Research Article

The Influences of the Model Configuration on the Simulation of Stratospheric Northern-Hemisphere Polar Vortex in the CMIP5 Models

Zelin Cai,^{1,2} Ke Wei,¹ Luyang Xu,^{1,2} Xiaoqing Lan,¹ Wen Chen,¹ and Debashis Nath¹

¹Center for Monsoon System Research, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100190, China

²University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence should be addressed to Ke Wei; weike@mail.iap.ac.cn

Received 27 May 2017; Revised 13 September 2017; Accepted 11 October 2017; Published 8 November 2017

Academic Editor: Annalisa Cherchi

Copyright © 2017 Zelin Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As a basic part of the atmosphere, the stratosphere plays an important role in the tropospheric climate and weather systems, especially during the winter, when the stratosphere and troposphere have their strongest interactions. This study assesses the abilities of the Fifth Phase of the Coupled Model Intercomparison Project (CMIP5) and CMIP3 models to simulate the boreal winter stratospheric polar vortex. Analysis indicates that the models with well-resolved stratospheres, that is, with a high model top (HTOP) covering the whole stratosphere, a high vertical resolution (HVer) of the stratosphere, and nonorographic gravity wave drag (NOG), rank higher in both the temporal scoring system and the spatial scoring system. The extreme cold polar vortex bias, which was found in the CMIP3 models, vanishes in the CMIP5 models with HTOP, HVer, and NOG but persists in the other CMIP5 models. A dynamical analysis shows that the heat flux propagating into the stratosphere is stronger in models with HTOP, HVer, and NOG, but these propagations are still weaker than those in the ERA40 reanalysis, indicating the lack of variability in the current CMIP5 models.

1. Introduction

Climate models are very effective tools for improving our mechanistic understanding of climate variability and for performing future climate projections. In recent decades, great progress has been made in the current climate models to have higher resolutions (both horizontal and vertical), more complete physical processes, and more sophisticated parameterization schemes. This has greatly improved model performance and applicability for climate change projections on global to regional scales [1–3].

Most of the published climate change studies are primarily focused on the surface and lower troposphere. Accordingly, most of the model improvements and tunings that have been made are related to the troposphere. However, in recent decades, studies have revealed that the stratosphere can play an important role in near-surface climate variability and change. Stratospheric forcing, resulting from internal

climate variability or from external factors such as massive injections of volcanic aerosols into the stratosphere, can be an important driver of tropospheric climate. Among others, Perlwitz and Graf [4, 5] and Perlwitz and Harnik [6] found that the propagation of planetary waves from the troposphere can be reflected by the stratospheric layers back into the troposphere when the stratospheric background is strong westerly winds [4, 7] or when there is negative wind shear in the westerly wind regime near the stratopause [6]. This can further affect the tropospheric weather systems [8–13] and circulation patterns [14]. Nath et al. [15] showed this link for the 500 hPa height anomalies over the Northern Pacific, which lead to cold weather over North America. Black [16] and Scaife et al. [17] showed that variations of the winter stratospheric polar vortex could have impacts on the surface climate via the impacts of these variations on the Arctic Oscillation/North Atlantic Oscillation. Baldwin and Dunkerton [8, 18] reported that both the strong and

the weak polar vortex anomalies could propagate within the stratosphere down to the tropopause. These anomalies can persist for up to 60 days, favor tropospheric anomalies with the same signs as those in the stratosphere, and influence surface storm tracks, surface pressure [19], and North Atlantic sea surface temperatures [20]. Furthermore, Manzini et al. [21] showed that the sea level pressure, surface temperature, and sea ice coverage anomalies in the Northern Hemispheric mid- and high latitudes can be traced back to the long-lasting stratospheric vortex anomalies on interdecadal time scales. Improved representation of the stratosphere was shown to create a first-order change in climate projections of winter sea level pressure in multiple climate models [22] and single model [23]. For the future climate projection, the spread in climate sensitivity may be reduced by reducing the spread of stratospheric changes [24].

Stratosphere-resolving models are also crucial for the simulation of the Brewer-Dobson circulation (BDC), which is likely to strengthen in the future with a rising greenhouse gas concentration in the atmosphere [25–28]. Since the BDC transports ozone and other trace gases from the tropics to the polar regions by means of the stratospheric overturning circulation, model performance in the stratosphere is important for understanding the formation of the ozone hole and its recovery [29], as well as the variations of the radiation balance [1, 30].

However, previous studies on the stratosphere circulation have shown some common model biases. For example, the “cold pole” problem has prevailed in most stratosphere-resolving coupled general climate models (GCMs) for a long time [31]. Stronger and colder stratospheric polar vortices are closely associated with fewer perturbations of the stratosphere and lower frequencies of sudden stratospheric warming events [32]. Cordero and Forster [33] evaluated the performance of 19 CMIP3 models, which were used for the Intergovernmental Panel on Climate Change Fourth Assessment Report (AR4). Their analysis indicated that the 19 GCMs exhibited a colder stratosphere than the NCEP reanalysis data. Although the models with higher altitudes and more stratospheric levels had better skills when simulating the polar temperature, they still had a cold bias of approximately 4–7 K at 10 hPa. Ren et al. [34] showed that this “cold pole” problem prevailed in their climate model, with an overestimated polar night jet speed, less zonal asymmetry, a much lower SSW frequency (only 3 events in the 60 perpetual-January model months), and a much weaker variability across the polar stratospheric region.

Until recently, this cold pole bias remained in most of the models. An analysis of CMIP5 [35] model simulations showed that the polar stratospheric cold bias in the spring temperatures still persists in most models, which leads to a delay in the final stratospheric warming of both hemispheres [36]. Another bias is the persistent annular mode pattern in the troposphere, which is common and found both in the CMIP3 coupled climate models [37] and in the CCMVal (Chemistry-Climate Model Validation Activity) chemistry-climate models [38]. This bias was considered to be related to the bias of the underestimated intraseasonal variability in the stratosphere [38].

In this study, we utilized a ranking system to study the individual model performances in simulating the stratospheric circulation during the boreal winter, and the question of how to simulate a better stratospheric polar vortex in the northern hemisphere is discussed. We compared several crucial processes, which might contribute to the increase in model performance in the stratosphere, including increases in the model top, the introduction of nonorographic gravity wave drag in the model, increases in the horizontal/vertical resolution of the stratosphere of the model, upgrading the model from climate model to earth system model (ESM), and the inclusion of chemical processes in the stratosphere.

In this study, 42 climate models from CMIP5 that were used in AR5 are used to evaluate their performances in simulating the Northern Hemispheric stratospheric polar vortex (SPV). Meanwhile, the results from 22 CMIP3 models, which were used for the AR4 evaluation, are also utilized for comparison purposes in order to examine the possible improvements in the simulations of SPV climatology. The paper is structured as follows. In Section 2, we will discuss the models, observation-based data, and the methodology used in this study. In Section 3, CMIP5 SPV climatology will be presented. A two-sample Kolmogorov-Smirnov test (K-S test) is performed, and the models will be ranked according to their p values from the K-S test. The dynamic reasons for the differences between the subensembles of the models will be discussed. In addition, a summary and discussion will be presented in Section 4.

2. Models and Data

2.1. Model Output Availability. Model intercomparison projects (MIPs) provide the best opportunities to evaluate how the current state-of-the-art climate models characterize the variability of the climate system. The most prominent and successful MIP is probably the Coupled Model Intercomparison Project (CMIP), which compares simulations from global coupled climate models and provides a standard set of model simulations aiming to evaluate how well the models simulate and project future climate change and to understand some of the factors responsible for the differences in model projections (<http://cmip-pcmdi.llnl.gov/cmip5/>). These simulation results were used in the climate change assessment report of the Intergovernmental Panel on Climate Change (IPCC). Since its inception in 1995, CMIP has evolved to its fifth phase (CMIP5), which was widely used in the fifth assessment report of the IPCC.

According to the IPCC AR5 report, there are a total of 50 CMIP5 models from 25 institutions, which is twice as many as those used in AR4 and CMIP3 (24 models). The CMIP5 models distinguish themselves from the CMIP3 models in many ways. About two-thirds of the CMIP5 models are Earth System Models (ESMs) that include the global carbon cycle [39]. However, all the CMIP3 models use a prescribed globally averaged CO_2 concentration. Hence, these models were called atmosphere-ocean general circulation models (AOGCMs) or climate system models (CSMs). Hereafter,

AOGCMs and CSMs are called GCMs for short. The horizontal resolutions for both the atmospheric and the oceanic components of most of the CMIP5 models are higher than those in the CMIP3 models. Moreover, about half of the CMIP5 models can fully resolve the stratosphere, and the vertical resolutions of some of the models (CESM-WACCM, CMCC-CMS, HadGEM2-CC, MIROC-ESM, MPI-ESM-MR, and MIROC-ESM-CHEM) have the ability to simulate the QBO.

In this study, we adopted 42 couple climate model outputs from the Program for Climate Model Diagnosis and Intercomparison (PCMDI, <http://cmip-pcmdi.llnl.gov/>). The names of the analyzed models are listed in Table 1, along with other model information. Table 1 is compiled using information that the participating modeling centers provided to the PCMDI (see http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php for more information about the models).

To compare with previous versions of CMIP and confirm model progress, the model outputs from 24 CGCMs, released by CMIP3 and used for the AR4 evaluation, are used in this study. The model names of the CMIP3 models are listed in Table 2. These model results were made available by the IPCC Data Distribution Center (<https://esgf.llnl.gov/>) and the Hadley Centre for Climate Prediction and Research.

Since our study is mainly focused on the boreal winter, when the stratosphere and troposphere have their strongest interactions [40], we use the monthly mean model outputs from the midwinter months (December–February). To compare these results with the reanalysis data and CMIP3 results, we have chosen a period of 30 years from 1969/70 to 1998/99. For our analysis, we use the zonal wind, meridional wind, and air temperature data from the r1i1p1 (r for the indicator of control run where historical runs initialized, i for initialization method indicator, and p for perturbed physics number) ensemble [35] for the outputs of the historical runs.

2.2. Observation-Based Data. We utilize the European Centre for Medium-Range Weather Forecasts 40-Year Reanalysis (ERA40, [41]) as our reference data, which covers the period from September 1957 to August 2002. The main reason we have chosen the ERA40 data is that it covers the period of 1969–1999 and reaches the level of 1 hPa. In addition to ERA40, there are other reanalysis datasets; for example, the Japanese 55-year reanalysis (JRA55, [42]) satisfies the necessary criteria. We have validated the JRA55 variables against those from ERA40 to demonstrate reanalysis uncertainties, and the results show that the differences between JRA55 and ERA40 are negligible under the circumstances of the present study. Thus, the results of comparing the model results with the JRA55 data are not shown in this paper.

2.3. Methodology

2.3.1. Basic Treatments. CMIP5 models have a variety of horizontal and vertical resolutions (Table 1), with MIROC4h having a resolution of T213 ($0.5625^\circ \times 0.5625^\circ$) and HadCM3 having a resolution of N48 ($3.75^\circ \times 2.5^\circ$). Hence, we regridded all the model outputs onto a $2.5^\circ \times 2.5^\circ$ grid in the horizontal domain with 23 vertical levels to match the resolution of ERA40. However, the highest level of some models is at

10 hPa, and thus there are only 18 levels for these models after the regridding.

The 42 CMIP5 models and 22 CMIP3 models used in the present study are listed in Tables 1 and 2, respectively. To compare the model performances with or without one specific process, the CMIP5 models can be classified into the following groups according to the schemes of these models.

HTOP: high-top model, that is, with a model top greater than 1 hPa, 20 models

LTOP: low-top model, that is, with a model top lower than 1 hPa, 22 models

NOG: model with nonorographic gravity (NOG) wave parameterization, 20 models

nNOG: model without nonorographic gravity (NOG) wave parameterization, 20 models

CHEM: models with stratospheric chemistry processes, 14 models

nCHEM: models without stratospheric chemistry processes, 28 models

ESM: earth system model that includes the global carbon cycle, 23 models

GCM: model without the global carbon cycle, 19 models

HHor: horizontal high-resolution (HHor) model, with a resolution finer than $1.9^\circ \times 1.9^\circ$ ($\sim T63$), 22 models

LHor: horizontal low-resolution (LHor) model, with a resolution sparser than $1.9^\circ \times 1.9^\circ$ ($\sim T63$), 20 models

HVer: vertical high-resolution (HVer) model of the stratosphere, with more than 19 levels above 200 hPa, 18 models

LVer: vertical low-resolution (LVer) model of the stratosphere, with fewer than 19 levels above 200 hPa, 24 models

The HTOP models are usually associated with an NOG scheme. Of the 20 high top models, 17 models have nonorographic wave parameterization schemes, while only 5/22 of the low top models have this parameterization. Meanwhile, of the 20 HTOP models, 18 models are classified as HVer models, while the LTOP models are all LVer models. Therefore, with the increase of the model top, the vertical resolution of the stratosphere is also increased, and the nonorographic wave parameterization scheme is implemented.

To demonstrate the differences in the vertical resolutions and vertical coverages of the atmospheric heights between the HTOP and LTOP models, the vertical layer profiles of the individual models are shown in Figure 1, which can be viewed as an updated version of Figure 1 of Cordero and Forster [33]. Though the CMIP5 models show an overall improvement over the CMIP3 models, the CMIP5 intermodel differences remain. The lowest model top is 10 hPa (~ 32 km, e.g., INM-CM4, CMCC-CM, and CNRM-CM5), while the highest is 5.1×10^{-4} Pa (~ 140 km, CESM-WACCM) and the rest of the models lie in between. Most of the LTOP models have

TABLE 1: Detailed information about the 42 CMIP5 models used in this study. Atmosphere indicates the atmospheric components of the model. Lid top is the altitude of the model top. >200 hPa indicates how many levels exist above 200 hPa in the model. GCM/ESM (L/H) classifies the models as GCM or ESM (low-top or high-top) models. Y indicates that the model incorporates nonorographic wave parameterization (interactive chemistry) or not (N). Lon $\times$ lat gives the horizontal resolution of the atmosphere model.

Model number	Model	Institute	Lid top	GCM/ESM	Nonorographic wave	Interactive chemistry	L/H top	Lon $\times$ lat
1	CanCM4	CCCMA	0.5 hPa	GCM	N	N	High	$1.875^\circ \times 1.875^\circ$
2	CanESM2	CCCMA	0.5 hPa	ESM	N	N	High	$1.875^\circ \times 1.875^\circ$
3	CESM1-WACCM	NOAA, USA	5.1×10^{-6} hPa	ESM	Y	Y	High	$1.9^\circ \times 2.5^\circ$
4	CMCC-CESM	CMCC	0.01 hPa	ESM	Y	N	High	$3.75^\circ \times 3.75^\circ$
5	CMCC-CMS	CMCC	0.01 hPa	GCM	Y	N	High	$1.875^\circ \times 1.875^\circ$ (T63)
6	GFDL-CM3	GFDL, USA	0.01 hPa	GCM	Y	Y	High	$\sim 1.8^\circ \times 1.8^\circ$
7	GISS-E2-H	NASA GISS, USA	0.01 hPa	GCM	Y	N	High	$2.5^\circ \times 2.0^\circ$
8	GISS-E2-R-CC	NASA GISS, USA	0.01 hPa	GCM	Y	N	High	$2.5^\circ \times 2.0^\circ$
9	GISS-E2-R	NASA GISS, USA	0.01 hPa	GCM	Y	N	High	$2.5^\circ \times 2.0^\circ$
10	HadGEM2-CC	HC, UK	85 km	ESM	Y	N	High	$1.875^\circ \times 1.25^\circ$
11	IPSL-CM5A-LR	IPSL, France	0.04 hPa	ESM	Y	Y	High	$1.9^\circ \times 3.75^\circ$
12	IPSL-CM5A-MR	IPSL, France	0.04 hPa	ESM	Y	Y	High	$1.25^\circ \times 2.5^\circ$
13	IPSL-CM5B-LR	IPSL, France	0.04 hPa	ESM	Y	Y	High	$1.9^\circ \times 3.75^\circ$
14	MIROC4h	MIROC, Japan	0.9 hPa	GCM	N	N	High	$0.5625^\circ \times 0.5625^\circ$ (T213)
15	MIROC-ESM-CHEM	MIROC, Japan	0.0036 hPa	ESM	Y	Y	High	$2.8^\circ \times 2.8^\circ$
16	MIROC-ESM	MIROC, Japan	0.0036 hPa	ESM	Y	Y	High	$2.8125^\circ \times 2.8125^\circ$ (T42)
17	MPI-ESM-LR	MPI-M, German	0.01 hPa	ESM	Y	N	High	$1.9^\circ \times 1.9^\circ$ (T63)
18	MPI-ESM-MR	MPI-M, German	0.01 hPa	ESM	Y	N	High	$1.9^\circ \times 1.9^\circ$ (T63)
19	MPI-ESM-P	MPI-M, German	0.01 hPa	ESM	Y	N	High	$1.9^\circ \times 1.9^\circ$ (T63)
20	MRI-CGCM3	MRI, Japan	0.01 hPa	GCM	Y	N	High	$1.1^\circ \times 1.1^\circ$ (TL159)
21	ACCESS1-0	CAWCR, Australia	39 km	GCM	Y	N	Low	$1.875^\circ \times 1.25^\circ$
22	ACCESS1-3	CAWCR, Australia	39 km	GCM	Y	N	Low	$1.875^\circ \times 1.25^\circ$
23	bcc-csm1	BCC, China	2.97 hPa	ESM	N	N	Low	$2.8^\circ \times 2.8^\circ$
24	BNU-ESM	BNU, China	2.19 hPa	ESM	N	N	Low	$2.8^\circ \times 2.8^\circ$ (T42)
25	CCSM4	NOAA, USA	2.19 hPa	GCM	N	Y	Low	$0.9^\circ \times 1.25^\circ$
26	CESM1-BGC	NOAA, USA	2.19 hPa	ESM	N	Y	Low	$0.9^\circ \times 1.25^\circ$
27	CESM1-CAM5	NOAA, USA	2.19 hPa	ESM	Y	Y	Low	$0.9^\circ \times 1.25^\circ$
28	CESM1-FASTCHEM	NOAA, USA	2.19 hPa	ESM	N	Y	Low	$0.9^\circ \times 1.25^\circ$
29	CMCC-CM	CMCC, Italy	10 hPa	GCM	N	N	Low	$0.75^\circ \times 0.75^\circ$ (T159)
30	CNRM-CM5	CNRMCFRACS	10 hPa	GCM	N	Y	Low	$1.4^\circ \times 1.4^\circ$ (TL127)
31	CSIRO-Mk3-6-0	CSIRO-QCCCE	4.5 hPa	GCM	N	N	Low	$1.8^\circ \times 1.8^\circ$
32	FGOALS-g2	LASG-CES	2.19 hPa	GCM	N	N	Low	$2.8^\circ \times 2.8^\circ$
33	FGOALS-s2	LASG-IAP	2.19 hPa	GCM	N	N	Low	$2.8^\circ \times 1.6^\circ$
34	FIO-ESM	FIO	2.19 hPa	ESM	N	N	Low	$2.8^\circ \times 2.8^\circ$
35	GFDL-ESM2G	GFDL, USA	3 hPa	ESM	N	N	Low	$2.5^\circ \times 2.0^\circ$
36	GFDL-ESM2M	GFDL, USA	3 hPa	ESM	N	N	Low	$2.5^\circ \times 2.0^\circ$
37	HadCM3	HC, UK	10 hPa	GCM	N	Y	Low	$3.75^\circ \times 2.5^\circ$ (N48)
38	HadGEM2-AO	HC, UK	40 km	ESM	N	N	Low	$1.875^\circ \times 1.25^\circ$ (N96)
39	HadGEM2-ES	HC, UK	40 km	ESM	N	N	Low	$1.875^\circ \times 1.25^\circ$
40	INMCM4	INM, Russia	10 hPa	GCM	Y	N	Low	$2.0^\circ \times 1.5^\circ$
41	MIROC5	MIROC, Japan	3 hPa	GCM	N	N	Low	$1.41^\circ \times 1.41^\circ$ (T85)
42	NorESM1-M	NCC, Norway	3.54 hPa	ESM	N	Y	Low	$2.5^\circ \times 1.9^\circ$

TABLE 2: Information about the 22 CMIP3 models used in the present study.

Model number	Model	Institute	Lid top	L/H top	Lon × lat
43	BCCR_BCM2.0	BCCR, Norway	10 hPa	Low	(~1.9° × 1.9°) T63L31
44	CGCM3.1 (T47)	CCCMA, Canada	1 hPa	Low	(~2.8° × 2.8°) T47L31
45	CGCM3.1 (T63)	CCCMA, Canada	1 hPa	Low	(~1.9° × 1.9°) T63
46	CNRM_CM3	CNRM, France	0.05 hPa	High	(~1.9° × 1.9°) T63L45
47	CSIRO_MK3.0	CSIRO, Australia	4.5 hPa	Low	(~1.9° × 1.9°) T63L18
48	CSIRO_MK3.5	CSIRO, Australia	4.5 hPa	Low	(~1.9° × 1.9°) T63L18
49	GFDL_CM2.0	NOAA GFDL, USA	3 hPa	Low	(2.0° × 2.5°) L24
50	GFDL_CM2.1	NOAA GFDL, USA	3 hPa	Low	(2.0° × 2.5°) L24
51	GISS_AOM	NASA GISS, USA	10 hPa	Low	(3.0° × 4.0°) L12
52	GISS_EH	NASA GISS, USA	0.1 hPa	High	(4.0° × 5.0°) L20
53	GISS_ER	NASA GISS, USA	0.1 hPa	High	(4.0° × 5.0°) L20
54	IAP_FGOALS1.0_G	LASG/IAP, China	2.2 hPa	Low	(~2.8° × 2.8°) T42
55	INGV_ECHAM4	INGV, Italy	10 hPa	Low	(~1.1° × 1.1°) T106L19
56	INM_CM3.0	INM, Russia	10 hPa	Low	(4.0° × 5.0°) L21
57	IPSL_CM4	IPSL, France	4 hPa	Low	(2.5° × 3.75°) L19
58	MIROC3.2_HIRES	CCSR/JAMSTEC, Japan	40 km	Low	(~1.1° × 1.1°) T106L56
59	MIROC3.2_MEDRES	CCSR/JAMSTEC, Japan	30 km	Low	(~2.8° × 2.8°) T42L20
60	MPIECHAM5	MPI, Germany	10 hPa	Low	(~1.9° × 1.9°) T63L31
61	MRI_CGCM2.3.2	MRI, Japan	0.4 hPa	High	(~2.8° × 2.8°) T42
62	NCAR_CCSM3.0	NCAR, USA	2.2 hPa	Low	(~1.4° × 1.4°) T85L26
63	NCAR_PCM1	NCAR, USA	2.2 hPa	Low	(~2.8° × 2.8°) T42L26
64	UKMO-HADCM3	MOHC, UK	5 hPa	Low	(2.5° × 3.75°) L19

a top at approximately 3 hPa, and 13 out of the 20 HTOP models have a top at approximately 80 km, which is the altitude of the mesopause. Overall, in both the troposphere and the stratosphere, the number of vertical layers of the CMIP5 models is more than those of the CMIP3 models (Figure 1, [33]). However, it is worth mentioning that these 42 models are not absolutely independent of each other. They primarily originated from CAMs (i.e., BCC-ESM, FGOALS, FIO-ESM, and NorESM1-M), the Hadley Center Models (i.e., the ACCESS models), and the ECHAM models (i.e., MPI-ESMs, CMCC-ESM). More specific information is listed in Table 1.

2.3.2. Ranking System. A temporal ranking system is adopted to study the individual model performances in simulating the stratospheric circulation, based on the work of Brands et al. [43]. This ranking system employs the two-sample Kolmogorov-Smirnov test (K-S test, [44]). The K-S test is a nonparametric test that compares the cumulative PDFs of two samples (e.g., the outputs of a CMIP5 model and ERA40) to determine if two samples have the same distribution. The null hypothesis is that the two samples come from the same theoretical probability distribution [43]. The algorithm is written as follows:

$$p_{ks} = \max_{i=1}^{2n} |E(x_i) - I(x_i)|, \quad (1)$$

where E and I denote the empirical cumulative distribution frequencies of a given model and ERA40, respectively, with x_i being the i th value of the sample and n being the sample size

(here $n = 30$). The value of p is bounded between 0 and 1. As the value approaches 1, the distributions of the two samples are more similar, and vice versa. If p is lower than 0.05, the null hypothesis is denied and the two samples are assumed to be from different distributions at a confidence level of >95%.

For a variable, such as the zonal-mean temperature, the ranking is performed at each grid point of the 42 models according to their respective p values from the K-S test (without the removal of climatology), with higher p values indicating a better rank. If a model is ranked number 1 at some grid point (latitude and height), this means that this model performs best at this grid point. Then, we average the ranks at all the grid points of a model to get the model rank. Finally, the 42 models are ranked according to their model ranks.

We also used the spatial score system, adopted from Douglass et al. [45], Waugh and Eyring [46], and Gettelman et al. [47]. It is based on three statistical metrics:

$$g_m = \max \left(0, 1 - \frac{1}{n} \sum_{i=1}^n \frac{|X_{i\text{mod}} - X_{i\text{ERA40}}|}{n_g \sigma_{i\text{ERA40}}} \right),$$

$$g_c = \frac{(\text{Cor}(X'_{i\text{mod}}, X'_{i\text{ERA40}}) + 1)}{2}, \quad (2)$$

$$g_v = \max \left(0, 1 - \frac{1}{n} \sum_{i=1}^n \frac{|\sigma_{i\text{mod}} - \sigma_{i\text{ERA40}}|}{n_g \sigma_{i\text{ERA40}}} \right),$$

where $X_{i\text{mod}}$ stands for a variable of a model at the i th time step and $X_{i\text{ERA40}}$ for ERA40. σ and Cor stand for the standard deviation and correlation coefficient, respectively. In

FIGURE 1: The vertical coverage and layers of the 42 CMIP5 models used in this study, with the blue lines indicating the tropopause and stratopause. The overall height of a box represents the altitude of the atmospheric model top and the solid lines in the boxes indicate the altitudes of the vertical layers. The 14 leftmost models are lacking specific layer information.

addition, n represents the total time steps and is set to 1 (DJF), and n_g is a scaling factor set to 3, as in Gettelman et al. [47]. Then, g_m , g_c , and g_v are synthesized into a term:

$$G_s = \frac{g_m + g_c + g_v}{3}. \quad (3)$$

The method was proved to be robust and to have some advantages by Grewe and Sausen [48] and Gettelman et al. [47].

3. Results

The 30-year zonal-mean climatology of the temperature and zonal winds of the boreal winter (DJF) are shown in Figure 2. The ERA40 temperature is well characterized, with a cold center located in the polar mid stratosphere called the polar vortex and another centered in the tropical tropopause. In the troposphere, the temperature drops with increasing altitudes and latitudes. The isotherm line of 205 K constrains the cold core of the polar vortex between the altitudes of 70 hPa

and 20 hPa. Given the zonal-mean zonal wind, two jets can be easily distinguished in ERA40 (Figure 2(a)), CMIP3 (Figure 2(b)), and CMIP5 (Figure 2(c)), with one jet being the polar night jet, centered at approximately 65°N, 10 hPa, and the other being the subtropical jet, centered at approximately 30°N, 200 hPa. Easterlies prevail in both the tropical stratosphere and the tropical lower troposphere. The CMIP3 ensemble means show a vortex of similar strength as that seen in ERA40, but the cold core moves upwards, stretching into the upper stratosphere. The polar vortex cold bias between CMIP3 MME and ERA40 is greater than -8 K (Figure 2(c)) in the mid and upper stratosphere. The displaced polar vortex leaves a slight warm bias in the lower stratosphere. Compared to CMIP3 MME, the isotherm line of 205 K of the CMIP5 ensemble is closer to that of the reanalysis, thus constraining the structure of the polar vortex to the mid stratosphere, though it still stretches into the higher levels. The temperature anomalies of CMIP5 at 10 hPa (Figure 2(e)) are around -4 K, much smaller than those of the CMIP3 models (Figure 2(d)). The polar night jet simulated in the CMIP3 models extends

FIGURE 2: The 30-year climatology (left panel) and discrepancies between each ensemble's (right panel) boreal winter (DJF) zonal-mean temperatures (shadings, unit: K) and zonal-mean zonal winds (contours, unit: m s^{-1}). The contour intervals of the left and right panels are 5 m s^{-1} and 1 m s^{-1} , respectively.

equatorward, causing large westerly biases in the midlatitudes and subtropical regions (Figure 2(d)). The extension of this jet in the CMIP5 MME is smaller (Figure 2(e)) because of the more moderate polar vortex in this model. In the tropical stratosphere, both ensembles show easterly biases.

This was probably induced by the lack of the QBO in the model simulations. Only 6 of the 64 employed models (42 CMIP5 and 22 CMIP3) are able to simulate the QBO or QBO-like oscillations. The tropical stratosphere is dominated by easterlies all the year round in the simulations of the other

FIGURE 3: Box plots of the G_s rankings of the zonal-mean zonal winds (left) and zonal-mean temperatures (right) for the polar cap region (60° – 90° N, 50–10 hPa) from the CMIP5 (red) and CMIP3 (blue) models. The middle values are the multimodel means, and the dots are the median values of the models; the upper and lower values are shown for the uncertainty range of the mean, and the uppermost and lowermost bars are the maximum and minimum values.

FIGURE 4: Similar to Figure 3 but for the temporal rankings. The primes represent rankings with the 30-year means removed from each grid.

59 models. Without interannual variability, the easterlies in the models are stronger than those observed.

To derive quantitative information about the performances of the models, the CMIP3 and CMIP5 ensembles are ranked according to spatial and temporal ranking systems, respectively. The results are shown in Figures 3 and 4. The upper and lower values of the box plots indicate the uncertainty range of the mean value, represented by $\sigma/\sqrt{n}$. For the spatial ranking system (Figure 3), both ensembles exhibit similar distributions, but the medians and averages of the CMIP5 models are slightly better than those of the CMIP3 models for both the zonal-mean zonal wind and the zonal-mean temperature. However, the differences in the temporal variabilities of the two ensembles (Figure 4) are significant.

FIGURE 5: Box plot of the rankings of the models classified into 6 pairs of groups: HT/LTOP models, with/without nonorographic wave drag, high/low vertical resolution in the stratosphere, ESM/GCM, interactive/noninteractive chemistry, and high/low horizontal resolution. In each group, the middle value is the multimodel mean, and the dot is the median value; the upper and lower values are shown for the uncertainty range of the mean, and the uppermost and lowermost bars show the maximum and minimum values.

TABLE 3: The median value of the model rankings of the mean temperatures in the stratospheric polar vortex region, classified into 6 pairs of groups: HT/LTOP models, with/without nonorographic wave drag, high/low vertical resolution in the stratosphere, ESM/GCM, interactive/noninteractive chemistry, and high/low horizontal resolution.

		Ranking	Number of models
Lid top	HT	15.1	20
	LT	22.55	22
Nonorographic wave	Y	14.40	21
	N	23.90	21
Interactive chemistry	Y	18.3	14
	N	20.1	28
GCM/ESM	ESM	17.3	23
	GCM	23.9	19
Horizontal resolution	H	18.7	22
	L	19.4	20
Vertical resolution	H	14.65	18
	L	24.2	24

For both variables, the rankings of CMIP5 ensemble tend to skew towards small values, yet they leave a long tail towards large values. This phenomenon indicates that some of the CMIP5 models have improved simulations and some did not, as compared to the results of the CMIP3 ensemble.

To further determine the processes that improve model simulations of the stratosphere, the CMIP5 models are sorted into groups, as discussed in Section 2.3.1. The temporal rankings of the zonal-mean temperatures for each of the subensembles are shown in Figure 5 and Table 3. Using t -test, the ranking differences between HTOP and LTOP, NOG and nNOG, HVer and LVer, and ESM and GCM are significant

TABLE 4: The median value of the model rankings of the mean temperatures in the stratospheric polar vortex region, classified into 4 pairs of groups for the high-top models and low-top models. The numbers in brackets are the number of models in each model group.

	Nonorographic wave		Interactive chemistry		GCM/ESM		Horizontal resolution	
	Y	N	Y	N	ESM	GCM	High	Low
HT	14.4 (17)	29.8 (3)	13.35 (6)	17.05 (14)	14.25 (12)	19.55 (8)	13.1 (10)	16.55 (10)
LT	17.35 (4)	22.55 (18)	23.65 (8)	22 (14)	20.1 (11)	24.5 (11)	26.9 (12)	20.65 (10)

at 97.68%, 99.95%, 99.90%, and 86.03%, respectively, while the ranking differences between CHEM and nCHEM and HHor and LHor are only significant at 15.11% and 13.88%, respectively. The median ranking of the HTOP models is 15.1, which is smaller than that of the LTOP models (22.6). In the CMIP3 (AR4) archive, only 5 of the 23 models have tops above 1 hPa [33], but that ratio increases to 22 of the 50 models in the CMIP5 suite [36]. In addition to the stratospheric polar vortex climatology, raising the model top can also increase the stratospheric variability on daily and interannual time scales and increase the frequency of major sudden stratospheric warming events [36].

Gravity waves (GWs) are generated by convection, topography, instability, tropical cyclones, and various adjustment processes in the atmosphere [49, 50]. It is now believed that vertically propagating gravity waves can provide the strong mechanical forcing that keeps the middle atmosphere away from radiative equilibrium and helps maintain the residual circulation between the equator and poles [51, 52]. The incorporation of nonorographic gravity wave drag could alleviate the cold pole problem in the stratosphere [53]. However, even at the CMIP5 stage, almost half of the models do not include nonorographic gravity wave parameterization. The NOG models rank higher than the nNOG models, with a median ranking of 14.4 for the 21 NOG models and 23.9 for the 21 nNOG models. The introduction of nonorographic gravity wave schemes usually requires a high model top [54]. Among the 21 NOG models, 17 models are HTOP models, which should be one critical reason why the HTOP models have more realistic polar vortices.

Moreover, the vertical resolutions of the stratosphere and the incorporation of carbon cycles can help to improve simulations of the polar vortex. The median ranking of the 18 HVer models is 14.7, which is much better than the median value of 24.2 of the 22 LVer models. Higher vertical resolutions allow subgrid waves to propagate vertically instead of being damped. For example, the successful simulation of the QBO requires an 800 m vertical resolution [55, 56]. For the ESM and GCM groups, the median ranking of the 23 ESMs is 17.3, which is much better than the 23.9 median ranking of the 19 GCMs. The ESMs are reported to have better simulations of the polar vortex than those of GCMs not only in their climatology but also for the intraseasonal variations of the vortex [57]. The inclusion of carbon cycle, especially the ocean biogeochemistry, has an influence on the global oceanic radiation flux, oceanic chlorophyll, and energy balance, which will help to simulate a better ocean mean state and better SST. This may influence the stratospheric circulation via large-scale planetary waves. However, this should be confirmed by further study.

Meanwhile, the treatments of stratospheric chemistry processes are not vital to the rankings of the models, indicating that, in current models, the temperature of the polar vortex does not strongly rely on interactive chemistry processes. The ranking differences between CHEM and nCHEM models are minor (18.3 for the 14 CHEMs versus 20.1 for the 28 nCHEMs). Therefore, using interactive chemistry or semioffline calculated chemistry in CHEM models can only lead to a very small change in the model ranking compared to using the prescribed ozone in the nCHEM models. This is coincident with the results of Eyring et al. [58], who noted that both the CHEM and the nCHEM multimodel means agree well with observations in terms of their absolute values and the trends of both temperature and zonal wind.

It is worth noting that improvements of the horizontal resolutions barely show a better polar vortex simulation. The median ranking for the 22 high-resolution models (those with a model resolution finer than $1.9^\circ \times 1.9^\circ$) is 18.7, while the median ranking for the other models is 19.4, suggesting that improving the model resolution results in a very limited improvement of the polar vortex temperature simulation.

Raising the model top can be the first step in improving the simulations of stratospheric circulation. For the HTOP models, the inclusions of the nonorographic gravity waves (NOGs), interactive chemistry (CHEM), and carbon cycle (ESM), as well as an increased horizontal resolution, can all lead to model ranking improvements (Table 4). The median ranking value rises from 29.8 (nNOG) to 14.4 (NOG), 17.05 (nCHEM) to 13.35 (CHEM), 19.55 (GCM) to 14.25 (ESM), and 16.55 (LHor) to 13.1 (HHor) in the HTOP models. For the LTOP models, however, the introduction of CHEM and finer resolutions leads to setbacks in the model rankings. The median ranking value drops from 22 (nCHEM) to 23.65 (CHEM) and from 20.65 (LHor) to 26.9 (HHor) in the LTOP models. Therefore, the development of models to include CHEM and finer horizontal resolution must also be accompanied by an increase of the model top to achieve a better simulation of the NH stratospheric polar vortex. The inclusion of NOG in LT and HT modes remains robust, with both subensembles showing a model ranking increase compared with the nNOG models.

The combination of the high model tops and high horizontal resolutions leads to better performances, with a median model ranking of 13.1 for the 10 models. We noted that 8 of the 10 models are models with high vertical resolutions in the stratosphere. The median rank of the 8 models with high model tops, high horizontal resolutions, and high vertical resolutions is 11.5, indicating that improvements of the dynamical model configuration can lead to considerable improvements in the simulation of the stratospheric polar

vortex. Furthermore, 7 of the above 8 models are models with nonorographic wave parameterization. The median value of the 7 models is 11.2, which is the best overall performance of all the combinations.

We compare these 7 best models (GFDL-CM3, CMCC-CMS, HadGEM2-CC, MPI-ESM-LR, MPI-ESM-MR, MPI-ESM-P, and MRI-CGCM3), which all have high model tops, high horizontal resolutions, high vertical resolutions, and nonorographic wave parameterization, to the other models and ERA40, as shown in Figures 6 and 7. The polar vortex of the 7-model ensemble (7ME) is well structured and confined to the proper altitudes. The cold pole problem was almost absent. The maximum temperature anomalies between the 7ME and the ERA40 are approximately 2 K at 10 hPa in the polar region and are only -0.5 K at approximately 20 hPa, while the temperature differences between the other models' ensemble (OME) and ERA40 are above -5 K at approximately 10 hPa over the polar cap, indicating a severe cold bias in these models. Accordingly, the zonal-mean zonal wind of the 7ME is very close to that of the reanalysis (Figure 6(d)), with very small departures; meanwhile, OME has stronger meridional temperature gradients around the polar vortex, leading to stronger zonal winds from the polar stratosphere to the subtropical tropopause in the models (Figure 6(e)). Positive temperature anomalies between 7ME and OME dominate the stratospheric polar region and are centered at 10 hPa (Figure 6(f)); the zonal wind difference has most of its largest value within the stratosphere, indicating that improvements of the dynamical model configuration and the inclusion of nonorographic waves can influence the stratospheric circulation simulation.

Another zonal wind anomaly center in both 7ME and OME is located in the tropical stratosphere. The maximum easterly in this area of 7ME is ~ 4 m/s and is ~ 16 m/s in the other model ensemble. The decrease of the easterly in the tropical stratosphere is caused by the successful simulation of the QBO in 3 of the 7 models (CMCC-CMS, HadGEM2-CC, and MPI-ESM-MR).

The EP-flux diagnosis was used to explore the dynamic differences between the 7ME and the OME. As shown in Figures 7(d)-7(e), both the 7ME and the OME exhibit suppressed heat fluxes when compared with that of ERA40. This indicates that the planetary wave activities in both ensembles are relatively weak. The 7ME has a weaker tropospheric wave guide (Figure 7(f)) than that of the OME, yet its stratospheric wave guide is slightly stronger than that of the OME. According to Chen et al. [59], this kind of dipole structure in EP-flux anomalies could also lead to the strengthening of the subtropical jet and the weakening of the polar night jet. Compared to that of ERA40, the convergence area of the 7ME models in the mid stratosphere has expanded into higher latitude regions, and thus the divergence center in the polar stratosphere found in ERA40 has almost vanished in the 7ME. In Figure 7(d), an anomalous convergence center is shown to exist in the polar vortex area, adjacent to the warm anomaly centers. Though the simulated heat fluxes (F_z) in the 7ME are weaker than those in ERA40, they are stronger than those in the OME (Figure 7(f)). Moreover, there are poleward propagating momentum flux anomalies in both

Figures 7(d) and 7(f). As a result, the EP-flux anomalies and the convergences in the polar vortex regions lead to relatively warmer temperatures in the 7ME models.

4. Summary and Discussion

In this study, the abilities of the CMIP3 and SMIP5 models to simulate the stratospheric polar vortex of the boreal winter are assessed. Considerable improvements were shown in the CMIP5 models, with decreased cold biases in the polar region and more realistic wind configurations. The grouped ranking analysis shows that at least 3 aspects contribute to these improvements: higher model tops, the inclusion of nonorographic gravity wave drag, and higher vertical resolutions in the stratosphere. The nonorographic gravity wave drags are vital for simulating a more realistic polar vortex. The forcing of the nonorographic wave drag together with heat flux (F_z) will warm the polar stratosphere and lead to a moderate polar vortex in the models. In fact, the introduction of nonorographic wave drag requires a high model top and a high vertical resolution in the stratosphere. A high vertical resolution of the stratosphere helps to portray the dynamics of this region. In addition, the incorporation of global carbon cycle can also help to improve the simulations of the polar vortex.

Raising the model top can be the first step in improving the simulations of stratospheric circulation. Analysis shows that including the nonorographic waves (NOGs), interactive chemistry (CHEM), carbon cycle (ESM), and improved horizontal resolution can lead to better simulations of the polar vortex, under the condition that the model top is increased to include the whole stratosphere. However, for low-top models, the modifications of the above processes lead to very limited improvements and even setbacks of the model rankings.

The ensemble of the 7 models with high model tops, nonorographic waves, and finer horizontal and vertical resolutions almost overcomes the cold-pole problem. Therefore, the temperature gradient between the high and midlatitudes is weakened, leading to a more realistic polar night jet. Although the EP flux is still weaker than that observed, it is stronger than that modeled by the CMIP3 models and the other CMIP5 models. The inclusion of finer vertical and horizontal resolutions resolves more small-scale waves in the model. Meanwhile, the NOG parameterization provides extra wave forcing in the stratosphere. All these help to damp the polar night jet and produce a better simulation. Compared to the HTOP models, the stratosphere in LTOP models is more exposed to outer space and the outgoing long wave radiation may be higher than that of the HTOP models. On the other hand, GHGs and ozone in the upper stratosphere of the HTOP models can absorb the long wave radiation from the lower atmosphere and thus keep the stratosphere warmer.

It is worth noting that the complexity of the current coupled models makes it very difficult to determine which process is responsible for the improvement of a specific system. The comparisons of each model group with others are "ensembles of opportunity" rather than "clean" comparisons, because any of the comparisons made will certainly contain

FIGURE 6: The 30-year climatology (left panel) and discrepancies between each ensemble (right panel) of the boreal winter (DJF) zonal-mean temperatures (shadings, unit: K) and zonal-mean zonal winds (contours, unit: m s^{-1}). (a) ERA40; (b) ensemble of the 7 CMIP5 models (with HT, NOG, HVer, and HHor); (c) ensemble of the other CMIP5 models; (d) the difference between (b) and (a); (e) the difference between (c) and (a); (f) the difference between (c) and (b). The contour intervals of the left and right panels are 5 m s^{-1} and 1 m s^{-1} , respectively.

FIGURE 7: The EP flux cross section of planetary waves (vectors) and its divergence (contours) for (a) ERA40, (b) CMIP5 7ME (ensemble of the 7 models with HTOP, NOG, HVer, and HHor), (c) CMIP5 OME (ensemble of the other models except the above models), (d) the difference between CMIP5 7ME and ERA40, (e) the difference between CMIP5 OME and ERA40, and (f) the difference between CMIP5 7ME and CMIP5 OME. The p fluxes are scaled by the inverse of the air density and the units are kg s^{-2} . The EP flux divergence contour interval is $0.6 \text{ m s}^{-1} \text{ day}^{-1}$.

other important differences. However, the comparison of the stratospheric circulation between CIMP5 models can help model users to understand the model uncertainties in stratospheric processes, and also provide useful information for future model improvements on model ability in describing the stratospheric circulation.

The improvements made to the simulation of the polar vortex will further influence the simulation of the troposphere climate since the stratosphere has significant influences on the tropospheric climate, as introduced in Section 1. The climate change projections based on models with cold biases in the boreal polar vortex need to be reconsidered. For example, Wei et al. [60] have shown that the discrepancies in the winter surface air temperature in East Asia between the HTOP and LTOP models can be up to 1.3 K by the end of the 21st century under different RCP scenarios. Hence, global and regional climate projections need to be reevaluated based on models that are better able to characterize the stratospheric circulation; otherwise, adaptation and mitigation policies might produce insufficient responses to the effects of anthropogenic global warming over the coming decades.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant no. 41375046), the National Key Research and Development Project of China (Grant no. 2016YFA0600600), and the Youth Innovation Promotion Association of Chinese Academy of Sciences (no. 2014064). In this study, the ERA40 data were provided by the ECMWF and are available online at <http://apps.ecmwf.int/datasets/data/>.

References

- [1] S. Solomon, K. H. Rosenlof, R. W. Portmann et al., "Contributions of stratospheric water vapor to decadal changes in the rate of global warming," *Science*, vol. 327, no. 5970, pp. 1219–1223, 2010.
- [2] K. R. Sperber, H. Annamalai, I.-S. Kang et al., "The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century," *Climate Dynamics*, vol. 41, no. 9–10, pp. 2711–2744, 2013.
- [3] K. Wei, T. Xu, Z. Du, H. Gong, and B. Xie, "How well do the current state-of-the-art CMIP5 models characterise the climatology of the East Asian winter monsoon?" *Climate Dynamics*, vol. 43, no. 5–6, pp. 1241–1255, 2014.
- [4] J. Perlwitz and H.-F. Graf, "The statistical connection between tropospheric and stratospheric circulation of the Northern Hemisphere in winter," *Journal of Climate*, vol. 8, no. 10, pp. 2281–2295, 1995.
- [5] J. Perlwitz and H.-F. Graf, "Troposphere-stratosphere dynamic coupling under strong and weak polar vortex conditions," *Geophysical Research Letters*, vol. 28, no. 2, pp. 271–274, 2001.
- [6] J. Perlwitz and N. Harnik, "Observational evidence of a stratospheric influence on the troposphere by planetary wave reflection," *Journal of Climate*, vol. 16, no. 18, pp. 3011–3026, 2003.
- [7] J. G. Charney and P. G. Drazin, "Propagation of planetary-scale disturbances from the lower into the upper atmosphere," *Journal of Geophysical Research: Atmospheres*, vol. 66, no. 1, pp. 83–109, 1961.
- [8] M. P. Baldwin and T. J. Dunkerton, "Propagation of the Arctic Oscillation from the stratosphere to the troposphere," *Journal of Geophysical Research: Atmospheres*, vol. 104, no. 24, pp. 30937–30946, 1999.
- [9] K. Kodera, H. Mukougawa, and S. Itoh, "Tropospheric impact of reflected planetary waves from the stratosphere," *Geophysical Research Letters*, vol. 35, no. 16, Article ID L16806, 2008.
- [10] S. C. Hardiman, N. Butchart, T. J. Hinton, S. M. Osprey, and L. J. Gray, "The effect of a well-resolved stratosphere on surface climate: Differences between CMIP5 simulations with high and low top versions of the met office climate model," *Journal of Climate*, vol. 25, no. 20, pp. 7083–7099, 2012.
- [11] K. Kodera, H. Mukougawa, and A. Fujii, "Influence of the vertical and zonal propagation of stratospheric planetary waves on tropospheric blockings," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 15, pp. 8333–8345, 2013.
- [12] D. Nath, W. Chen, L. Wang, and Y. Ma, "Planetary wave reflection and its impact on tropospheric cold weather over Asia during January 2008," *Advances in Atmospheric Sciences*, vol. 31, no. 4, pp. 851–862, 2014.
- [13] J. Kidston, A. A. Scaife, S. C. Hardiman et al., "Stratospheric influence on tropospheric jet streams, storm tracks and surface weather," *Nature Geoscience*, vol. 8, no. 6, pp. 433–440, 2015.
- [14] K. Coughlin and K. K. Tung, "Tropospheric wave response to decelerated stratosphere seen as downward propagation in northern annular mode," *Journal of Geophysical Research: Atmospheres*, vol. 110, no. 1, pp. 1–9, 2005.
- [15] D. Nath, W. Chen, C. Zelin, A. I. Pogoreltsev, and K. Wei, "Dynamics of 2013 Sudden Stratospheric Warming event and its impact on cold weather over Eurasia: Role of planetary wave reflection," *Scientific Reports*, vol. 6, Article ID 24174, 2016.
- [16] R. X. Black, "Stratospheric forcing of surface climate in the Arctic oscillation," *Journal of Climate*, vol. 15, no. 3, pp. 268–277, 2002.
- [17] A. A. Scaife, J. R. Knight, G. K. Vallis, and C. K. Folland, "A stratospheric influence on the winter NAO and North Atlantic surface climate," *Geophysical Research Letters*, vol. 32, no. 18, Article ID L18715, pp. 1–5, 2005.
- [18] M. P. Baldwin and T. J. Dunkerton, "Stratospheric harbingers of anomalous weather regimes," *Science*, vol. 294, no. 5542, pp. 581–584, 2001.
- [19] K. Walter and H. F. Graf, "The North Atlantic variability structure, storm tracks, and precipitation depending on the polar vortex strength," *Atmospheric Chemistry and Physics*, vol. 5, no. 1, pp. 239–248, 2005.
- [20] H.-F. Graf and K. Walter, "Polar vortex controls coupling of North Atlantic Ocean and atmosphere," *Geophysical Research Letters*, vol. 32, no. 1, pp. 1–4, 2005.
- [21] E. Manzini, C. Cagnazzo, P. G. Fogli, A. Bellucci, and W. A. Mueller, "Stratosphere-troposphere coupling at inter-decadal time scales: Implications for the North Atlantic Ocean," *Geophysical Research Letters*, vol. 39, no. 5, Article ID L05801, 2012.
- [22] A. A. Scaife, T. Spanghel, D. R. Fereday et al., "Climate change projections and stratosphere-troposphere interaction," *Climate Dynamics*, vol. 38, no. 9–10, pp. 2089–2097, 2012.

- [23] A. Y. Karpechko and E. Manzini, "Stratospheric influence on tropospheric climate change in the Northern Hemisphere," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. 5, Article ID D05133, 2012.
- [24] E. Manzini, A. Y. Karpechko, J. Anstey et al., "Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 13, pp. 7979–7998, 2014.
- [25] N. Butchart and A. A. Scaife, "Removal of chlorofluorocarbons by increased mass exchange between the stratosphere and troposphere in a changing climate," *Nature*, vol. 410, no. 6830, pp. 799–802, 2001.
- [26] N. Butchart, A. A. Scaife, and M. Bourqui, "Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation," *Climate Dynamics*, vol. 27, no. 7-8, pp. 727–741, 2006.
- [27] R. R. Garcia and W. J. Randel, "Acceleration of the brewer-dobson circulation due to increases in greenhouse gases," *Journal of the Atmospheric Sciences*, vol. 65, no. 8, pp. 2731–2739, 2008.
- [28] T. Iwasaki, H. Hamada, and K. Miyazaki, "Comparisons of Brewer-Dobson circulations diagnosed from reanalyses," *Journal of the Meteorological Society of Japan*, vol. 87, no. 6, pp. 997–1006, 2009.
- [29] S. Solomon, "Stratospheric ozone depletion: A review of concepts and history," *Reviews of Geophysics*, vol. 37, no. 3, pp. 275–316, 1999.
- [30] D. W. J. Thompson and S. Solomon, "Understanding recent stratospheric climate change," *Journal of Climate*, vol. 22, no. 8, pp. 1934–1943, 2009.
- [31] S. Pawson, K. Kodera, K. Hamilton et al., "The GCM-reality intercomparison project for SPARC (GRIPS): Scientific issues and initial results," *Bulletin of the American Meteorological Society*, vol. 81, no. 4, pp. 781–796, 2000.
- [32] A. J. Charlton and L. M. Polvani, "A new look at stratospheric sudden warmings. Part I. Climatology and modeling benchmarks," *Journal of Climate*, vol. 20, no. 3, pp. 449–469, 2007.
- [33] E. C. Cordero and P. M. D. F. Forster, "Stratospheric variability and trends in models used for the IPCC AR4," *Atmospheric Chemistry and Physics*, vol. 6, no. 12, pp. 5369–5380, 2006.
- [34] R. Ren, G. Wu, M. Cai, and J. Yu, "Winter season stratospheric circulation in the SAMIL/LASG general circulation model," *Advances in Atmospheric Sciences*, vol. 26, no. 3, pp. 451–464, 2009.
- [35] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An overview of CMIP5 and the experiment design," *Bulletin of the American Meteorological Society*, vol. 93, no. 4, pp. 485–498, 2012.
- [36] A. J. Charlton-Perez, M. P. Baldwin, T. Birner et al., "On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 6, pp. 2494–2505, 2013.
- [37] E. P. Gerber, L. M. Polvani, and D. Ancukiewicz, "Annular mode time scales in the intergovernmental panel on climate change fourth assessment report models," *Geophysical Research Letters*, vol. 35, no. 22, Article ID L22707, 2008.
- [38] E. P. Gerber, M. P. Baldwin, H. Akiyoshi et al., "Stratosphere-troposphere coupling and annular mode variability in chemistry-climate models," *Journal of Geophysical Research: Atmospheres*, vol. 115, no. 18, Article ID D00M06, 2010.
- [39] P. Friedlingstein, M. Meinshausen, V. K. Arora et al., "Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks," *Journal of Climate*, vol. 27, no. 2, pp. 511–526, 2014.
- [40] M. P. Baldwin, D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, "Stratospheric memory and skill of extended-range weather forecasts," *Science*, vol. 301, no. 5633, pp. 636–640, 2003.
- [41] S. M. Uppala, P. W. allberg, A. J. Simmons, U. Andrae, and V. da Costa Bechtold, "The ERA-40 re-analysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, pp. 2961–3012, 2005.
- [42] S. Kobayashi, Y. Ota, Y. Harada et al., "The JRA-55 reanalysis: general specifications and basic characteristics," *Journal of the Meteorological Society of Japan*, vol. 93, no. 1, pp. 5–48, 2015.
- [43] S. Brands, S. Herrera, J. Fernández, and J. M. Gutiérrez, "How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa?: A performance comparison for the downscaling community," *Climate Dynamics*, vol. 41, no. 3-4, pp. 803–817, 2013.
- [44] D. Wilks, *Statistical Methods in the Atmospheric Sciences*, vol. 100, Academic press, San Diego, Calif, USA, 2011.
- [45] A. R. Douglass, M. J. Prather, T. M. Hall et al., "Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft," *Journal of Geophysical Research: Atmospheres*, vol. 104, no. D22, pp. 27545–27564, 1999.
- [46] D. W. Waugh and V. Eyring, "Quantitative performance metrics for stratospheric-resolving chemistry-climate models," *Atmospheric Chemistry and Physics*, vol. 8, no. 18, pp. 5699–5713, 2008.
- [47] A. Gettelman, M. I. Hegglin, S.-W. Son et al., "Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends," *Journal of Geophysical Research: Atmospheres*, vol. 115, no. 20, Article ID D00M08, 2010.
- [48] V. Grewe and R. Sausen, "Comment on "quantitative performance metrics for stratospheric- resolving chemistry-climate models" By Waugh and Eyring (2008)," *Atmospheric Chemistry and Physics*, vol. 9, no. 23, pp. 9101–9110, 2009.
- [49] S. Watanabe, Y. Kawatani, Y. Tomikawa, K. Miyazaki, M. Takahashi, and K. Sato, "General aspects of a T213L256 middle atmosphere general circulation model," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. 12, Article ID D12110, 2008.
- [50] D. Chen, Z. Y. Chen, and D. R. Lü, "Simulation of the stratospheric gravity waves generated by the Typhoon Matsa in 2005," *Science China Earth Sciences*, vol. 55, no. 4, pp. 602–610, 2012.
- [51] R. J. Murgatroyd and F. Singleton, "Possible meridional circulations in the stratosphere and mesosphere," *Quarterly Journal of the Royal Meteorological Society*, vol. 87, no. 372, pp. 125–135, 1961.
- [52] D. G. Andrews and M. E. McIntyre, "Planetary waves in horizontal and vertical shear. I. The generalized Eliassen-Palm relation and the mean zonal acceleration," *Journal of the Atmospheric Sciences*, vol. 33, no. 11, pp. 2031–2048, 1976.
- [53] A. A. Scaife, N. Butchart, C. D. Warner, and R. Swinbank, "Impact of a spectral gravity wave parameterization on the stratosphere in the met office unified model," *Journal of the Atmospheric Sciences*, vol. 59, pp. 1473–1489, 2002.
- [54] C. McLandress and J. F. Scinocca, "The GCM response to current parameterizations of nonorographic gravity wave drag," *Journal of the Atmospheric Sciences*, vol. 62, no. 7, pp. 2394–2413, 2005.

- [55] K. Hamilton, R. J. Wilson, and R. S. Hemler, "Middle atmosphere simulated with high vertical and horizontal resolution versions of a GCM: Improvements in the cold pole bias and generation of a QBO-like oscillation in the tropics," *Journal of the Atmospheric Sciences*, vol. 56, no. 22, pp. 3829–3846, 1999.
- [56] K. Hamilton, R. J. Wilson, and R. S. Hemler, "Spontaneous stratospheric QBO-like oscillations simulated by the GFDL SKYHI general circulation model," *Journal of the Atmospheric Sciences*, vol. 58, no. 21, pp. 3271–3292, 2001.
- [57] J. Rao, R. Ren, and Y. Yang, "Parallel comparison of the northern winter stratospheric circulation in reanalysis and in CMIP5 models," *Advances in Atmospheric Sciences*, vol. 32, no. 7, pp. 952–966, 2015.
- [58] V. Eyring, J. M. Arblaster, I. Cionni et al., "Long-term ozone changes and associated climate impacts in CMIP5 simulations," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 10, pp. 5029–5060, 2013.
- [59] W. Chen, M. Takahashi, and H. F. Graf, "Interannual variations of stationary planetary wave activity in the northern winter troposphere and stratosphere and their relations to NAM and SST," *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D24, 2003.
- [60] K. Wei, Z. Cai, W. Chen, and L. Xu, "The effect of a well-resolved stratosphere on East Asian winter climate," *Climate Dynamics*, pp. 1–14, 2016.