting that the improvements are related to variations in T_{TROPO}. As given in Table 2, the correlations between T_{TROPO} and U_{TROPO} are -0.6 in both ensembles because of the thermal wind relationship, leading to similar correlations between T_{TROPO}/U_{TROPO} and RI^2_{TROPO}. The correlation between T_{TROPO} and f_{SSW} is as expected negative, but it is also quite weak ($r \sim$ 0.2–0.3; Table 2), reflecting the complicated chain of nonlinear dynamical processes that link T_{TROPO} to f_{SSW}, and which involve, among others, RI^2_{TROPO} and FZ_{100}.

TABLE 2. Correlation of the intermodel spread in various indices. For each index, the top number is for CMIP5, and the bottom number is for CMIP6. Correlations followed by an asterisk (*) are based on absolute differences in T_{TROPO}, U_{TROPO}, and RI^2_{TROPO} with respect to their MMM. Correlations significant at the 95% level are shown in boldface and at the 90% level in italic.

	T_{TROPO}	U_{TROPO}	RI^2_{TROPO}	FZ_{100}	FZ_{10}	f_{SSW}
T_{TROPO}		**−0.6**	**−0.5**	*−0.4*	*−0.4*	−0.2
		−0.6	**−0.7**	**−0.6**	*−0.4*	−0.3
U_{TROPO}			**0.8**	**0.7**	*0.4*	0.3
			0.7	**0.6**	0.1	**0.5**
RI^2_{TROPO}				**0.7**	*0.4*	**0.5**
				0.8	*0.4*	**0.6**
$r_{vertical}$	0.3*	0.1*	0.1*	**−0.6**	**−0.6**	*−0.4*
	0.6*	0.3*	0.1*	**−0.5**	**−0.7**	*−0.4*

We want to point out that our analysis of Table 2 and Fig. 7 is solely based on correlations, which do not establish causality. To partly address the question of cause and effect between T_{TROPO}/U_{TROPO} and FZ, we examine lagged correlations from daily ERA-40 (Fig. 8). The correlations are generally small ($r = -0.2$), but given that these are daily data for 44 years, the correlations are significant. The correlations are strongest when T_{TROPO} leads by about one week (Fig. 8a), suggesting

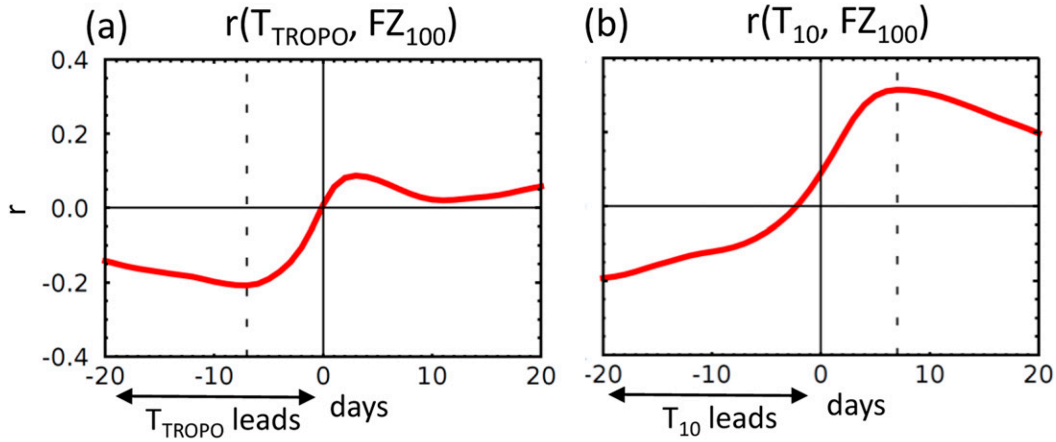

FIG. 8. ERA-40 lagged correlation between (a) FZ_{100} and T_{TROPO} and (b) FZ_{100} and T_{10} (the 10-hPa temperatures averaged from 50° to 90°N), using daily data from November to March. Dashed lines mark the time of the strongest correlation.

that variations in T_{TROPO}/U_{TROPO} from synoptic variability indeed influence the upward propagation of planetary waves into the stratosphere. On the other hand, the correlations between FZ_{100} and T_{10}, the 10-hPa temperatures averaged from 50°–90°N, peak when FZ_{100} leads (Fig. 8b), indicating that midstratospheric warming is a consequence of upward-propagating waves and their poleward heat transports.

e. The role of vertical model resolution

The final question we attempt to address is whether perhaps a model's vertical resolution is connected to the simulation of stratospheric variability, similar to the finding that the location of a model's top is important (Shaw and Perlwitz 2010; Charlton-Perez et al. 2013). First, we investigate potential causes for the biases in T_{TROPO}. Such biases must be accompanied by variations in the height of the thermally defined tropopause, as has been previously reported for CMIP5 (Kim et al. 2013; Ao et al. 2015). The mean tropopause pressure (Figs. 9a,b) has a larger intermodel spread in CMIP5 than in CMIP6, but the MMM of it (dashed black) is similar between the two ensembles and low-biased with respect to the reanalysis (solid black). The upward shifted tropopause corresponds to a cold bias above and/or a warm bias below the tropopause. Similar cold biases and too high tropopause heights have been reported before (SPARC 2010; Charlton-Perez et al. 2013), indicating that these are common climate model problems.

As mentioned in the introduction, sufficient vertical resolution is an important prerequisite for properly representing the UTLS region, the location of the tropopause, and the propagation of waves. We use the vertical resolution in the stratosphere ($r_{vertical}$) to examine potential influences of it on the simulation of T_{TROPO}. Figures 9c and 9d show the correlation between $r_{vertical}$ and the absolute temperature difference ($|\Delta T|$) with respect to the MMM at each grid point. Here, we use absolute differences because a coarse vertical resolution can lead to either upward or downward displacements of the tropopause. The region of reddish colors above the extratropical tropopause (~250 hPa) indicates that a coarse resolution (larger $r_{vertical}$) is indeed associated with larger errors in T_{TROPO}. However, the correlation of $r_{vertical}$ with U_{TROPO} and RI^2_{TROPO} is much smaller (Table 2), perhaps because of the complicated nonlinear nature of RI^2_{TROPO}. Repeating this analysis using the vertical resolution in the UTLS region leads to very similar results (not shown).

Last, we examine how $r_{vertical}$ is connected to the simulated stratospheric variability. Table 2 demonstrates that coarse vertical resolution is associated with reduced upward wave propagation (FZ_{100} and FZ_{10}) and f_{SSW}. However, $r_{vertical}$ is only weakly correlated with $|\Delta U_{TROPO}|$ and $|\Delta RI^2_{TROPO}|$, suggesting that the associations between $r_{vertical}$ and FZ_{100}, FZ_{10}, and f_{SSW} are more related to the improved representation of stratospheric wave dynamics when the vertical resolution is high, and not so much to the influence of resolution on biases in T_{TROPO} and U_{TROPO}.

4. Summary, discussion, and conclusions

We investigate the intermodel variations in the SSW frequency simulated by the two latest generations of climate models, CMIP5 and CMIP6. Both generations of models show a wide spread in their frequencies, with some models producing no SSW, and others having more than one SSW per year. This raises concerns about the models' ability to simulate a realistic stratospheric circulation variability, the downward influence of this variability into the troposphere, and potential shift of this variability under future climate change. The main goal of this study is to identify basic indicators that can be used to understand why so many models produce unrealistic SSW frequencies and how to improve the models. To this end, we compare the intermodel spread of various quantities. We find that the climatological mean strength of the polar vortex is the single most reliable indicator for the SSW frequency, followed by the upward-propagating wave activity flux in the lower stratosphere. In many models, this flux undergoes an unrealistic reduction in the lower stratosphere, which is connected to biases

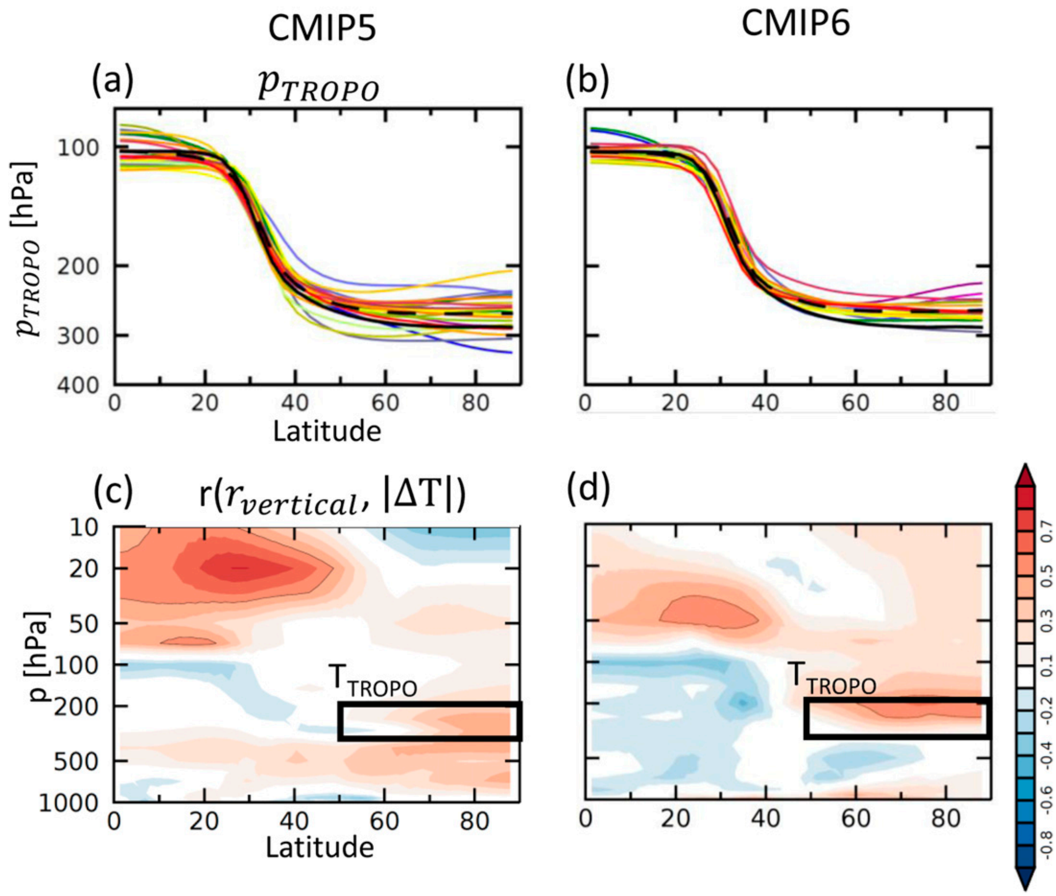

FIG. 9. (a),(b) Tropopause pressure (hPa) by latitude; solid black is for ERA-40 and dashed black is for the MMM. (c),(d) Correlation between intermodel variations in vertical resolution in the stratosphere (r_{vertical}) and magnitude of local temperature differences of each model with respect to the MMM ($|\Delta T|$).

in the simulation of the atmospheric background in this region and also to a too coarse vertical model resolution. In agreement with earlier studies (Scott and Polvani 2004; Son et al. 2007), we identify a critical region in the lower stratosphere just above the subtropical jet, where the index of refraction for planetary waves has a local minimum and acts as a valve controlling the vertical wave propagation into the stratosphere. This is indicated by a robust correlation between the upward wave activity flux and the index of refraction in this region. The index is highly sensitive to subtle inaccuracies in the simulation of the zonal-mean zonal wind, which in turn are connected to temperature biases above the high-latitude tropopause through the thermal wind relationship. Overall, our results are also consistent with findings with idealized models (Jucker 2016; Martineau et al. 2018). However, it is unclear to what extent a model's high-latitude temperatures actively influence the SSWs through the thermal wind relation, or whether it is primarily the zonal-mean zonal wind that controls the SSWs and the temperatures over the high latitudes simply follow along.

Our results for CMIP5 and CMIP6 are often similar, for example in terms of the sign and magnitude of the relationships and the location of key regions. This similarity adds to the credibility of our findings. Since we perform our analysis separately for CMIP5 and CMIP6, we are also able to judge the overall performance of the two ensembles. We find that CMIP6 as a whole is more realistic than CMIP5. This is most obvious from the smaller rms error in the simulated SSW frequency (26% vs 37%), but various correlations in CMIP6 are also more robust and physical than in CMIP5. This indicates that the stratospheric simulation performance in CMIP5 is influenced by a more diverse range of errors than in CMIP6, making it more difficult to linearly relate errors to specific sources. We believe that besides enhanced parameterization, finer resolutions and higher model tops are important reasons for the improvements seen in CMIP6: the mean of all CMIP6 models has a top at 1 hPa and a vertical resolution in the stratosphere (UTLS region) of ~1.8 km (~900 m), whereas in CMIP5 the mean top is located at 2.6 hPa and the vertical resolution is ~2.9 km (~1300 m). The finer vertical resolution in CMIP6 creates a more realistic lower stratospheric background, improves the simulation of the stratospheric wave dynamics, and reduces the negative biases in the vertical wave activity flux, and hence in SSW frequency. In an earlier study, Charlton-Perez et al. (2013) argued that models with a low model lid (below 1 hPa) tend to reproduce too few SSWs, and that a high lid is needed for a model to produce a realistic frequency. Indeed, we find that the SSW frequency and the height of the model lid are correlated at $r \sim 0.6$ in CMIP5. In CMIP6, however, this correlation reduces to $r \sim 0.3$, which is smaller than $r \sim -0.4$ between the vertical resolution and the SSW frequency (Table 2). There is also some linear relationship between the vertical resolution and the height of the model lid ($r \sim -0.6$ in both ensembles), making it difficult to determine whether the improvements in the SSW frequency of CMIP6 are due to the finer resolution, a higher lid, or a combination of both.

There are also some limitations to our study. For example, our results depend on a mix of CMIP models. These models form ensembles of opportunity and do not span the range of uncertainties desirable for this study, creating unavoidable imperfections. Another potential limitation to our study is the use of all available simulation years to calculate the needed climatologies and estimate the frequency of SSWs. We did so because no prior knowledge about a model's SSWs is needed. Some climatologies, like the strength of the stratospheric winds, are influenced by SSWs and are therefore not independent estimators. An alternative approach would have been to only include years without SSWs, but we found that doing so does not fundamentally change our conclusions. Our study is also limited in the sense that the reasons for the intermodel spread in polar vortex strength remain unclear, despite our finding that this strength is crucial for a model's SSW frequency. However, biases in polar vortex strength may be related to a near endless list of reasons, and in most cases it is impossible to diagnose these reasons from the available model output. One such reason may be the treatment of gravity waves. For example, Fig. 1 shows that there are surprisingly large differences in the SSW frequencies produced by the Canadian CanESM2 (110%) and CanESM5 (40%), despite the structural similarities of these two models. These differences are likely related to changes in the orographic gravity wave drag parameterizations (J. Anstey 2019, personal communication). Data for the gravity wave drag tendencies, which could be used to clarify the role of gravity wave drag for SSWs, are available from the CMIP6 archives, but this task must be left to future research. Another shortcoming of our study is that we did not take into account variations in the amount of the tropospheric wave generation, which may also play a role for the number of SSWs. However, these variations appear to be small compared to the variations in the stratosphere (Fig. 2). The calculation of the vertical wave activity fluxes from vertically coarsely resolved model data also creates uncertainties, but these are unavoidable and, based on our tests, quite small. And last, our study relies on linear regression and correlation to examine the dynamical relationships, but these relationships are usually complex and nonlinear. We nevertheless believe that our linear approach is reasonably successful.

An important message of this study is that not only the atmospheric background UTLS region is important for the simulation of SSWs but also a sufficient vertical resolution in the stratosphere. Two aspects are associated with the vertical resolution: first, finer resolution improves the simulation of the atmospheric background, especially in the UTLS region, where the vertical gradients are strong, and second, finer resolution improves the representation of wave dynamics in the entire stratosphere. Most models exhibit in the lower stratosphere a reduction of the wave activity flux entering from below (Fig. 2), and the wave activity flux is significantly negatively correlated with the vertical resolution (Table 2), helping to explain the negative correlation with the SSW frequency. Similar to Charlton-Perez et al. (2013), we also find that models with a higher lid tend to produce more SSWs. This behavior is more pronounced in CMIP5, probably because CMIP6 has a relatively high percentage (65%) of high-top models. However, the upward wave activity flux in the lower stratosphere is only weakly correlated with the height of the model top ($r \sim 0.5$ in CMIP5 and $r \sim 0.2$ in CMIP6), implying that the vertical model resolution is perhaps more important than the height of the model top.

Another lesson learned from this study is that it is difficult to pinpoint errors in a model's SSW frequency to other simulation biases, simply because the frequency depends on the subtle interplay of a variety of complicated factors. Simulating a reasonable SSW frequency can therefore be considered as a sensitive indicator for a model's overall stratospheric simulation performance.

A potential application of the results of this study is that a model's expected SSW frequency can be estimated from only two basic modeling aspects: vortex strength and lower stratospheric wave activity flux. A model developer can derive both quantities at reasonable accuracy from relatively short simulations, avoiding the necessity of much longer simulations for more robust SSW statistics.

Acknowledgments. We thank the National Science Foundation under Grant 1446292 and the Department of Atmospheric Sciences at the University of Utah for their support. The use of the computing infrastructure from the Center for High Performance Computing at the University of Utah is gratefully acknowledged. We thank Isla Simpson for valuable suggestions and comments, and Andrew Charlton-Perez and the other two anonymous reviewers for their comments that helped to improve this paper. We acknowledge ECMWF for providing the ERA-40 data, and the Coupled Model Intercomparison Project, the World Climate Research Programme, and the various modeling groups for producing and making available the model output used in this study.

REFERENCES

Ao, C. O., and Coauthors, 2015: Evaluation of CMIP5 upper troposphere and lower stratosphere geopotential height with GPS radio occultation observations. *J. Geophys. Res. Atmos.*, **120**, 1678–1689, https://doi.org/10.1002/2014JD022239.

Ayarzagüena, B., and Coauthors, 2018: No robust evidence of future changes in major stratospheric sudden warmings: A multi-model assessment from CCMI. *Atmos. Chem. Phys.*, **18**, 11 277–11 287, https://doi.org/10.5194/ACP-18-11277-2018.

——, and Coauthors, 2020: Uncertainty in the response of sudden stratospheric warmings and stratosphere–troposphere coupling to quadrupled CO_2 concentrations in CMIP6 models. *J. Geophys. Res. Atmos.*, **125**, e2019JD032345, https://doi.org/10.1029/2019JD032345.

Birner, T., 2006: Fine-scale structure of the extratropical tropopause region. *J. Geophys. Res.*, **111**, D04104, https://doi.org/10.1029/2005JD006301.

——, 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.

Butchart, N., and Coauthors, 2011: Multimodel climate and variability of the stratosphere. *J. Geophys. Res.*, **116**, D05102, https://doi.org/10.1029/2010JD014995.

Butler, A. H., J. P. Sjoberg, 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*, A. W. Robertson and F. Vitart, Eds., Elsevier, 223–241.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate*, **20**, 449–469, https://doi.org/10.1175/JCLI3996.1.

——, and Coauthors, 2007: A new look at stratospheric sudden warmings. Part II: Evaluation of numerical model simulations. *J. Climate*, **20**, 470–488, https://doi.org/10.1175/JCLI3994.1.

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.

Chen, P., and W. A. Robinson, 1992: Propagation of planetary waves between the troposphere and stratosphere. *J. Atmos. Sci.*, **49**, 2533–2545, https://doi.org/10.1175/1520-0469(1992)049<2533:POPWBT>2.0.CO;2.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

de la Cámara, A., J. R. Albers, and T. Birner, 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.

Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis*. Wiley-Interscience, 706 pp.

Edmon, H. J., B. J. Hoskins, and M. E. McIntyre, 1981: Eliassen–Palm cross sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616, https://doi.org/10.1175/1520-0469(1980)037<2600:epcsft>2.0.co;2.

Eliassen, A., and E. Palm, 1961: On the transfer of energy in stationary mountain waves. *Geofys. Publ.*, **3**, 1–23.

Gerber, E. P., and L. M. Polvani, 2009: Stratosphere–troposphere coupling in a relatively simple AGCM: The importance of stratospheric variability. *J. Climate*, **22**, 1920–1933, https://doi.org/10.1175/2008JCLI2548.1.

——, and E. Manzini, 2016: The Dynamics and Variability Model Intercomparison Project (DynVarMIP) for CMIP6: Assessing the stratosphere–troposphere system. *Geosci. Model Dev.*, **9**, 3413–3425, https://doi.org/10.5194/GMD-9-3413-2016.

Gettelman, A., L. L. Pan, W. J. Randel, P. Hoor, T. Birner, and M. I. Hegglin, 2011: The extratropical upper troposphere and lower stratosphere. *Rev. Geophys.*, **49**, 1–31, https://doi.org/10.1029/2011RG000355.

Hegglin, M. I., and Coauthors, 2010: Multimodel assessment of the upper troposphere and lower stratosphere: Extratropics. *J. Geophys. Res.*, **115**, D00M09, https://doi.org/10.1029/2010JD013884.

Horan, M. F., and T. Reichler, 2017: Modeling seasonal sudden stratospheric warming climatology based on polar vortex statistics. *J. Climate*, **30**, 10 101–10 116, https://doi.org/10.1175/JCLI-D-17-0257.1.

Jucker, M., 2016: Are sudden stratospheric warmings generic? Insights from an idealized GCM. *J. Atmos. Sci.*, **73**, 5061–5080, https://doi.org/10.1175/JAS-D-15-0353.1.

——, S. Fueglistaler, and G. K. Vallis, 2014: Stratospheric sudden warmings in an idealized GCM. *J. Geophys. Res. Atmos.*, **119**, 11 054–11 064, https://doi.org/10.1002/2014JD022170.

Karoly, D. J., and B. J. Hoskins, 1982: Three-dimensional propagation of planetary waves. *J. Meteor. Soc. Japan*, **60**, 109–123, https://doi.org/10.2151/JMSJ1965.60.1_109.

Karpechko, A. Y., P. Hitchcock, 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.

Kim, J., K. M. Grise, and S. W. Son, 2013: Thermal characteristics of the cold-point tropopause region in CMIP5 models. *J. Geophys. Res. Atmos.*, **118**, 8827–8841, https://doi.org/10.1002/JGRD.50649.

Lee, Y., and R. X. Black, 2015: The structure and dynamics of the stratospheric northern annular mode in CMIP5 simulations. *J. Climate*, **28**, 86–107, https://doi.org/10.1175/JCLI-D-13-00570.1.

Martineau, P., G. Chen, S. W. Son, and J. Kim, 2018: Lower-stratospheric control of the frequency of sudden stratospheric warming events. *J. Geophys. Res. Atmos.*, **123**, 3051–3070, https://doi.org/10.1002/2017JD027648.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.*, **27**, 871–883, https://doi.org/10.1175/1520-0469(1970)027<0871:VPOSPW>2.0.CO;2.

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, 38 pp., https://www.ecmwf.int/sites/default/files/elibrary/2019/19084-control-stratospheric-temperature-ifs-resolution-and-vertical-advection.pdf.

Reichler, T., M. Dameris, and R. Sausen, 2003: Determining the tropopause height from gridded data. *Geophys. Res. Lett.*, **30**, 2042, https://doi.org/10.1029/2003GL018240.

Scaife, A. A., T. Woollings, J. Knight, G. Martin, and T. Hinton, 2010: Atmospheric blocking and mean biases in climate models. *J. Climate*, **23**, 6143–6152, https://doi.org/10.1175/2010JCLI3728.1.

Schimanke, S., T. Spangehl, H. Huebener, and U. Cubasch, 2013: Variability and trends of major stratospheric warmings in simulations under constant and increasing GHG concentrations. *Climate Dyn.*, **40**, 1733–1747, https://doi.org/10.1007/S00382-012-1530-X.

Scott, R. K., and L. M. Polvani, 2004: Stratospheric control of upward wave flux near the tropopause. *Geophys. Res. Lett.*, **31**, L02115, https://doi.org/10.1029/2003GL017965.

Shaw, T. A., and J. Perlwitz, 2010: The impact of stratospheric model configuration on planetary-scale waves in Northern Hemisphere winter. *J. Climate*, **23**, 3369–3389, https://doi.org/10.1175/2010JCLI3438.1.

——, ——, 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.

Sigmond, M., and J. F. Scinocca, 2010: The influence of the basic state on the Northern Hemisphere circulation response to climate change. *J. Climate*, **23**, 1434–1446, https://doi.org/10.1175/2009JCLI3167.1.

——, ——, 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.

Simpson, I. R., M. Blackburn, and J. D. Haigh, 2009: The role of eddies in driving the tropospheric response to stratospheric heating perturbations. *J. Atmos. Sci.*, **66**, 1347–1365, https://doi.org/10.1175/2008JAS2758.1.

Sjoberg, J. P., and T. Birner, 2014: Stratospheric wave–mean flow feedbacks and sudden stratospheric warmings in a simple model forced by upward wave activity flux. *J. Atmos. Sci.*, **71**, 4055–4071, https://doi.org/10.1175/JAS-D-14-0113.1.

Son, S. W., S. Lee, and S. B. Feldstein, 2007: Intraseasonal variability of the zonal-mean extratropical tropopause height. *J. Atmos. Sci.*, **64**, 608–620, https://doi.org/10.1175/JAS3855.1.

SPARC, 2010: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. SPARC Rep. 5, WCRP-30/2010, WMO/TD-40, www.sparc-climate.org/publications/sparc-reports/.

Taguchi, M., 2017: A study of different frequencies of major stratospheric sudden warmings in CMIP5 historical simulations. *J. Geophys. Res. Atmos.*, **122**, 5144–5156, https://doi.org/10.1002/2016JD025826.

Tripathi, O. P., and Coauthors, 2015: The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quart. J. Roy. Meteor. Soc.*, **141**, 987–1003, https://doi.org/10.1002/qj.2432.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, https://doi.org/10.1256/qj.04.176.

Wang, W., M. Shangguan, W. Tian, T. Schmidt, and A. Ding, 2019: Large uncertainties in estimation of tropical tropopause temperature variabilities due to model vertical resolution. *Geophys. Res. Lett.*, **46**, 10 043–10 052, https://doi.org/10.1029/2019GL084112.

Wu, Z., and T. Reichler, 2019: Surface control of the frequency of stratospheric sudden warming events. *J. Climate*, **32**, 4753–4766, https://doi.org/10.1175/JCLI-D-18-0801.1.