The Climatology and the Midwinter Suppression of the Cold-Season North Pacific Storm Track in CMIP6 Models

MINGHAO YANG, a CHONGYIN LI, a,b XIONG CHEN, a YANKE TAN, c XIN LI, a CHAO ZHANG, a and GUIWAN CHEN b

a College of Meteorology and Oceanography, National University of Defense Technology, Nanjing, China
b State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
c Department of Atmospheric and Oceanic Sciences, Institute of Atmospheric Sciences, Fudan University, Shanghai, China

(Manuscript received 11 May 2020, in final form 5 May 2021)

ABSTRACT: The reproducibility of climatology and the midwinter suppression of the cold-season North Pacific storm track (NPST) in historical runs of 18 CMIP6 models is evaluated against the NCEP reanalysis data. The results show that the position of the climatological peak area of 850-hPa meridional eddy heat flux ($v' T_{850}$) is well captured by these models. The spatial patterns of climatological $v' T_{850}$ are basically consistent with the NCEP reanalysis. Generally, NorESM2-LM and CESM2-WACCM present a relatively strong capability to reproduce the climatological amplitude of $v' T_{850}$ with lower RMSE than the other models. Compared with CMIP5 models, the intermodel spread of $v' T_{850}$ climatology among the CMIP6 models is smaller, and their multimodel ensemble is closer to the NCEP reanalysis. The geographical distribution in more than half of the selected models is farther south and east. For the seasonal variability of $v' T_{850}$ nearly half of the models exhibit a double-peak structure. In contrast, the apparent midwinter suppression in the NPST represented by the 250-hPa filtered meridional wind variance ($v' v'_{250}$) is reproduced by all the selected models. In addition, the present study investigates the possible reasons for simulation biases regarding climatological NPST amplitude. It is found that a higher model horizontal resolution significantly intensifies the climatological $v' v'_{250}$ and the intensity of the East Asian winter monsoon (EAWM). However, the climatological $v' T_{850}$ is not sensitive to the model grid spacing. Additionally, the climatological low-tropospheric atmospheric baroclinicity is uncorrelated with climatological $v' v'_{250}$. The stronger climatological baroclinic energy conversion is associated with the stronger climatological $v' T_{850}$.

KEYWORDS: Storm tracks; Climate models; Model errors; Model evaluation/performance

1. Introduction

The storm track refers to the region with the strongest prevalent synoptic-scale perturbations in the midlatitudes (Chang and Fu 2002) and represents baroclinic wave activity (Wallace et al. 1988; Booth et al. 2017), which plays a crucial role in the weather and climate of the midlatitudes. The transient eddy activity in the storm track has an important contribution to preserving the global climatological mean temperature field and relaxing the equator-to-pole temperature gradient (Trenberth and Stepaniak 2003). There is a symbiotic relationship between the transient eddy and planetary-scale basic flow anomaly (Cai and Mak 1990; Branstator 1995; Song et al. 2016), and the interaction between them can cause low-frequency variations in midlatitude atmospheric circulation (Lau 1988; Luo et al. 2011). Therefore, variations in the intensity and migration of storm tracks are fundamental in determining the equilibrium state of the climate system (Zappa et al. 2013). Observational studies show that both the North Pacific storm track (NPST) and the North Atlantic storm track (NAST) have significant subseasonal, seasonal, interannual, and interdecadal variations (Nakamura 1992; Chang and Fu 2002; Nakamura et al. 2002; Lee et al. 2012a; Nakamura et al. 2012a; Wang et al. 2017). In general, the intensity of the storm track in the Northern Hemisphere attains its minimum in summer and the seasonal variations are broadly in alignment with the variations in the atmospheric baroclinicity. Based on observational datasets from 1955 to 1984, Nakamura (1992) documented that although the background baroclinicity marked by the vertical shear of zonal wind peaks during midwinter, the intensity of the NPST in midwinter is weaker than that in late autumn and early spring. This counterintuitive phenomenon is known as the “midwinter suppression” phenomenon of the NPST (Nakamura 1992; Nakamura et al. 2002). The existence of this phenomenon was confirmed by Christoph et al. (1997) with longer time-span observations and a general circulation model. In contrast, the NAST generally reaches its maximum in midwinter, which corresponds to the atmospheric baroclinicity (Nakamura 1992; Lee et al. 2012b). It should be noted that the midwinter minimum can also occur for the NAST, especially in winters of an anomalously strong North Atlantic jet (Deng and Mak 2006; Penny et al. 2013; Afargan and Kaspi 2017; Park and Lee 2020).

The mechanisms that lead to midwinter suppression have been extensively studied from the perspective of upper-tropospheric jets (Nakamura 1992; Chang 2001; Nakamura and Sampe 2002; Harnik and Chang 2004; Afargan and Kaspi 2017; Yuval et al. 2018;...
Novak et al. (2020), the topography of central Asian mountains (Park et al. 2010; Lee et al. 2013), the eddy energy budget (Deng and Mak 2006; Schemm and Schneider 2018; Zhao and Liang 2019), diabatic processes (Chang 2001; Chang and Zurita-Gotor 2007), the upstream seeding effect (Penny et al. 2010; Chang and Guo 2012), and time-mean flow (Nakamura et al. 2002; Park and Lee 2020). For example, a stronger upper-tropospheric jet is accompanied by more obvious midwinter suppression in the NPST due to the advection and trapping of eddies by the excessively increased group velocity (Nakamura 1992; Nakamura et al. 2002). Recently, based on the theories of canonical transfer (Liang 2016) and localized multiscale energy and vorticity analysis (Liang and Robinson 2005) by using multiscale window transform (Liang and Anderson 2007), Zhao and Liang (2019) investigated the causes and underlying dynamic processes of the midwinter suppression in the NPST. They found that the weakening of baroclinic canonical transfer in midwinter is due to the large distance between the NPST and the upper-tropospheric jet, which attenuates their interaction. Novak et al. (2020) used an idealized aquaplanet general circulation model to demonstrate that a midwinter suppression can also occur in the activity of a statistically zonally symmetric storm track. In addition, with a global perspective, Park and Lee (2020) proposed a new mechanism that partly attributes the midwinter suppression to global planetary-scale waves consuming the zonal available potential energy. A more detailed overview on the mechanisms that lead to midwinter suppression can be found in Schemm and Schneider (2018), Zhao and Liang (2019), and Novak et al. (2020).

Using sophisticated climate models has been an effective way to investigate the dynamics of storm tracks (Chang and Guo 2007), the impact of oceanographic processes on storm tracks (Small et al. 2014; Wang et al. 2016; Ma et al. 2017), and the response of end-of-century storm changes to global warming (Yin 2005; Ulbrich et al. 2008; Nishii et al. 2009; Chang et al. 2012; Lehmann et al. 2014). Before conducting these scientific studies, it is essential to evaluate how well these climate models perform in reproducing storm tracks under current climate conditions (Chang et al. 2013). Ulbrich et al. (2008) analyzed the spatial distribution of the Northern Hemisphere storm track under present-day forcing conditions in 16 climate models from the Coupled Model Intercomparison Project (CMIP) phase 3 (CMIP3). Chang et al. (2013) evaluated the reproducibility of 17 CMIP3 models in terms of the climatological global storm-track activity and seasonal cycle and found that nearly half of the selected models present prominent biases in the intensity of the storm track. Based on the objective feature-tracking algorithm, Zappa et al. (2013) investigated the intensity and spatial patterns in the summer and winter NASTs in detail in CMIP5 phase 5 (CMIP5) models. Yang et al. (2018) documented the climatology and interannual variability in the winter NPST in CMIP5 models. Booth et al. (2017) studied the wintertime surface storm tracks represented by 24-h differences in the daily mean 10-m meridional wind data over western boundary currents and their extensions in CMIP5 models. However, little attention has been paid to the evaluation of midwinter suppression in the NPST. By defining the metric as the ratio of the standard deviation to the root-mean-square error, Nishii et al. (2009) assessed the reproducibility of CMIP3 models on the seasonal evolution of the NPST and analyzed the future projection of the NPST in midwinter.

The CMIP phase 6 (CMIP6; Eyring et al. 2016) models give us the opportunity to study global and regional climate change with newly released data. Evaluating the performance of climate models to characterize the storm track is crucial for increasing the confidence in future projections (Zappa et al. 2013). Most recently, Priestley et al. (2020) evaluated the globally extratropical storm tracks in CMIP6 historical simulations in detail by applying the Lagrangian method. They found that the primary biases of the storm-track state in CMIP5 still exist in CMIP6, although to a lesser extent. Based on the Eulerian perspective, the specific aims of the present study are the following:

1) to document the climatological characteristics of the lower-tropospheric NPST in CMIP6 models and conduct a comparison with the CMIP5 ensemble;
2) to evaluate the reproducibility of the midwinter suppression of NPST in CMIP6 models; and
3) to discuss the possible factors influencing the simulations.

The structure of the present study is as follows: a brief description of the data and methods is presented in section 2. The climatological amplitude, geographical distribution, and midwinter suppression of the NPST are presented in section 3. Finally, summaries and discussions are given in section 4.

2. Data and methods

a. Data

According to Chang et al. (2013), the storm-track amplitude in the Northern Hemisphere is consistent among the different reanalysis data. The daily atmospheric data used as a reference in the present study, including meridional wind and air temperature, are provided by the National Centers for Environmental Prediction (NCEP) global atmospheric reanalysis datasets with a horizontal resolution of 2.5° × 2.5° (Kalnay et al. 1996). The results of historical runs in 18 CMIP6 models are selected, and brief information about them is presented in Table 1. For models with multiple runs, only a single run, r1i1p1f1 (in which the model integration starts from 1 January 1861 with the external forcings, and the first nine integration years are recognized as the spinup time; He et al. 2020), is used in our research.

In addition, the historical runs in 13 CMIP5 models (Yang et al. 2018) are compared with the results of the CMIP6 models when the comparison with the multimodel ensemble (MME) is conducted. The time span of the data used in the present study is 1951–2005. The cold season refers to the period from November to March containing two peaks of the NPST. Furthermore, two models (IPSL-CM6A-ATM-HR and IPSL-CM6A-LR with horizontal resolutions of 50 and 250 km, respectively) that participate in the High Resolution
Table 1. List of 18 selected CMIP6 climate models. The horizontal resolution (lon × lat) refers to the dimensions of the model output of historical runs on a Gaussian grid. More details are available online at https://esgf-node.llnl.gov/search/cmip6/.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Institution</th>
<th>Horizontal resolution</th>
<th>Vertical levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC-CSM2-MR</td>
<td>Beijing Climate Centre (BCC), China Meteorological Administration (CMA), China</td>
<td>320 × 160</td>
<td>46</td>
</tr>
<tr>
<td>BCC-ESM1</td>
<td>BCC, CMA, China</td>
<td>128 × 64</td>
<td>26</td>
</tr>
<tr>
<td>CanESM5</td>
<td>Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Canada</td>
<td>128 × 64</td>
<td>49</td>
</tr>
<tr>
<td>CESM2</td>
<td>National Center for Atmospheric Research (NCAR), Climate and Global Dynamics Laboratory (CGDL), United States</td>
<td>288 × 192</td>
<td>32</td>
</tr>
<tr>
<td>CESM2-FV2</td>
<td>NCAR, CGDL, United States</td>
<td>144 × 96</td>
<td>32</td>
</tr>
<tr>
<td>CESM2-WACCM</td>
<td>NCAR, CGDL, United States</td>
<td>288 × 192</td>
<td>70</td>
</tr>
<tr>
<td>EC-Earth3</td>
<td>European Earth System Model (EESM), European Community</td>
<td>512 × 256</td>
<td>91</td>
</tr>
<tr>
<td>EC-Earth3-Veg</td>
<td>EESM, European Community</td>
<td>512 × 256</td>
<td>91</td>
</tr>
<tr>
<td>FGOALS-f3-L</td>
<td>Chinese Academy of Sciences (CAS), Institute of Atmospheric Physics (IAP), China</td>
<td>360 × 180</td>
<td>32</td>
</tr>
<tr>
<td>FGOALS-g3</td>
<td>CAS, IAP, China</td>
<td>180 × 80</td>
<td>26</td>
</tr>
<tr>
<td>GFDL-CM4</td>
<td>National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, United States</td>
<td>360 × 180</td>
<td>33</td>
</tr>
<tr>
<td>INM-CM4-8</td>
<td>Institute for Numerical Mathematics (INM), Russian Academy of Science (RAS), Russia</td>
<td>180 × 120</td>
<td>21</td>
</tr>
<tr>
<td>INM-CM5-0</td>
<td>INM, RAS, Russia</td>
<td>180 × 120</td>
<td>73</td>
</tr>
<tr>
<td>IPSL-CM6A-LR</td>
<td>Institut Pierre Simon Laplace, France</td>
<td>144 × 143</td>
<td>79</td>
</tr>
<tr>
<td>MPI-ESM1-2-HR</td>
<td>Max Planck Institute for Meteorology, Germany</td>
<td>384 × 192</td>
<td>95</td>
</tr>
<tr>
<td>NESM3</td>
<td>Nanjing University of Information Science and Technology, China</td>
<td>192 × 96</td>
<td>47</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>Norwegian Climate Centre, Norway</td>
<td>144 × 96</td>
<td>32</td>
</tr>
<tr>
<td>SAM0-UNICON</td>
<td>Seoul National University, South Korea</td>
<td>288 × 192</td>
<td>30</td>
</tr>
</tbody>
</table>

Model Intercomparison Project (HighResMIP) (Haarsma et al. 2016) are used in section 3d.

b. Methods

The synoptic-scale filtered eddy variance is easy to calculate and can well reflect the interaction between the storm track and the time-mean flow (Chang 2009). The storm track can be represented by eddy quantities such as geopotential height variances, the eddy kinetic energy, heat fluxes, and momentum fluxes (Chang and Fu 2002). Given the baroclinic nature of synoptic-scale disturbances particularly in the lower troposphere (Nakamura et al. 2002), the NPST is characterized by the meridional eddy heat flux at 850 hPa ($\vec{y} T_{850}$) in the present study, and the following sections on the climatological amplitude and geographical distribution focus on this eddy indicator. Considering that the NPST midwinter suppression is more obvious in the upper troposphere (Nakamura 1992; Lee et al. 2011), another commonly used storm-track metric, the filtered meridional wind velocity variance at 250 hPa ($\vec{y} u_{250}$), is also included in sections 3c and 3d. The Lanczos bandpass filter (Duchon 1979) with 13 daily weights is used to isolate synoptic-scale (2.5–6-day) disturbances from the daily data. The storm-track strength index (STSI) used in our research is the same as that used by Yang et al. (2020a), which sets a threshold defined as the median of the NPST amplitudes of all of the grids within a domain (25°–65°N, 130°E–120°W), and $\text{NPST}_{\text{amp}}$ denotes the amplitude at the grid point.

According to Deng and Mak (2006), the subseasonal variability index (SVI), which quantifies the strength of the NPST subseasonal variability, can be written as

$$\text{SVI} = \frac{\text{ST}_1 - (\text{ST}_1 + \text{ST}_3)/2}{(\text{ST}_1 + (\text{ST}_1 + \text{ST}_3)/2)^{1/2}}$$

where $\text{ST}_1$, $\text{ST}_2$, and $\text{ST}_3$ denote the intensity of the NPST in November, January, and March, respectively. A positive (negative) SVI indicates the maximum (minimum) intensity of NPST in midwinter. Some scholars find it convenient to use the SVI to measure the midwinter suppression of storm tracks (e.g., Mak and Deng 2007). When the SVI is negative, the midwinter suppression phenomenon is thought to occur. The larger the absolute value of negative SVI, the stronger the midwinter suppression. In addition, it should be noted that even when the SVI is zero or slightly positive, the NPST may be partly suppressed compared to the seasonality with a distinct midwinter peak.

The East Asian winter monsoon (EAWM) index (EAWMI) defined by Li and Yang (2010) using the zonal wind at 200 hPa can be represented as follows:

$$\text{EAWMI} = \frac{1}{2} \left( 2 \times U_{200}[30°–35°N, 90°–160°E] - U_{200}[5°–60°N, 70°–170°E] - U_{200}[5°S–10°N, 90°–160°E] \right)$$

X. Jiang et al. (2013) and Gong et al. (2014) argued that this EAWMI can describe most characteristics and physical processes related to the EAWM. The EAWMI is used to investigate the relationship between the NPST and EAWM.
3. Results

a. Climatological amplitude

Figure 1 shows the climatological amplitude of the NPST marked by $v'$ at $850$ in the CMIP6 models. All these models can basically reproduce the climatological distribution of $v'$, which consists of a zonally elongated pattern from East Asia to the west coast of North America over the North Pacific basin. The peak intensity of $v'$ is found eastward of Japan in the NCEP reanalysis (Fig. 2a). The locations of peak intensity of $v'$ in all models are basically consistent with the NCEP reanalysis, but there are significant differences in intensity among these models. Specifically, INM-CM4-8 overestimates the intensity of almost whole $v'$ region, with the deviation center exceeding 4 K m s$^{-1}$, while the simulation results in some models (EC-Earth3-Veg, FGOALS-f3-L, IPSL-CM6A-LR, and SAM0-UNICON) are monolithically weaker than those in the NCEP reanalysis. The simulated amplitude in INM-CM5-0 is overestimated in the northern flank of the climatological mean and underestimated in the southern flank, while the simulation result in GFDL-CM4 is basically the opposite of the overestimated area in the southern flank and the underestimated area in the northern flank. In addition, it is noteworthy that from Fig. 1, the amplitudes simulated in almost all these CMIP6 models are weaker in the southwestern flank of the climatological mean but stronger on the southeastern flank compared with those in the NCEP reanalysis, which can also be found in the pattern in their MME (Fig. 2b) and may indicate systematic deviations in the CMIP6 models.

Figure 3 shows the Taylor diagram of the climatological North Pacific $v'$ of the CMIP6 models and 13 selected CMIP5 models, which can objectively evaluate the capability
of each model in simulating the $v' T_{850}$. We see that except for NESM3, the spatial correlation coefficients between the pattern in the climatological amplitude simulated by the remaining CMIP6 models and that derived from the NCEP reanalysis are all over 0.9, especially those of CESM2, CESM2-WACCM, and NorESM2-LM, which are as high as 0.98, showing a better capability to simulate the spatial pattern of climatological $v' T_{850}$. However, in the CMIP5 models, FGOALS-g2 can reproduce the spatial pattern in the climatological $v' T_{850}$ as well as these CMIP6 models, but the spatial correlation coefficient between nearly half of the CMIP5 models (INM-CM4, IPSL-CM5B-LR, CanESM2, GFDL-CM3, and MPI-ESM-LR) and the NCEP reanalysis is less than 0.9, especially INM-CM4, which is only 0.65, indicating a relatively large deviation in the spatial pattern.

For the standard deviation of the climatological $v' T_{850}$ in these CMIP6 models, the only simulation result from INM-CM4-8 is greatly overestimated, and the result from NESM3 is almost consistent with the NCEP reanalysis. However, the $v' T_{850}$ amplitudes in the remaining models are smaller than those in the NCEP reanalysis. The standard deviations of the climatological $v' T_{850}$ in most CMIP5 models are similar to those in the CMIP6 models with negative deviation, and there are three models (FGOALS-g2, ACCESS1.3, and IPSL-CM5B-LR) whose results are larger than those in the NCEP reanalysis. The root-mean-square error (RMSE) measured by the distance from the reference point (REF) to each scatter in Fig. 3 is used to describe the simulation deviation quantitatively. The CMIP6 models whose scatter fall in the first circle with REF as the center are BCC-CSM2-MR, CanESM5, CESM2, CESM2-WACCM, and NorESM2-M; there is also one CMIP5 model in this category (FGOALS-g2), suggesting that the simulated climatological $v' T_{850}$ in these models is closer to that of the NCEP reanalysis. However, a CMIP6 model, INM-CM4-8, and two CMIP5 models, INM-CM4 and IPSL-CM5B-LR, have relatively large RMSEs compared with the other models. In addition, from Fig. 3, we can also see some improvement in the CMIP6 models relative to the CMIP5 models. For example, the performance of CanESM5 is obviously better in simulating the observed climatological pattern of $v' T_{850}$ than that of CanESM2. The results of INM-CM4-8 are more consistent with the NCEP reanalysis than INM-CM4, especially in terms of the spatial pattern. The reproducibility of NorESM2-LM and GFDL-CM4 is also better than that of NorESM1-M and GFDL CM3, respectively. However, the simulation results of FGOALS-g2 are more successful at spatial pattern and have smaller RMSE relative to those of FGOALS-g3. It is worth noting that the CMIP6 models have more concentrated scatters in the Taylor diagram than the CMIP5 models, demonstrating that the differences among the CMIP6 models are smaller than those among the CMIP5 models.
To describe the regional differences between models more clearly, Fig. 4 shows the intermodel spreads estimated with the standard deviations of these models. The standard deviations among the CMIP6 models (Fig. 4a) are primarily concentrated in the northeastern flank of the MME of the climatological $v'T_{850}$ with the standard deviation center located at 40°–50°N, 160°E–170°W. The standard deviations among the CMIP5 models (Fig. 4b) are similar to those among the CMIP6 models, but the magnitude of the standard deviation is larger than that in the CMIP6 models.

The MME peak area magnitude of the climatological $v'T_{850}$ in the CMIP5 models is underestimated and smaller than that in the CMIP6 models (Fig. 2c). The simulation biases in CMIP6 ensemble are similar to those in CMIP5 ensemble but to a lesser extent, which is consistent with Priestley et al. (2020). It can also be found from Fig. 3 that the scatter for CMIP6 MME is closer to REF compared with the CMIP5 MME and any single models. In addition, the spatial correlation coefficients of the MME and NCEP reanalysis of CMIP6 and CMIP5 are 0.99 and 0.97, respectively, suggesting that the MME of CMIP6 is slightly better in simulating the observed climatological pattern of $v'T_{850}$ than the MME of CMIP5. In general, the capability of the CMIP6 model to reproduce the climatology of $v'T_{850}$ is obviously improved compared to the CMIP5 models, which is in agreement with the result of cyclone intensity in the Northern Hemisphere (Priestley et al. 2020).

### Geographical distribution

It can be seen from the above analysis that most CMIP6 and CMIP5 models achieve reasonable reproducibility of the spatial pattern in $v'T_{850}$ climatology, especially CESM2, CESM2-WACC, and NorESM2-LM from CMIP6 and FGOALS-g2 from CMIP5. To more carefully investigate the simulation results of $v'T_{850}$ geographical distribution, following the same method as in the definition of the STSI, the latitude index (LAI) and the longitude index (LOI) of the NPST can be expressed as

$$\text{LAI} = \frac{\sum_{1}^{N} \text{NPST}_{\text{lat}}}{N}, \quad (4)$$

$$\text{LOI} = \frac{\sum_{1}^{N} \text{NPST}_{\text{lon}}}{N}, \quad (5)$$

where NPST$_{\text{lat}}$ and NPST$_{\text{lon}}$ represent the latitude and longitude on the grid point, respectively. The LAI and LOI help us quantitatively describe the geographical distribution of the NPST. A positive difference from the NCEP reanalysis in the LAI (LOI) marks a northward (eastward) shift of the storm track. The LAIs from the MME of the CMIP6 models and BCC-ESM1 are basically consistent with the NCEP reanalysis with little bias (Fig. 5a). The LAIs from the MME of the CMIP5 models and more than half of the selected CMIP6 models are underestimated, revealing that the positions of $v'T_{850}$ in them are farther south. The meridional positions in five CMIP6 models (CanESM5, CESM2-FV2, INM-CM4-8, INM-CM5-0, and NESM3) are farther north compared with that in the NCEP reanalysis.

For the zonal structure, the NPST not only has a polycentric distribution but also the strongest centers are often found east of 160°W (Li et al. 2010). In the region east of 160°W over the North Pacific, the reason for the occurrence of the NPST peak center is relatively complex and is difficult to explain by the existing dynamics mechanism of the storm track alone, and the influence of external forcing cannot be ignored and may even be the most important factor (Zhu et al. 2013). The anomalies

![Fig. 4. Maps showing the standard deviations of $v'T_{850}$ among (a) the CMIP6 models and (b) the CMIP5 models (shading; unit: K m s$^{-1}$). The contours are the MME of climatological amplitude with intervals of 1 K m s$^{-1}$.](image)

![Fig. 5. The differences of the long-term mean (a) latitude index and (b) longitude index of the NPST represented by $v'T_{850}$ in NDJFM between the individual CMIP6 models and the NCEP reanalysis. Positive differences are in red and negative in blue.](image)
of the NPST in its eastern flank can even cause anomalies in the nearby Aleutian low and blocking (Nakamura and Wallace 1990; Huang et al. 2002). For the LOI of $v'^T_{850}$ (Fig. 5b), except for CESM2-WACCM, INM-CM5-0, and NorESM2-LM, the remaining individual CMIP6 models and both the MMEs of the CMIP6 and CMIP5 models overestimate the LOI, indicating that the $v'^T_{850}$ from them is farther east, especially IPSL-CM6A-LR and SAM0-UNICON. The results of CESM2-FV, EC-Earth3, FGOALS-f3-L, and INM-CM4-8 are more consistent with the NCEP reanalysis. In general, the MME of the CMIP6 models not only simulates the LAI better than the MME of the CMIP5 models but also reproduces the LOI with less bias.

c. Midwinter suppression

Figure 6 shows the latitude–time map of the monthly mean march of the climatological-mean NPST amplitude characterized by the lower-tropospheric $v'^T$, which contributes to investigating the capability of CMIP6 models to reproduce the midwinter suppression. Nearly half of the CMIP6 models (BCC-ESM1, CESM2-FV2, FGOALS-f3-L, GFDL-CM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, NESM3 and SAM0-UNICON) exhibit obvious double-peak structures, which appear in late autumn (November) or early winter (December) and early spring (March) or midspring (April), respectively. Among those more than half of the CMIP6 models with a single maximum (excluding INM-CM4-8, whose peak occurs in midwinter), the peaks occur in late autumn or early winter. In addition, the strength of the early spring $v'^T_{850}$ in most models is weaker than that of the NCEP reanalysis.

Figure 7a shows that there are two significant peaks near 40°N in the NCEP reanalysis in late autumn and early spring, respectively, and the intensity in early spring is greater than...
FIG. 7. The latitude–time maps of the monthly mean march of climatological-mean NPST amplitude (shading) represented by the (a),(b),(e),(f) $\psi' T_{850}$ (unit: K m s$^{-1}$) averaged over 140°E–160°W and (c),(d),(g),(h) $\psi' v_{250}$ averaged over 130°E–160°W, and the averaged upper-tropospheric westerly wind at 250 hPa (deep pink contours) over 130°–160°E with contours at 20, 30, 40, 50, 60, and 70 m s$^{-1}$ in (a),(c) the NCEP reanalysis, (b),(d) the MME of the CMIP6 models, (e),(g) the better-simulated group, and (f),(h) the general-simulated group. The black contours denote the differences between the (b)(d) MME and (e)–(h) two groups of the CMIP6 models and the NCEP reanalysis, with intervals of 0.5 K m s$^{-1}$ in (b), (e), and (f) and 5 m$^2$ s$^{-2}$ in (d), (g), and (h); the black bold line represents the contour with the zero value.
that in late autumn. Only NESM3 successfully reproduces this phenomenon, revealing a relatively strong capability to simulate the midwinter suppression of $\nu_T$ at 850. The MME of the CMIP6 models has barely double peaks (Fig. 7b), and the strength in late autumn is obviously larger than that in early spring. The minimum and maximum cold-season $\nu_T$ intensities in the NCEP reanalysis occur in February and March, respectively. The minimum intensity in half of the CMIP6 models also appears in February, but the maximum intensity in more than half of the model occurs in December, which means that the capability of the CMIP6 models to simulate the subseasonal variability of the lower-tropospheric NPST remains to be improved.

The SVIs from the CMIP6 models are presented in the vertical coordinates of Fig. 8. Except for the positive SVI from INM-CM4-8, the SVIs of $\nu_T$ from the other models are negative (Fig. 8a), showing that the minimum intensity occurs in midwinter or there is a midwinter suppression phenomenon in these models. Although the criterion proposed by Deng and Mak (2006) to judge midwinter suppression can successfully identify it in most CMIP6 models, there may be some potential problems with such a criterion. Particularly, in the selection of months, more than half of the models, including the MME and the NCEP reanalysis, show that the minimum intensity of $\nu_T$ occurs in February, and four models (BCC-CSM2-MR, CESM2, EC-Earth3, INM-CM5-0) show the minimum intensity occurring in March instead of January. Apart from INM-CM4-8, in which the second peak appears in early spring, the second peak in the remaining models appears in midspring. The intensities of the first peak in most models are underestimated. The MME underestimates the intensity of the two peaks, especially the midspring peak. If we want to use the SVI to measure the midwinter suppression in the upper-level troposphere, we need to redefine it, changing the months from November and March to October and April. It can be found from Fig. 8b that the SVIs of $\nu_T$ from all these CMIP6 models are negative, revealing the successful reproduction of midwinter suppression in the upper-level troposphere.

Studies have shown that midwinter suppression and upper-tropospheric jets have certain relationships (Nakamura 1992; Chang et al. 2002; Nakamura et al. 2002; Harnik and Chang 2004; Nishii et al. 2009; Afargan and Kaspi 2017); more specifically, a stronger upper-tropospheric jet is accompanied by more obvious midwinter suppression in the NPST. The scatter diagram of the SVI derived from the $\nu_T$ and the upper-tropospheric westerly wind at 250 hPa over 25°–35°N, 130°–160°E in DJF (abscissa). The asterisk (pound sign) indicates statistical significance at the 5% (10%) level.

Compared with the midwinter suppression of NPST represented by the $\nu_T$ (Fig. 6), that in the NPST represented by the $\nu_T$ is more significant (Fig. 9), which is consistent with the observations of Nakamura (1992) and Lee et al. (2010). All these models, including the MME (Fig. 7d), successfully reproduce the midwinter suppression with the double-peak pattern and weakening in midwinter. In addition, it is worth noting that the midwinter suppression in the upper-level troposphere is somewhat different from that in the lower-level troposphere; not only it is more obvious, but the double peaks also occur in midautumn (October) and midwinters (April) (Fig. 7c), respectively. The first peaks in four models (FGOALS-g3, IPSL-CM6A-LR, INM-CM4-8, and NESM3) occur in late autumn. Apart from INM-CM4-8, in which the second peak appears in early spring, the second peak in the remaining models appears in midwinter. The intensities of the first peak in most models are underestimated. The MME underestimates the intensity of the two peaks, especially the midspring peak. If we want to use the SVI to measure the midwinter suppression in the upper-level troposphere, we need to redefine it, changing the months from November and March to October and April. It can be found from Fig. 8b that the first peaks of $\nu_T$ from all these CMIP6 models are negative, revealing the successful reproduction of midwinter suppression in the upper-level troposphere.
with statistical significance at the 10% level (Fig. 8b). However, it should be noted that the relationship between the midwinter suppression and the upper-tropospheric jet would be prominently impacted by the outliers (INM-CM4-8 and NESM3). A sensitivity test by removing the outliers (INM-CM4-8 and NESM3) shows that there are no significant relationships between the climatological intensity of the upper-tropospheric jet and the midwinter suppression of NPST represented by both $y_T^{850}$ and $y_T^{250}$. To better examine the association of the upper-tropospheric jet with the midwinter suppression, Fig. 10 shows the latitude–time maps of the westerly jet speed for the individual models. Although some models have relatively large simulation biases, like INM-CM4-8, INM-CM5-0, and EC-Earth3-Veg, all the CMIP6 models, including the MME (Figs. 7b,d), can well reproduce the subseasonal variability of the upper-tropospheric jet with the maximum amplitude in January (Fig. 7a), which provides a favorable condition for the midwinter suppression of NPST. However, the intensity of upper-tropospheric jet is remarkably underestimated by INM-CM4-8 (Fig. 10i), which may be responsible for its failure in reproducing midwinter suppression represented by $y_T^{850}$ (Fig. 6l).

From the bias distributions in Figs. 6 and 9, we can find that almost all the CMIP6 models, including its MME (Figs. 7b,d), underestimate the amplitudes of spring NPST in both the lower and upper troposphere and cannot reproduce the relative amplitude between the two peaks. The resulting obviously weakened spring peak may result from the systematic biases in these CMIP6 models. To further investigate the reproducibility of midwinter suppression in CMIP6 models, according to the performance of individual model in Taylor diagram (Fig. 3), these CMIP6 models are divided into the better-simulated group and the general-simulated group with eight members each. The better-simulated group includes BCC-CSM2-MR,
BCC-ESM1, CanESM5, CESM2, CESM2-WACCM, EC-Earth3, EC-Earth3-Veg, and NorESM2-LM, and the general-simulated group includes CESM2-FV2, FGOALS-g3, FGOALS-f3-L, GFDL-CM4, INM-CM5-0, IPSL-CM6A-LR, MPI-ESM1-2-HR, and SAM0-UNICON. The zonally averaged amplitudes of $v' T'_{850}$ (Fig. 7e) and $v' v'_{250}$ (Fig. 7g) in the better-simulated group are obviously more consistent with the NCEP reanalysis than those in the general-simulated group (Figs. 7f,h), especially with less simulation biases in two peaks. In addition, although both groups can successfully reproduce the midwinter suppression of $v' v'_{250}$, the result in the better-simulated group is more remarkable. The simulated upper-tropospheric jet in the better-simulated group is stronger than that in the general-simulated group, which may be responsible for the more remarkable midwinter suppression in the better-simulated group (Nakamura 1992).

d. Possible factors influencing the simulations

1) HORIZONTAL RESOLUTION

Previous studies revealed that the horizontal resolution has a significant influence on the simulation results (e.g., Senior 1995; Willison et al. 2013; Small et al. 2019). Here, the hypothesis is that the model grid spacing can significantly impact the climatological amplitude of storm track by air–sea coupling (Nakamura et al. 2008; Sampe et al. 2010) and resolving mesoscale latent heating (Willison et al. 2013). In the lower-level troposphere, the correlation coefficient between the model grid spacing and the climatological amplitude of the NPST represented by the $v' T'_{850}$ is 0.28 (Fig. 11a), failing to pass the significance test, which suggests that there is no significant relationship between the intensity of $v' T'_{850}$ and model grid spacing. However, in the upper-level troposphere, the
correlation coefficient between the model grid spacing and the intensity of the NPST characterized by the $y_{0}^{250}$ is 0.64 (Fig. 11b) with a 99% confidence level, which indicates that the higher the spatial resolution is, the larger the climatological $y_{0}^{250}$ amplitude is. The significant out-of-phase relation between the model grid spacing and the NPST intensity in the upper-level troposphere is consistent with previous studies (Chang et al. 2013). The model grid spacing can explain approximately 41% of the variance in NPST represented by the $y_{0}^{250}$.

The results above may reveal that the sensitivities of the NPST in the lower-level and upper-level troposphere to the model horizontal resolution are different. The decrease in the model grid spacing can significantly enhance the amplitude of the NPST characterized by the $y_{0}^{250}$ but has no obvious effect on the NPST represented by the $y_{0}^{T_{850}}$. To exclude the influences of physical processes between different models, two models from HighResMIP, IPSL-CM6A-ATM-HR, and IPSL-CM6A-LR with horizontal resolutions of 50 and 250 km, respectively, are used for further validation. We see that the $y_{0}^{T_{850}}$ in the high-resolution model IPSL-CM6A-ATM-HR (Fig. 12a) is essentially in agreement with that in the low-resolution model IPSL-CM6A-ATM-HR (Fig. 12b). The RMSEs of the high-resolution model IPSL-CM6A-ATM-HR and the low-resolution model IPSL-CM6A-LR are 0.75 and 0.76 K m s$^{-1}$, respectively. In the upper-level troposphere, the $v'v_{250}$ over the eastern North Pacific in the high-resolution model IPSL-CM6A-ATM-HR (Fig. 12c) is closer to the NCEP reanalysis than that in the low-resolution model IPSL-CM6A-LR (Fig. 12d), and their RMSEs are 7.57 and 10.94 m$^2$s$^{-2}$, respectively. It can be concluded that the $v'T_{850}$ is not as sensitive as the $v'v_{250}$ to the model grid spacing.

2) EAST ASIAN WINTER MONSOON

Previous studies showed that there is a remarkable out-of-phase relationship between the NPST and EAWM (Nakamura et al. 2002; Lee et al. 2010). It can be observed from Fig. 11c that the EAWM and NPST represented by the $v'T_{850}$ are negatively correlated and their correlation coefficient is $-0.60$. However, the removal of the particular outlier drastically impacts the result; if the INM-CM4-8 is excluded, their relationship will not be significant. In contrast, the correlation coefficient between the EAWM and the $y_{0}^{250}$ is 0.47 ($R^2 = 22\%$) with 95% confidence level, suggesting a prominently in-phase relation between them (Fig. 11d). The stronger the climatological EAWM is, the more the climatological $y_{0}^{T_{850}}$ in the upper-level troposphere intensifies. It needs to be emphasized that the interannual relationship between the EAWM and NPST should be significantly negative regardless of the way the storm track is characterized (Nakamura et al. 2002; Lee et al. 2010; Yang et al. 2020b). However, the relationship with the EAWM shows an inconsistency between the $y_{0}^{T_{850}}$ and $y_{0}^{250}$ once again. A possible reason is discussed in the next section.

3) ATMOSPHERIC BAROCLINICITY

The atmospheric baroclinicity marked by the Eady growth rate (EGR; Lindzen and Farrell 1980) can be written as

$$\sigma = 0.31N^{-1}f \frac{\partial u}{\partial z},$$  \hspace{0.5cm} (6)

where $f$ is the Coriolis parameter, $N$ is the Brunt–Väisälä frequency, $u$ is the horizontal wind, and $\sigma$ is the Eady growth rate. We calculated the EGR at 850 hPa and the vertical derivatives.
of wind speed at 850 hPa are obtained by central difference with 1000 and 700 hPa. According to the linear theory of baroclinic instability (Charney 1947; Eady 1949; Lindzen and Farrell 1980), the atmospheric baroclinicity measured by EGR is expected to feed synoptic-scale eddies to form a well-organized storm track (Hoskins and Valdes 1990). There is a significant positive relationship between the climatological NPST quantified by \( y_{0} T_{0}^{'850} \) and the lower-tropospheric atmospheric baroclinicity (Fig. 13a), and their correlation coefficient is 0.59. It must be noted that the climatological \( y_{0} T_{0}^{'850} \)-EGR relationship will not be significant if INM-CM4-8 is removed. In addition, the climatological \( y_{250} T_{250}^{'} \)-EGR relationship is also not significant (Fig. 13b), indicating that the enhanced climatological lower-tropospheric atmospheric baroclinicity does not coincide with the intensified climatological NPST. In fact, the inconsistency between the atmospheric baroclinicity and the NPST occurs in not only the midwinter suppression but also some previous studies (e.g., Kuwano-Yoshida and Minobe 2017; Ma et al. 2017; Park et al. 2010; Yang et al. 2020c), which suggests that their relationship is quite puzzling.

Based on empirical orthogonal function and singular value decomposition analyses performed on kinds of different eddy measures, including the geopotential height variance, stream-function variance, lower-tropospheric heat flux, meridional wind variance, and upper-tropospheric momentum flux, some noteworthy differences among different eddy measures were documented by Chang and Fu (2002). It can be observed from Fig. 13c that the climatological \( y_{0} T_{0}^{'850} \) is not consistent with that of the \( y_{250} T_{250}^{'250} \) without a significant correlation relationship. Combined with the differences in midwinter suppression, sensitivity to the model horizontal resolution and inconsistent relationship with EAWM discussed above, we think that there may be internal differences between these two eddy variables. Some additional discussions on the inconsistency between the climatological \( y_{0} T_{0}^{'850} \) and \( y_{250} T_{250}^{'250} \) can be found in the last paragraph of this section. In addition, some previous studies show that different choices of variables used to represent the storm-track activity will lead to different results (Inatsu and Hoskins 2004; Chang et al. 2012; Small et al. 2014). For example, the zonal-mean \( y_{250} T_{250}^{'250} \) shows an increasing trend in the north of 40°N during the boreal wintertime in the case of global warming, while a decreasing trend is noticed when using the variance of filtered sea level pressure as an indicator to quantify the storm tracks (Chang et al. 2012). Therefore, it is necessary to use additional alternative eddy variables to represent storm-track activity in the future study.

4) BAROCLINIC ENERGY CONVERSION

According to Cai et al. (2007), the baroclinic energy conversion (BEC) from eddy available potential energy to eddy kinetic energy can be expressed as

\[
\text{BEC} = - \frac{R}{g} \left( \frac{\rho}{\rho_0} \right)^{\gamma/c} \left( \omega \nabla T \right).
\]

(7)

The transient eddy transports heat not only poleward but also upward (Chang et al. 2002), both of which are closely related under the constraint of the thermal wind relationship. The intensity of the climatological mean \( y_{0} T_{0}^{'850} \) is positively correlated with the BEC, and their correlation coefficient is 0.77 with 99% confidence level (Fig. 13d), which means that the intensified \( y_{0} T_{0}^{'850} \) significantly corresponds to enhanced BEC. There are two reasons for their close relationship. On the one hand, they maintain the thermal wind balance through nongeostrophic circulation, in which warm air rises and cold air sinks. Nongeostrophic circulation is required for transient eddies in the vertically sheared zonal flow. In the storm-track region, the spatial distribution of the vertical eddy heat flux corresponds to the poleward eddy heat flux. On the other hand, through BEC, the vertical eddy heat flux can convert eddy
available potential energy to eddy kinetic energy, which directly contributes to the growth and development of the NPST. It is clear that these two eddy variables represent different physical properties of the storm track. Given that the \( y_0 T_0 \) is included in the energy conversion from mean available potential energy into eddy available potential energy and \( y_0 y_0 \) is included in the eddy kinetic energy, the nonsignificant correlation between \( y_0 T_0 \) and \( y_0 y_0 \) (Fig. 13c) suggests that energy conversion from mean available potential energy into eddy available potential energy does not have a large influence on eddy kinetic energy. Although parameterizations of diabatic processes may have a large influence, it can be inferred that eddy available potential energy generation through diabatic heating processes may be responsible for the stronger \( y_0 y_0 \), which is consistent with the result about the horizontal resolution of the models and previous studies (e.g., Willison et al. 2013). In addition, according to the local kinetic energy budget of high-frequency eddies (T. Jiang et al. 2013), the lower-tropospheric NPST and upper-tropospheric NPST may be dominated by different physical process, which may be responsible for these inconsistent results between \( y_0 T_0 \) and \( y_0 y_0 \).

4. Summary and discussion

Based on the NCEP reanalysis, the present study evaluates the climatological amplitude of the lower-tropospheric NPST quantified by \( y_0 T_0 \) and its geographical distribution derived from the historical runs in 18 CMIP6 models and conducts a comparison with the MME of 13 CMIP5 models. The midwinter suppression in these selected CMIP6 models is also evaluated. In addition, we explore the possible factors that may impact the simulation results from the perspective of model grid spacing, the EAWM, atmospheric baroclinicity, and BEC. The main conclusions are as follows:

In general, all the selected CMIP6 models can capture the major characteristics of the cold-season \( y_0 T_0 \) well, but the differences among them are evident. The position of the climatological amplitude peak area over eastern Japan is well reproduced by these CMIP6 models. Except for the overall overestimation of the climatological amplitude by INM-CM4-8, the simulated amplitudes in almost all the CMIP6 models are underestimated in the southwestern flank of the climatological mean but overestimated in the southeastern flank. The spatial pattern in the CMIP6 models, especially CESM2, CESM2-WACCM, and NorESM2-LM, is basically consistent with that in the NCEP reanalysis with spatial correlation coefficients as high as 0.98. Generally, NorESM2-LM and CESM2-WACCM show a relatively strong capability to simulate the \( y_0 T_0 \) climatological amplitude with low RMSE. In addition, although there is an exception (FGOALS-g3), the results in most CMIP6 models, including the MME, are more consistent with the NCEP reanalysis than those in previous versions participating in CMIP5. Furthermore, the differences among the CMIP6 models are smaller than those among the CMIP5 models.

The LAIs from the MME of the CMIP6 models and BCC-ESM1 are basically consistent with that from the NCEP reanalysis with little bias. More than half of the selected CMIP6 models and the MME of CMIP5 underestimate the LAI.
Except for CESM2-WACCM, INM-CM5-0, and NorESM2-LM, the remaining individual CMIP6 models and both the MME of the CMIP6 models and the MME of the CMIP5 models overestimate the LOI.

For the subseasonal variability represented by the $v'/T'_{250}$, nearly half of the CMIP6 models exhibit a double-peak structure, suggesting obvious midwinter suppression. Except for the peak of INM-CM4-8, which occurs in midwinter, the remaining CMIP6 models have single maximum peaks in late autumn or early winter. The apparent midwinter suppression in the NPST represented by the $v'/T'_{250}$ is reproduced by all the selected CMIP6 models. In addition, there is a positive relationship between the midwinter suppression and the intensity of the upper-tropospheric jet; however, their relationship will not be significant if the outliers (INM-CM4-8 and NEM3) are removed.

In the lower-level troposphere, the NPST climatological amplitude quantified by the $v'/T'_{850}$ has no significant relationship with the model grid spacing. In contrast, there is a significant out-of-phase relation between the NPST amplitude characterized by the $v'/T'_{250}$ and the model grid spacing, which indicates that the $v'/T'_{850}$ is not as sensitive as the $v'/T'_{250}$ to the model grid spacing. The NPST climatological amplitude marked by $v'/T'_{850}$ is negatively correlated with the intensity of the EAWM, but their relationship is dependent on the particular outlier (INM-CM4-8). In contrast, there is a significant in-phase correlation relationship between the climatological $v'/T'_{250}$ and the EAWM. In addition, although the climatological low-level atmospheric baroclinicity is positively correlated with the climatological $v'/T'_{850}$ when the outlier (INM-CM4-8) is not removed, it is uncorrelated with the climatological $v'/T'_{250}$, which suggests a quite puzzling relationship between the atmospheric baroclinicity and the NPST. Furthermore, the climatological amplitude of $v'/T'_{850}$ is significantly positively correlated with the BEC from the eddy available potential energy to the eddy kinetic energy.

According to the linear theory of baroclinic instability, the underestimated spring NPST is expected to associate with underestimated upper-tropospheric jet. However, it can be found that the bias distributions of upper-tropospheric jet (Fig. 10) are mismatched with those of NPST (Figs. 6 and 9), indicating that the underestimated spring NPST cannot be attributed to the biases in upper-tropospheric jet. Compared with the winter NPST, the variations of spring NPST receive much less attention by scientific community, and therefore it is indeed difficult to confirm the reasons of their biases. Recently, Yang et al. (2020c) found that only eastern Pacific El Niño episodes have a distinct influence on the strength and movement of the winter NPST. A further study finds a significant role of decaying phase of two types of El Niño, especially central Pacific El Niño, on the spring NPST. Therefore, the impact of sea surface temperature may shed light on the underestimated spring NPST in CMIP6 models, which remains to be further studied.

Apart from the possible difference between the lower-tropospheric $v'/T'$ and the upper-tropospheric $v'/v'$, our recent research finds that in some ways, the nature of the surface storm track significantly differs with the tropospheric storm track. For example, the structure of the surface storm-track response to El Niño events is different from that of the tropospheric storm track without intensification in the exit region and extension equatorward. In addition, the interdecadal change in the interannual relationship between the NPST and EAWM in the mid-1980s (Yang et al. 2021) excludes the surface storm track. Therefore, we believe that more in-depth research is needed to explore the differences in the representation of storm tracks measured by different kinds of eddy quantities. With the development of emergent constraints, we hope our study contributes to reducing the uncertainty in climate prediction regarding storm-track activity.

Acknowledgments. The authors thank three anonymous reviewers for their helpful and crucial comments, which improved the manuscript substantially. Many thanks are given to the editor Hisashi Nakamura for his help in finding excellent reviewers. This research was supported by the National Key Research and Development Program of China (Grant 2018YFC1505901), the National Natural Science Foundation of China (Grants 41490642, 4160501, and 4152014008), and the Key Research Program of National University of Defense Technology (ZK17-02-010). The WCRP CMIP6 datasets were obtained online (https://esgf-node.llnl.gov/search/cmip6/). The NCEP–NCAR reanalysis dataset was obtained online (https://www.esrl.noaa.gov/psd/data/gridded/reanalysis/).

REFERENCES


Yang, M., X. Li, R. Zuo, X. Chen, and L. Wang, 2018: Climatological and interannual variability of winter North Pacific storm track


