A forensic investigation of climate model biases in teleconnections: the case of the relationship between ENSO and the northern stratospheric polar vortex

Shen, X. ORCID: https://orcid.org/0009-0003-7116-1397, Kretschmer, M. ORCID: https://orcid.org/0000-0002-2756-9526 and Shepherd, T. G. ORCID: https://orcid.org/0000-0002-6631-9968 (2024) A forensic investigation of climate model biases in teleconnections: the case of the relationship between ENSO and the northern stratospheric polar vortex. Journal of Geophysical Research, 129 (19). e2024JD041252. ISSN 0148-0227 doi: https://doi.org/10.1029/2024JD041252 Available at https://centaur.reading.ac.uk/118492/

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2024JD041252

A Forensic Investigation of Climate Model Biases in Teleconnections: The Case of the Relationship Between ENSO and the Northern Stratospheric Polar Vortex

Xiaocen Shen[1], Marlene Kretschmer[1,2], and Theodore G. Shepherd[1]

[1]Department of Meteorology, University of Reading, Reading, UK, [2]Institute for Meteorology, University of Leipzig, Leipzig, Germany

Key Points:

- Although the model shows an opposite-signed teleconnection to observations, it successfully captures the known underlying causal pathway
- The existence of an additional pathway explains the discrepancy, with no strong observational evidence for or against it
- A bias adjustment of this additional pathway aligns the modeled relationship with observations and adjusts the related climate impacts

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Shen,
xiaocen.shen@reading.ac.uk

Citation:

Shen, X., Kretschmer, M., & Shepherd, T. G. (2024). A forensic investigation of climate model biases in teleconnections: The case of the relationship between ENSO and the northern stratospheric polar vortex. *Journal of Geophysical Research: Atmospheres*, *129*, e2024JD041252. https://doi.org/10.1029/2024JD041252

Received 26 MAR 2024
Accepted 17 SEP 2024

Author Contributions:

Conceptualization: Xiaocen Shen, Marlene Kretschmer, Theodore G. Shepherd
Formal analysis: Xiaocen Shen
Funding acquisition: Marlene Kretschmer, Theodore G. Shepherd
Investigation: Xiaocen Shen, Marlene Kretschmer, Theodore G. Shepherd
Project administration: Marlene Kretschmer, Theodore G. Shepherd
Software: Xiaocen Shen
Visualization: Xiaocen Shen
Writing – original draft: Xiaocen Shen

Abstract Teleconnections are crucial in shaping climate variability and regional climate change. The fidelity of teleconnections in climate models is important for reliable climate projections. As the observed sample size is limited, scientific judgment is required when models disagree with observed teleconnections. We illustrate this using the example of the relationship between El Niño-Southern Oscillation (ENSO) and the northern stratospheric polar vortex (SPV), where the MIROC6 large ensemble exhibits an ENSO-SPV correlation opposite in sign to observations. Yet the model well captures the upward planetary-wave propagation pathway through which ENSO is known to affect the SPV. We show that the discrepancy arises from the model showing an additional linkage related to horizontal stratospheric wave propagation. Observations do not provide strong statistical evidence for or against the existence of this linkage. Thus, depending on the research purpose, a choice has to be made in how to use the model simulations. Under the assumption that the additional linkage is spurious, a physically-based bias adjustment is applied to the SPV, which effectively aligns the modeled ENSO-SPV relationship with the observations, and thereby removes the model-observations discrepancy in the surface air temperature response. However, if one believed that the additional linkage was genuine and was undersampled in the observations, a different approach could be taken. Our study emphasizes that caution is needed when concluding that a model is not suitable for studying teleconnections. We propose a forensic approach and argue that it helps to better understand model performance and utilize climate model data more effectively.

Plain Language Summary Teleconnections are usually manifested as recurring patterns which link weather and climate anomalies over large distances across the globe. Therefore, they play an important role in shaping climate variability and regional climate change. The limited observational records make it difficult to draw robust conclusions. To avoid this concern, climate model simulations are widely used as they can provide a larger sample size. However, climate models do not always agree with the observed teleconnections. When this happens, scientific judgment is needed to assess the suitability of models for studying teleconnections. Here, based on an example, we introduce a forensic approach to assessing the fidelity of teleconnections in models, where a physically-based analysis is required to understand the origin of the model-observations discrepancy. By "forensic" we mean a scientific approach that considers various explanations for a phenomenon and attempts to disprove them through physical evidence, thereby narrowing down the possibilities. In this way, one can better understand the reason for the difference between model and observations and make an informed use of the climate model outputs.

1. Introduction

Teleconnections are essential parts of the climate system. They are usually manifested as recurring patterns which link weather and climate anomalies over large distances across the globe (Barnston & Livezey, 1987; Wallace & Gutzler, 1981), therefore strongly contributing to the variability of global-scale atmospheric circulation (Hoskins & Woollings, 2015). One striking example is El Niño-Southern Oscillation (ENSO), the most prominent year-to-year internal variability in the climate system. Although ENSO itself occurs in the tropical Pacific and is reflected as a fluctuation between unusually warm and cold conditions (Bjerknes, 1969), it excites wave-trains propagating into the extratropics, which then strongly affect the precipitation and temperature over mid-high latitudes (Trenberth et al., 1998; Van Loon & Madden, 1981). In this way, it can also modulate other teleconnections, for example, the surface response to the Quasi-Biennial Oscillation (e.g., Kumar et al., 2024). Moreover, the

Writing – review & editing: Marlene Kretschmer, Theodore G. Shepherd

extratropical circulation response will influence the upward propagation of planetary waves in the midlatitudes (Domeisen et al., 2019), thereby leading, for instance, to significant changes of the Northern Hemisphere stratospheric polar vortex (SPV), which can itself affect midlatitude circulation (Garfinkel & Hartmann, 2007; Manzini et al., 2006; Van Loon & Labitzke, 1987). ENSO, the SPV, as well as the ENSO-SPV teleconnection are suggested to have a strong influence on European winter climate (Brönnimann, 2007; Ineson & Scaife, 2009) and serve as key sources of predictability on both the seasonal and interannual timescale (Domeisen et al., 2015; Dunstone et al., 2016). Hence, studying their linkages not only helps to better understand the atmospheric circulation variability, but is also key to improving prediction skill.

Analyzing teleconnections in observations is often challenging. For example, it is sometimes debatable whether an observed pattern is truly a physical teleconnection or just a statistical artifact (Butler & Polvani, 2011; Runge et al., 2014; Wallace, 2000). Part of this debate arises from the fact that a correlation does not always represent a physical causal linkage, since a strong correlation can sometimes instead reflect co-varying changes in response to other factors (Kretschmer et al., 2021). In addition, there are large internal variabilities in the climate system which might result in unstable statistical relationships (Dong & McPhaden, 2017; Wang et al., 2023) beyond what would be expected from sampling uncertainty alone, and can be reflected in the nonstationarity of the relationships (Ayarzagüena et al., 2019; Butler & Polvani, 2011; Domeisen et al., 2019), making it difficult to draw solid conclusions. In this light, while the rather short observational data is invariably the starting point for the analysis of teleconnections, climate model simulations are used to provide a more robust understanding as they have large sample sizes (Kosaka et al., 2009; Monnin et al., 2022; Qasmi et al., 2017). To use a climate model to conduct the analysis for a specific purpose, it is common to first evaluate the usefulness of the model by comparing model statistics against observations (Sillmann et al., 2013; Willmott, 1982), with discrepancies in the teleconnections, for example, as evaluated in terms of correlations, usually considered to be a failure of the model to capture the teleconnection. However, caution is needed in concluding that the model is deficient because such discrepancies might be the result of internal variability, confounding factors, or/and state dependence (Ineson et al., 2023; Jain et al., 2023; Mudhar et al., 2024; Notz, 2015).

In the context of assessing model discrepancies in teleconnections, here we advocate that a more physically-based approach combined with the statistical diagnostics is required to better understand the origin of the model biases. With the evidential basis, a scientific judgment can then be made to decide whether a model is fit for a given scientific purpose. We refer to this approach as a *forensic* investigation, meaning a scientific approach that considers various explanations for a phenomenon and attempts to disprove them through physical evidence, thereby narrowing down the possibilities. In this work, we study the relationship between ENSO and the SPV as an example of such an approach.

The description of the datasets is provided in Section 2. The discrepancy between model and observations is analyzed in Section 3, followed by a physically-based analysis to understand its origin in Section 4. Section 5 presents the quantification of the strength of the observational evidence for or against the identified physical reason for the model-observations discrepancy. Section 6 discusses how to use the model output, providing a physically-based model bias adjustment approach as one option, along with application to the impacts on Northern Hemispheric winter surface temperature. A summary is presented in Section 7. Section 8 presents guidance on our forensic investigation and a related discussion.

2. Data

The monthly ERA5 reanalysis dataset is used as the observational reference for the atmospheric variables (Hersbach et al., 2020), and is used at a $2.5 \times 2.5°$ horizontal resolution. In order to have the largest possible sample size, the 1950 to 2023 period is used, since reanalysis from 1950 has been widely used for northern hemisphere studies (i.e., Yang et al., 2023; Yessimbet et al., 2022). The monthly averaged sea surface temperature (SST) from 1950 to 2023 is adopted from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) dataset (Rayner et al., 2003). In the observations, the long-term trend is removed by subtracting the linear trend. The climate model used here is the historical simulation from the sixth version of the Model for Interdisciplinary Research on Climate (MIROC6) large-ensemble dataset. It has 50 members covering the period from 1850 to 2014, making it one of the largest ensembles so far (Shiogama et al., 2023). Monthly data is used and the anomaly in the model is computed by extracting the ensemble mean for each year, by which the possible influence of long-term trends can be excluded. A

linear detrending (as with the observations) for individual ensemble members is also tested and leads to virtually identical results, but is a less precise form of trend removal when a model ensemble is available.

Here the Niño3.4 index is used, which is the SST averaged over [5°N–5°S, 120°–170°W] (Trenberth, 1997), with its mean anomaly value during November to January representing the ENSO state of that winter. The ENSO state is categorized with a ±0.5 standard deviation threshold when doing the classification analysis but the results are not sensitive to the threshold value. The strength of the SPV is denoted as the zonal mean zonal wind at 10 hPa and 60°N (Baldwin & Thompson, 2009; Charlton & Polvani, 2007). The eddy heat flux is computed as [v′T′] at 100 hPa averaged between [45°-75°N], where the prime is the deviation from the zonal mean and the bracket denotes the zonal mean, v denotes the meridional winds, and T denotes temperature. It represents the upward flux of planetary waves into the stratosphere (Polvani & Waugh, 2004). The zonal mean Eliassen-Palm (EP) flux (Andrews et al., 1987) is employed to illustrate the horizontal and vertical propagation of the waves. As monthly mean data is used, the wave propagation represents that of stationary planetary waves. In the following, ENSO, SPV, and [v′T′] are used to refer to the Niño3.4 index, SPV intensity index, and eddy heat flux, respectively, if not otherwise specified.

3. Relationship Between ENSO and SPV in the Model and Observations

Firstly, we check the average linear relationship between ENSO and SPV in the model and in the observations. Figure 1a shows the correlation coefficient between the November-January mean Niño3.4 index and the December-February averaged zonal mean zonal wind anomaly across all the members in the MIROC6 model. This chosen time lag reflects the fact that both the extratropical response and the resulting upward wave propagation are known to take time to establish themselves (Ineson & Scaife, 2009; Trenberth et al., 1998). Figure 1a shows a positive correlation between the northern polar stratospheric zonal mean zonal wind and ENSO, suggestive of a first-order linear relationship between ENSO and SPV with an intensified SPV after El Niño and a weakened SPV following La Niña. However, this is opposite in sign to the negative correlation seen in the observations (Figure 1b), which suggests instead that El Niño induces a weakening of the SPV and La Niña a strengthening, as documented by previous studies (e.g., Brönnimann, 2007; Manzini et al., 2006). In the Southern Hemisphere, on the contrary, the modeled polar stratospheric zonal mean zonal wind response is very similar to observations in the low to middle stratosphere (Figures 1a and 1b). In this study, we focus on the opposite-signed correlation between MIROC6 and observations in the Northern Hemispheric SPV response to ENSO, which is an apparent discrepancy.

This discrepancy in sign of the ENSO-SPV correlation between MIROC6 and observations (as also documented in Manzini et al., 2024) is further explored in Figures 1c and 1d. Although we choose the zonal wind at a single latitude (star in Figures 1a and 1b), the results are not sensitive to this definition, as the entire polar stratospheric zonal wind in the Northern Hemisphere shows a consistent response in both observations and model data (Figures 1a and 1b). There is clearly a lot of scatter in the relationships, illustrating that El Niño and La Niña can both be related to either weakening or intensification of the SPV. Whether the ENSO-SPV relationship is nonlinear is a subject of recent research (Jiménez-Esteve & Domeisen, 2019; Manzini et al., 2024). Certainly the scatter plot for observations (Figure 1d) is not particularly linear (reflected in the fact the regression coefficient is not statistically significant), but the limited sample size precludes firm conclusions in this respect (Domeisen et al., 2019). We thus restrict our attention to the linear relationship as a basic summary statistic. The regression coefficient (R) of SPV with respect to Niño 3.4 is 0.56 ms^{-1}/°C in MIROC6, which is opposite in sign to the value of −1.33 ms^{-1}/°C in ERA5. As the documented pathway for how ENSO influences SPV is via anomalous upward-propagating planetary waves (Domeisen et al., 2019), a possible reason for the opposite-signed correlation is that the model does not produce the correct planetary-wave forcing following ENSO, which is conceivable as the model presents a slight westward bias of its SST anomaly in ENSO years (Tatebe et al., 2019). However, the regression coefficient of the eddy heat flux ([v′T′]) averaged during December to February with respect to Niño 3.4 shows the same positive sign in the model (0.24 K·ms^{-1}/°C) and observations (0.64 K·ms^{-1}/°C), although the value is considerably smaller in MIROC6. The positive value means that enhanced upward-propagating planetary waves are associated with El Niño, whereas vertical wave propagation is suppressed following La Niña.

A closer inspection of the tropospheric response, that is, the geopotential height at 500 hPa, shows a similar pattern in the extratropics between MIROC6 and ERA5 (shading in Figures 2a and 2d), and the structure of the climatological planetary waves (contours in Figures 2a and 2d), computed as the deviation from the zonal mean

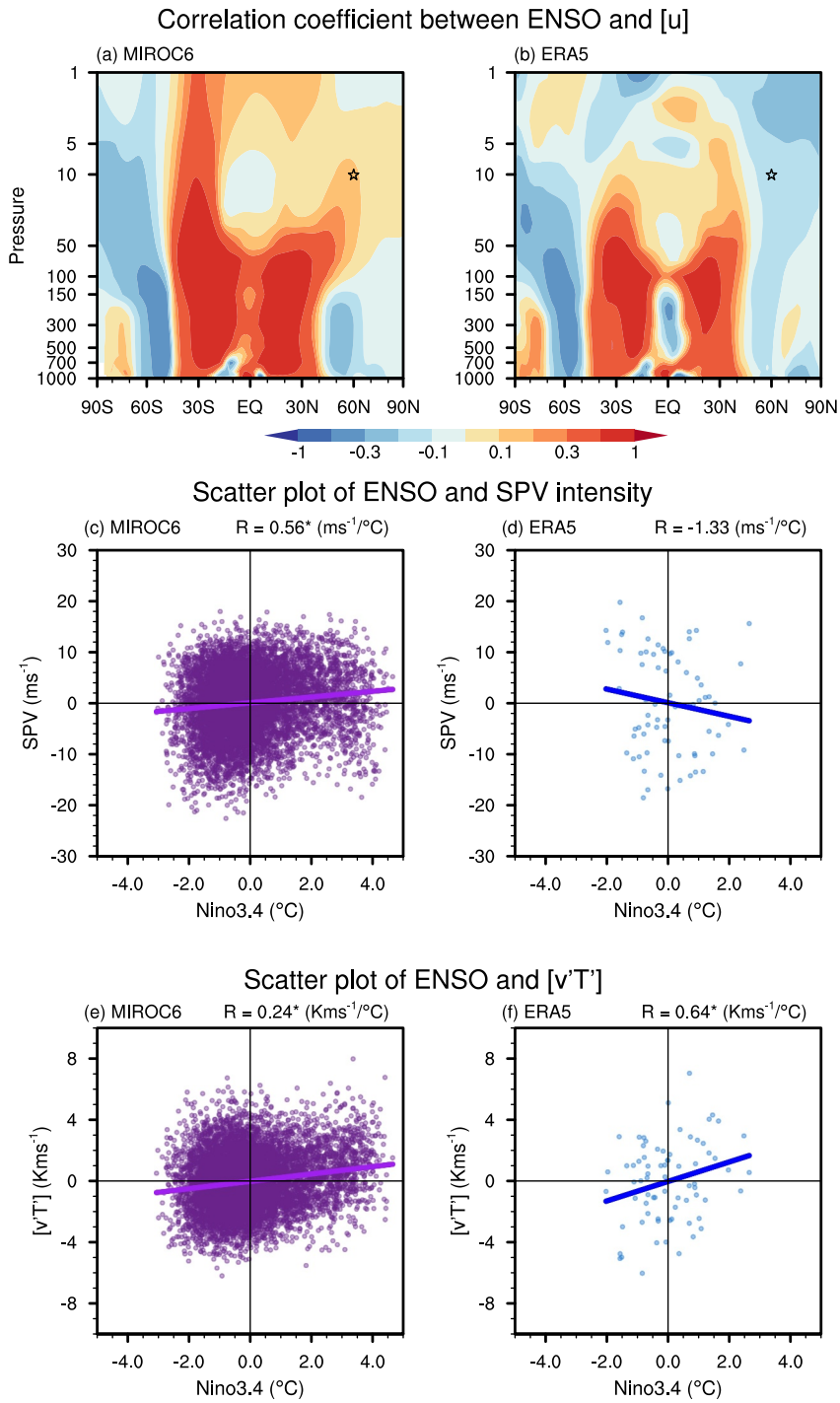

Figure 1. Relationship between ENSO and SPV in MIROC6 and ERA5. (a) and (b) Correlation coefficient between the November-January averaged Niño3.4 index and December-February averaged zonal mean zonal wind anomaly. The star denotes the location of zonal mean zonal wind used as the index of Northern Hemisphere SPV intensity. (c) and (d) Scatter plot of the Niño3.4 index versus SPV intensity. (e) and (f) Scatter plot of Niño3.4 index versus December-February averaged eddy heat flux at 100 hPa between [45°-75°N]. The solid lines denote the linear regression lines, with the regression coefficient (R) shown on the upper right. The asterisk denotes that the regression coefficient is statistically significant at the 95% confidence level. The left column shows the MIROC6 results, while the right column is based on ERA5.

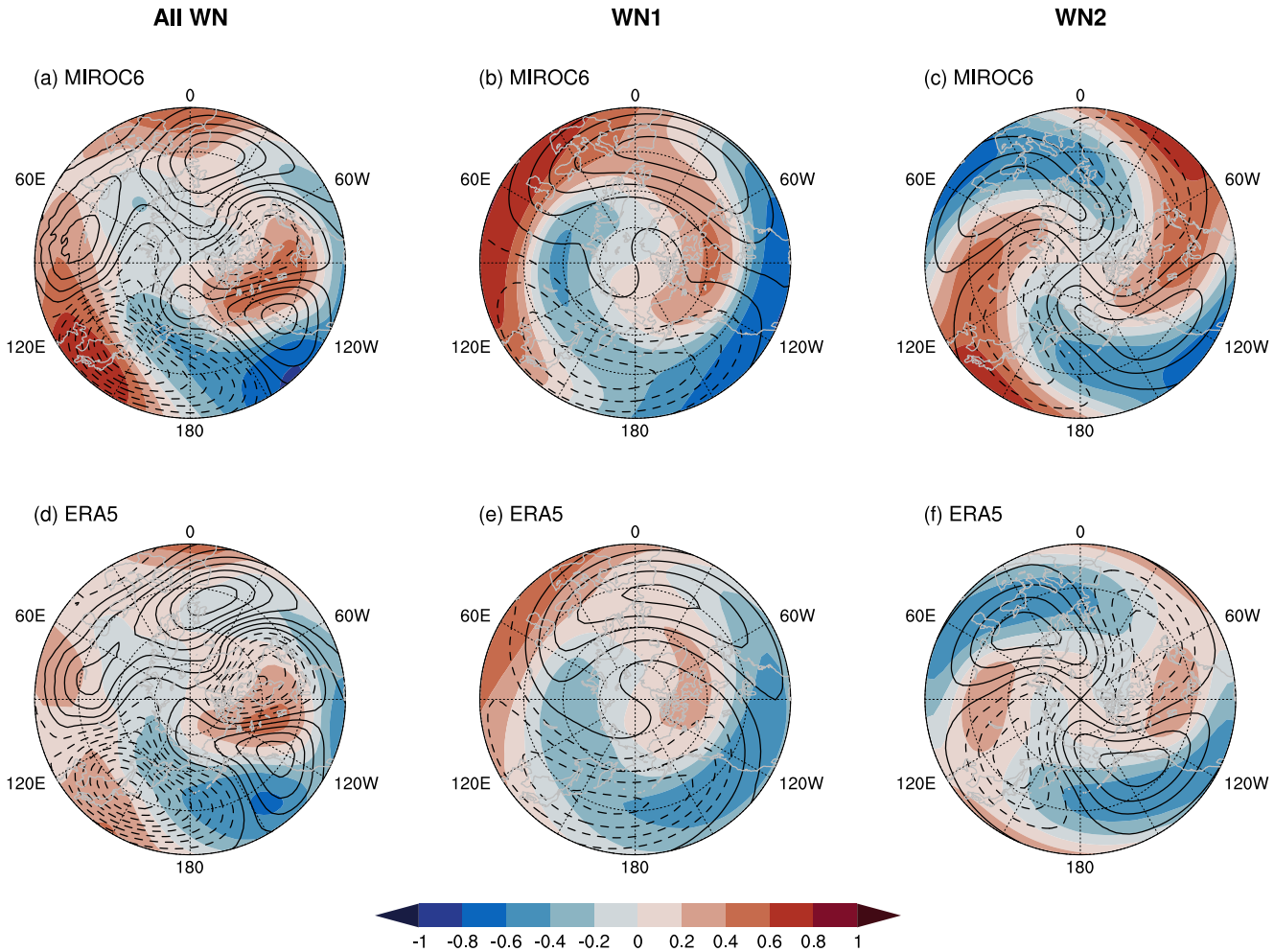

Figure 2. Extratropical response following ENSO. (a) and (d) Correlation coefficient between November-January averaged Niño3.4 index and December-February averaged Z500 (shading). Contours show the climatological planetary waves (contour interval is 25 gpm), computed as the deviation from the zonal mean. (b) and (e), and (c) and (f), as in (a) and (d), but for the wavenumber one (WN1) and wavenumber two (WN2) component, respectively. The upper panel shows the results in MIROC6, whereas the lower panel presents the ERA5 results.

averaged during December to January, is also consistent between the model and observations. Hence, the anomalous circulation shows a constructive linear interference with the climatological planetary waves in El Niño years, and a destructive linear interference in La Niña years, leading to the positive correlation between ENSO and upward wave forcing. A similar conclusion holds for the different wavenumber (WN) components, as the responses of both WN1 and WN2 are consistent with the observations (Figure 2). Thus, in terms of the linear relationship for the ensemble mean, MIROC6 successfully reproduces the expected wave forcing response to ENSO. However, this correct wave forcing nevertheless seems to lead to a stratospheric polar zonal flow response to ENSO that is inconsistent with observations, at least in terms of common statistical diagnostics.

One possibility is that the discrepancies between MIROC6 and ERA5 are simply a result of internal variability, as the sample size of the observations is small (Figure 1d). To assess the potential influence of internal variability, we apply the bootstrap method (Efron & Tibshirani, 1994) and sample without replacement in the model data across all the members to obtain a data set with the same sample size as ERA5. This procedure was repeated 1,000 times and the resulting distribution of the correlation coefficients between ENSO and SPV is shown in Figure 3a. For the model, even the 25th percentile is greater than zero, while the observed value is outside the 95% confidence interval. This indicates that although internal variability can lead to a wide range of ENSO-SPV correlations for a sample the size of ERA5, the observed value is still quite extreme. However, this is not the case for the relationship between ENSO and [v'T'], with the model overall producing a similar but weaker correlation

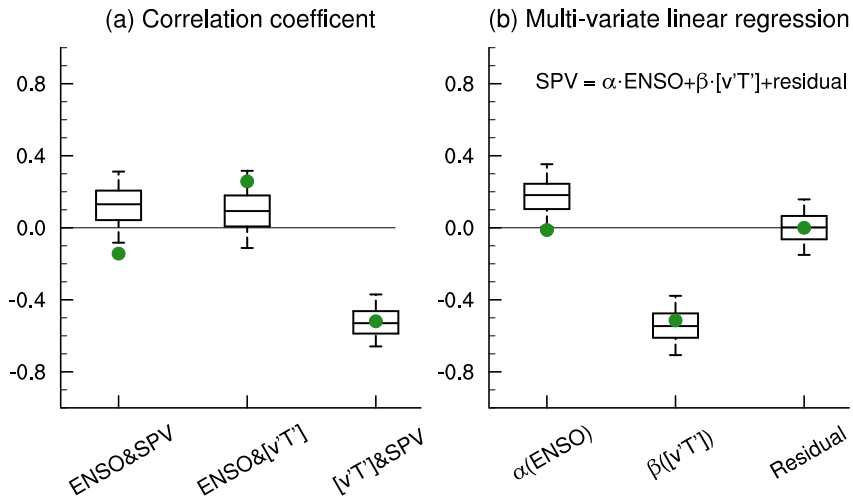

Figure 3. Distribution of the relationships in bootstrapped model groups with the observational sample size. (a) Correlation coefficients between ENSO and SPV, ENSO and [v′T′], and [v′T′] and SPV. (b) Multi-variate linear regression coefficients. The box and whiskers show the fifth percentile, 25th percentile, median, 75th percentile, and 95th percentile in the distribution. The green dots indicate the correlation and regression coefficients in observations. The indices have been standardized in the MLR for comparison with observations.

compared to observations, and the observational value lying within the 95% confidence interval. The distribution of the WN1 and WN2 eddy heat flux response also shows a consistent sign with observations (Fig. S1 in Supporting Information S1), denoting that the MIROC6 model well captures the observed eddy heat flux response to ENSO.

We next assess whether the relationship between the anomalous wave forcing and the SPV response is captured in the model. The negative correlation coefficient between [v′T′] and SPV intensity (Figure 3a) implies that enhanced wave forcing leads to a deceleration of the zonal flow, as expected (Matsuno, 1971), and the model is in perfect agreement with observations. In other words, MIROC6 successfully reproduces the wave-mean flow interaction process with a similar value to observations. Hence, a puzzle exists in that the consistent wave forcing following ENSO and the consistent dynamical process of wave-mean flow interaction nevertheless generates an inconsistent SPV response to ENSO, reflected in the opposite signed ENSO-SPV correlation.

One hypothesis for how this could arise is that in MIROC6, the influence of ENSO on SPV through wave forcing (negative in sign) is obscured by another process linking ENSO and SPV (positive in sign), which is larger in strength thereby leading to the overall positive ENSO-SPV correlation. To test this possibility, we use multi-variate linear regression (MLR) to quantify the influence of ENSO on SPV with the contribution via eddy heat flux excluded, written as

$$\mathrm{SPV} = \alpha \cdot \mathrm{ENSO} + \beta \cdot \left[\mathrm{v}'\mathrm{T}'\right] + \mathrm{residual}. \quad (1)$$

Here β can be interpreted as the influence on SPV by the eddy heat flux anomaly (including that induced by ENSO), and α denotes any other influence of ENSO apart from that mediated by the eddy heat flux anomaly. Note that these indices have been standardized for comparison with observations. We find the *residual* term in MIROC6 shows a median value close to zero (Figure 3b), which is in perfect agreement with observations, indicating that the variability of SPV is almost entirely explained by the variability of eddy heat flux and ENSO. For the β term, both observations and model show a negative value, with the observations falling within the interquartile range (IQR) of the model, confirming the physically consistent representation of wave-mean flow interaction (Figures 1e and 1f). In contrast, the parameter α shows on average a strong positive value (with a median value of 0.18) in MIROC6, but is close to zero in the observations. This confirms that in addition to the influence on the SPV via wave forcing, that is, from ENSO to [v′T′] to SPV, in this model there exists another effect from ENSO to SPV through processes not involving [v′T′], which seems absent in the observations. This other, positive influence of ENSO on SPV (represented by α) exceeds the negative effect mediated by the wave

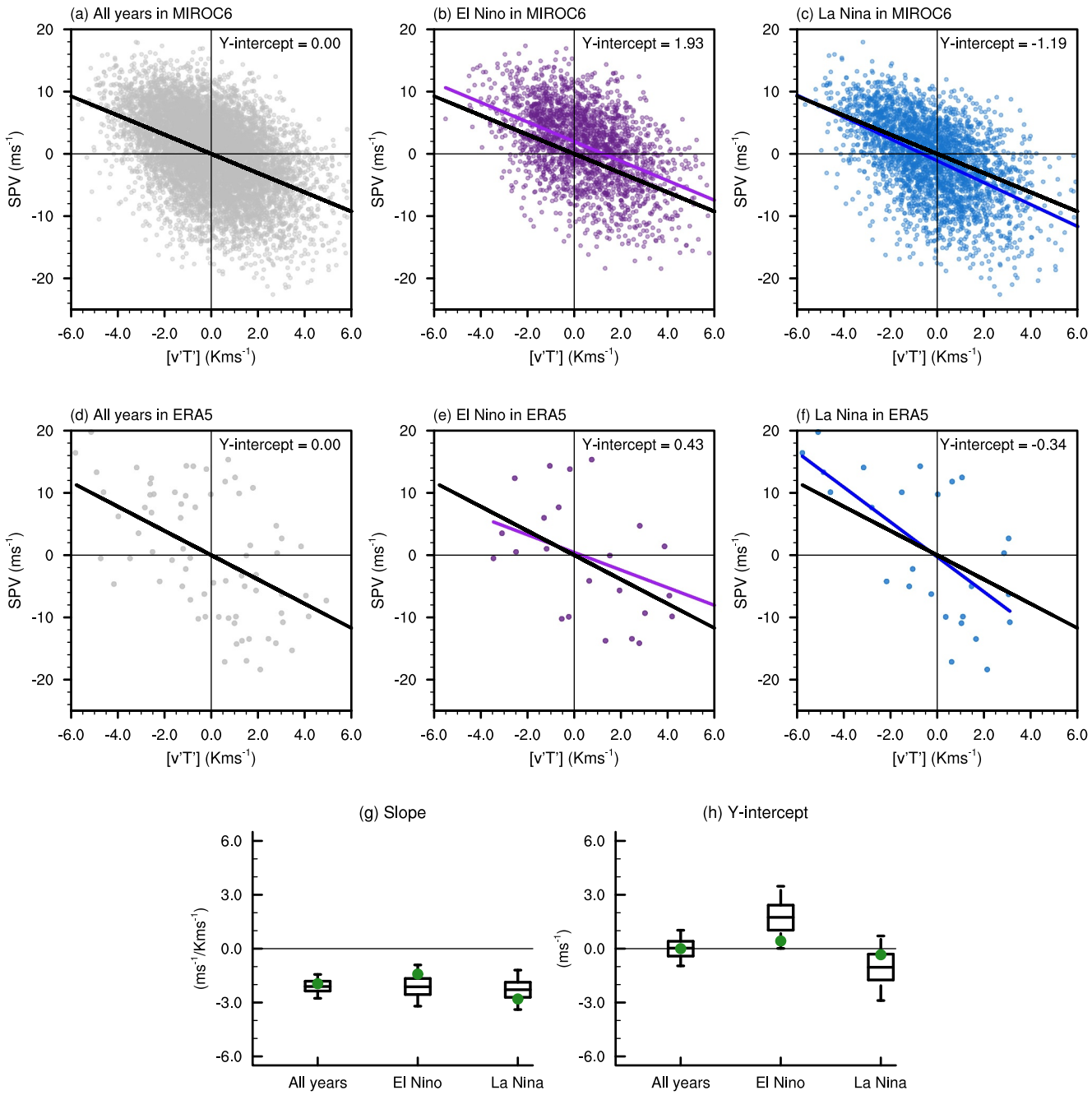

Figure 4. Relationship between SPV and [v′T′] in different backgrounds. (a)–(c) Scatter plot of SPV (*y*-axis) versus [v′T′] (*x*-axis) in all years, El Niño years, and La Niña years, respectively. The colored line indicates the linear fitted line, with the black lines indicating the all-years result. (d)–(f) As in (a)–(c), but for ERA5. (g) and (h) The distribution of slope and *y*-intercept from the linear regression of SPV with respect to [v′T′], which is obtained from the bootstrapped data with the same sample size as the observations (with the observations indicated by the green dots). The selection thresholds of El Niño and La Niña years are 0.5 and −0.5 standard deviations of the ENSO index, respectively.

forcing (represented by the product of the ENSO-[v′T′] correlation and β) and thereby leads to the overall positive correlation between ENSO and SPV in the model. The observational value of α is hardly an outlier, but is small enough in magnitude that the overall correlation in the observations is dominated by the influence via wave forcing, and thus is negative.

We next explore the ENSO-SPV relationship in more detail. To this end, we examine the scatter plots of eddy heat flux (*x*-axis) versus SPV (*y*-axis) in the model under different ENSO background conditions, together with the

linear fitting regression lines (Figures 4a–4c). When all years are considered, the regression line shows a negative slope with the intercept on the *y*-axis very close to zero (Figure 4a), indicating that the change in SPV is proportional to the change in eddy heat flux. However, in El Niño years and in La Niña years, while the slopes are similar to that for all years, the y-intercepts shift to 1.93 ms^{-1} and -1.19 ms^{-1}, respectively (Figures 4b and 4c). This positive *y*-intercept in El Niño years suggests that even when there is no anomalous wave forcing, the SPV tends to strengthen. The negative *y*-intercept in La Niña years indicates, on the contrary, a weakening of the SPV even without anomalous eddy heat flux. This implies a positive correlation between ENSO and SPV after controlling for the influence of ENSO on SPV via wave forcing, which is consistent with the result from the MLR method (non-zero coefficient α in Figure 3b). This difference between ENSO states is robust after considering the possible influence of internal variability. According to the bootstrapped result, the slopes are similar across those three different backgrounds (Figure 4g), showing that the wave-mean flow interaction remains unchanged. However, the *y*-intercept for the model data presents a clear positive shift in El Niño years and a negative shift in La Niña years (Figure 4h).

Thus, the existence of two opposing pathways from ENSO to SPV in MIROC6 can be interpreted as a state-dependence of the [v′T′]-SPV relationship in response to ENSO phase, reflected by the *y*-intercept shift. While the ENSO-induced wave forcing anomaly in MIROC6 would lead to an SPV change consistent with observations, ultimately the ENSO-SPV correlation is dominated by another mechanism which depends on the state of ENSO. The corresponding scatter plots in ERA5 are shown in Figures 4d–4f, in which the *y*-intercept is close to zero for all years, as in the model. For El Niño years, the *y*-intercept shifts to 0.43 ms^{-1}, whereas it shifts to -0.34 ms^{-1} for La Niña years. These observational *y*-intercept changes are of the same sign as in the model, and although smaller in magnitude, they lie within the sampling uncertainty range (Figure 4h). Hence, it appears that the *y*-intercept change in the model holds in observations in both El Niño and La Niña years.

4. Physical Understanding of the Discrepancy Between Model and Observations

In the previous section, the discrepancy between model and observations of the ENSO-SPV correlation was attributed to a state dependence of the [v′T′]-SPV relationship in the model, meaning that the ENSO-SPV relationship is not fully mediated by the wave forcing, as it appears to be in the observations. However, we cannot rule out the possibility that this difference is simply undersampling in the observational record, with the observations being well consistent with the model in terms of the *y*-intercept change (Figure 4), and the α term in Equation 1 is hardly an outlier compared to the model distribution (Figure 3b). Thus, it is important to understand the physical origin of the state dependence.

To that end, the zonal mean zonal wind response to ENSO, together with that of the EP flux and its divergence, are shown in Figure 5. While the plot shows the correlation coefficients, we discuss the El Niño state as an example for an easier understanding. In MIROC6, in response to El Niño, the planetary waves show an anomalous upward propagation in the midlatitudes (around 60°N), which reaches into the stratosphere (above 150 hPa, Figure 5a). Although this is also the case in observations (Figure 5b), an apparent difference appears within the stratosphere. In observations, the upward-propagating waves focus into the polar region, resulting in high-latitude EP flux convergence throughout the stratosphere (blue shading). Through wave-mean flow interaction, this leads to a deceleration of the westerly jet thus a weakening of the SPV (dashed contours). However, in MIROC6, the waves propagate equatorward into lower latitudes after entering the lower stratosphere, leading to high-latitude EP flux divergence and an acceleration of the westerly jet (red box region). Hence, the reason for the opposite SPV response to ENSO between model and observations is seen to be related to the different wave propagation within the stratosphere.

To further study the wave-mean flow interaction process in the polar stratosphere, we conduct a wave-activity budget analysis in this region [red boxes in Figure 5] (Kushner & Polvani, 2004; Sigmond & Shepherd, 2014). An opposite correlation coefficient is seen for the relationship between ENSO and EP flux divergence (EPFD) within the budget region (Figure 6a). In contrast, recall that the correlation between ENSO and [v′T′] has the same sign in the model and in observations (Figure 3a), as well as the correlation between EPFD and [v′T′] (not shown). Note that [v′T′] is proportional to the vertical component of EP flux (Fz, Andrews et al., 1987). As EPFD has the expected positive relationship with SPV intensity in the model (Figure 6a), its opposite relationship with ENSO can potentially explain the opposite correlation between ENSO and SPV.

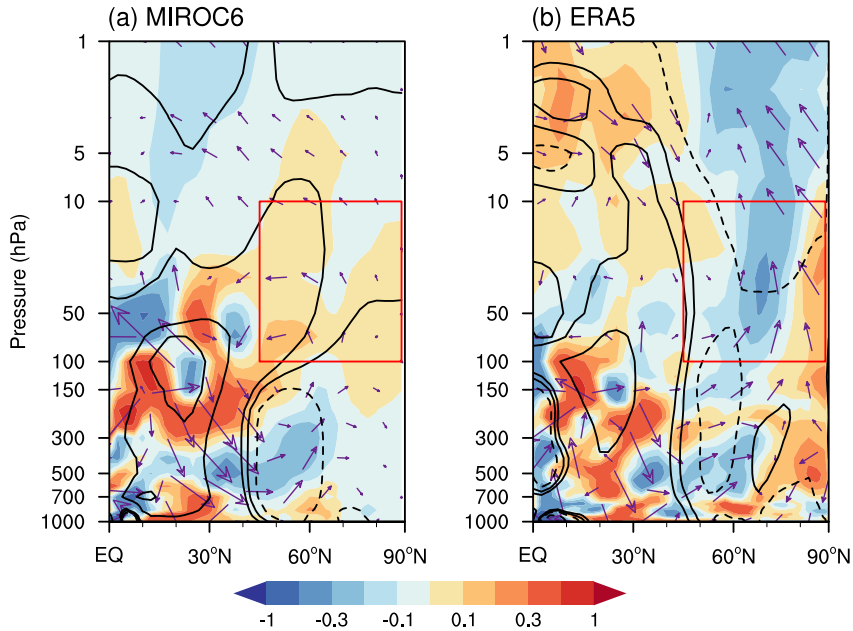

Figure 5. Wave-mean flow interaction in response to ENSO. (a) and (b) Correlation coefficients for MIROC6 and ERA5, respectively, of [u] and Niño3.4 index (contours), EP flux and Niño3.4 index (arrows), and EP flux divergence and Niño3.4 index (shading). Dashed lines denote negative values. The red boxes show the region [45–90°N, 100-10 hPa] used for the budget analysis.

To test if indeed this different ENSO-EPFD response is the cause of the opposite ENSO-SPV relationship in MIROC6, we again use MLR in the following form:

$$\mathrm{SPV} = \alpha \cdot \mathrm{ENSO} + \beta \cdot \mathrm{EPFD} + \mathrm{residual}, \tag{2}$$

where β can be interpreted as the influence of EPFD on SPV, and α denotes the rest of the influence of ENSO that is not mediated via EPFD. Figure 6b shows the standardized regression coefficients in this MLR model for the bootstrapped data, which suggest that the positive link between ENSO and SPV decreases on average to 0.11 after excluding the influence of EPFD in MIROC6, with the observational value now falling within the 95% confidence interval. Furthermore, the scatter plot of the correlation coefficient between ENSO and EPFD (CC[ENSO,

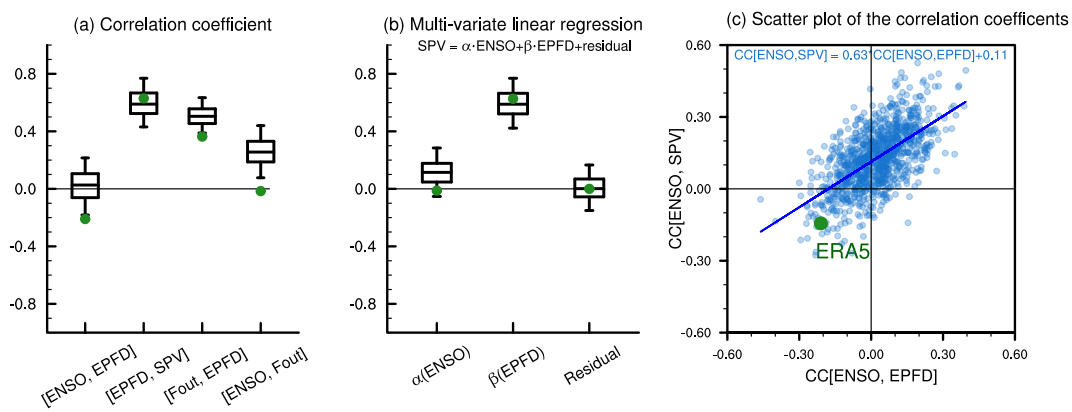

Figure 6. Role of EP-flux divergence. (a) Correlation coefficients between different variables. The conditional correlation between F_{out} and EPFD is computed after excluding the influence of incoming waves. (b) Multi-variate linear regression coefficients. (c) Scatter plot for CC[ENSO, EPFD] (*x*-axis) versus CC[ENSO, SPV] (*y*-axis). The green dot represents the value in ERA5. The distributions in (a) and (b) are obtained from the bootstrapped data with the same sample size as the observations, as are the single samples in (c).

EPFD], *x*-axis) versus the correlation coefficient between ENSO and SPV (CC[ENSO, SPV], *y*-axis), as shown in Figure 6c, suggests a very strong positive correlation between those two statistics with a regression coefficient of 0.63. In particular, the subgroups of the bootstrapped MIROC6 ensemble with a more positive CC[ENSO, EPFD] also tend to have a more positive CC[ENSO, SPV], confirming the role of the EPFD response in influencing the SPV change following ENSO.

We next assess what might contribute to the inconsistent EPFD between observations and model. The EPFD within a latitude-altitude box is influenced by both the inflow and outflow of the planetary wave flux across the boundaries of the box. To capture the outflow of the planetary wave flux, we compute the sum of the vertical wave propagation (Fz) at the upper boundary and the horizontal wave propagation (Fy) at the lower latitude boundary to generate a new index (F_{out}). The incoming wave forcing, described by [v′T′], shows the correct response following ENSO (Figure 3a). However, F_{out} shows an opposite-signed response compared to observations (Figure 6a), with the Fy at the lower latitude contributing the most. Meanwhile, the relationship between F_{out} and EPFD has the same sign as the observations. Note that the correlation between F_{out} and EPFD is strongly influenced by the incoming waves, as the outflow and inflow are not independent. The conditional correlation coefficient we show here is between F_{out} and EPFD after excluding the role of inflow based on partial regression, thereby reflecting their true relationship. This result is consistent with physical understanding, as the stronger outflows contribute to the EP flux divergence, yielding a positive correlation between them. Hence, the different F_{out} response following ENSO could be the reason for the different EPFD response. Figure 7 further shows the wave-mean flow interaction for different ENSO states in both MIROC6 and ERA5, which helps give more dynamical insight. In El Niño years, waves propagate upward into the stratosphere in the midlatitudes (around 60°N) in both model and observations. However, in MIROC6, waves then propagate equatorward, leading to high-latitude EP flux divergence and zonal wind acceleration (Figure 7a); whereas in ERA5, waves mostly propagate further upward, leading to a consistent convergence of EP flux and zonal wind deceleration throughout the stratospheric polar region (Figure 7d). The situation in La Niña years is essentially opposite to that in El Niño years (Figures 7b and e and 7c, f). Hence, although ENSO leads to a similar wave forcing into the stratosphere in the model and in the observations, the outflow from high to low latitudes is quite different, leading to the opposite-signed EPFD response and thus the opposite SPV response in the model compared to observations.

Together, this provides the full picture of how ENSO influences the SPV in the MIROC6 model (Figure 8). Taking an El Niño year as an example, the enhanced wave forcing ([v′T′]) contributes to an EP flux convergence (EPFD) in the polar stratosphere and thus to SPV weakening, and vice-versa in La Niña years. The sign of this direct eddy heat flux pathway is reflected as a positive correlation between ENSO and [v′T′], multiplied by a negative correlation between [v′T′] and EPFD, and then by a positive correlation between EPFD and SPV, leading overall to a negative correlation between ENSO and SPV. This pathway is consistent with observations and theory, and the model captures the known pathway successfully. However, in addition to the anomalous wave forcing into the stratosphere, the outflow of the planetary waves within the stratosphere shows a strong positive correlation with ENSO. In El Niño years, there tends to be strong outflow of the waves, leading to enhanced EPFD. This positive correlation between ENSO and EPFD dominates over the negative correlation via the eddy heat flux, leading to the positive ENSO-SPV correlation in MIROC6 (Figures 1 and 3). This result is consistent with the MLR analysis as discussed in the previous section, confirming that the causal linkage between [v′T′] and SPV variability is strongly state dependent on ENSO in MIROC6, with the wave propagation within the stratosphere being an important factor.

5. Observational Evidence for the Identified Physical Reason

So far, we have identified the physical reason behind the model-observations discrepancy. The state dependence of the [v′T′]-SPV relationship in the MIROC6 model is what leads to a ENSO-SPV correlation opposite in sign to the observations (Figures 3 and 4), and is mainly attributed to the different wave propagation within the stratosphere (Figures 5–7). While this linkage is physically plausible, an essential question is whether the observational record provides evidence for or against the existence of such a state dependence in the real world. This will be the evidential basis for us to decide whether the model can be used to study aspects of climate variability related to the ENSO-SPV teleconnection.

At first glance, the observations show some consistency with the modeled distribution in terms of this state dependence (Figures 4 and 6), but they are based on a very small number of ENSO years and thus can be very

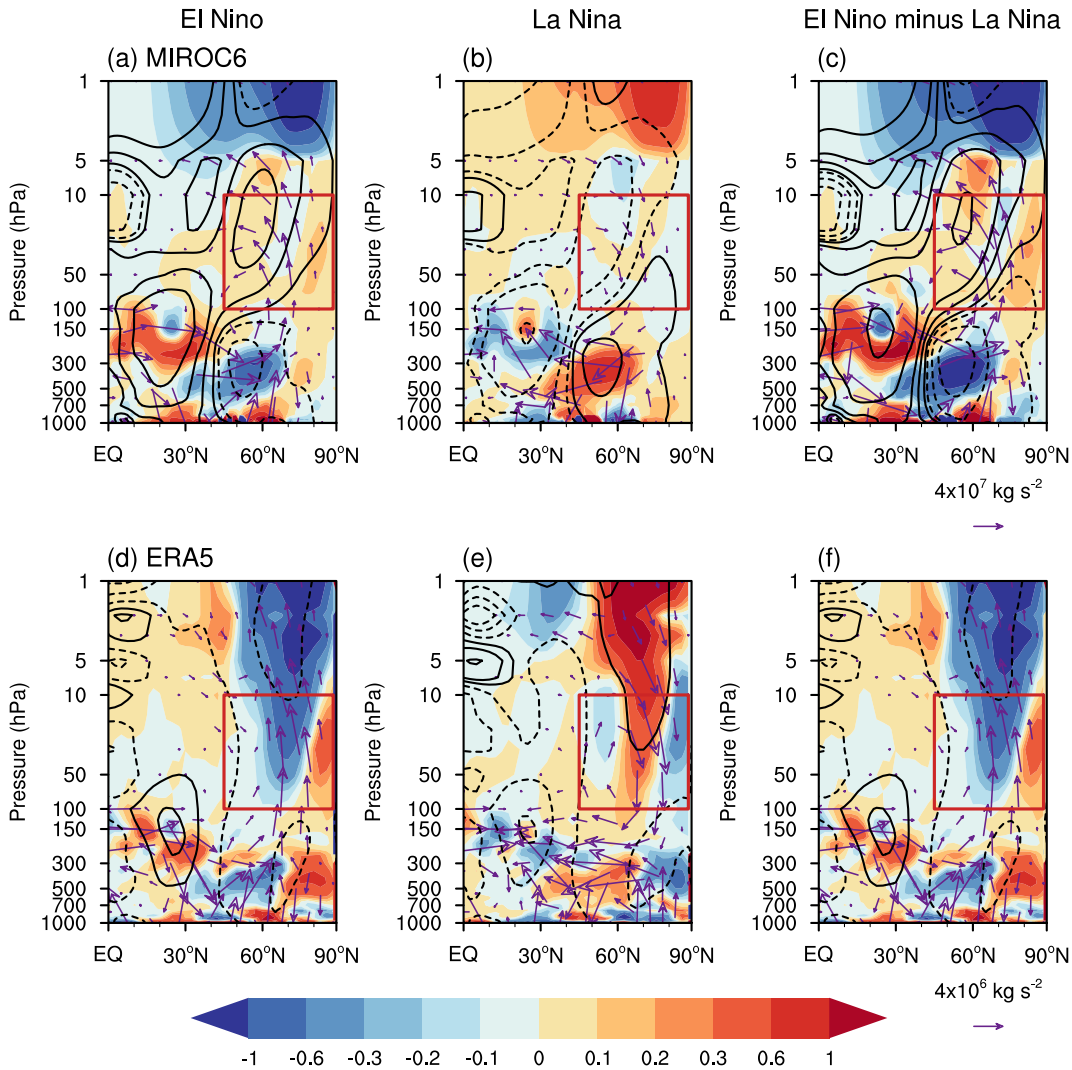

Figure 7. Wave-mean flow interaction in different ENSO states. (a) and (b) Composite field in El Niño and La Niña years in MIROC6. (c) Difference between (a) and (b). (d)–(f) same as (a)–(c), but for ERA5. EP flux anomaly is shown as a vector, with the shading denotes EP flux divergence anomaly. Contour shows the zonal mean zonal wind anomaly. Here the selection thresholds of El Niño and La Niña years are 0.5 and −0.5 standard deviations, respectively. For visualization purposes, the vertical component of EP flux was multiplied by 200.

uncertain. To quantify the strength of the observational evidence in the face of such uncertainty, we use the Bayes Factor (BF, Kass & Raftery, 1995; Kretschmer et al., 2020). The general idea is that we quantify how likely it is to see the data under two opposing hypotheses, with the BF being the ratio of these probabilities. Thus, the BF represents the relative strength of the evidence supporting the two hypotheses. It is preferable to the p-value in a Null Hypothesis Significance Test, which does not specify an alternative hypothesis. Note that the BF does not represent which hypothesis is more likely, but rather with which hypothesis the data is more consistent. In order to invert the conditional in that way, one would need to know the prior probabilities of the two hypotheses (Shepherd, 2021). Without strong priors, a BF that differs from unity by a very substantial margin is needed for the evidence to be considered influential either one way or the other.

First, we check if the observations support the existence of this additional pathway, represented by the non-zero α term in Equation 1 (Figure 3b). We consider two hypotheses: the alternative hypothesis H1 is "in the real world, the α term is the same as the modeled median value", and the null hypothesis H0 is "in the real world, the α term is zero". The detailed forms of these hypotheses are shown in Table 1. We then quantify whether the ERA5 data gives any reason to conclude that the observed relationship is more likely to have occurred under H1 than under

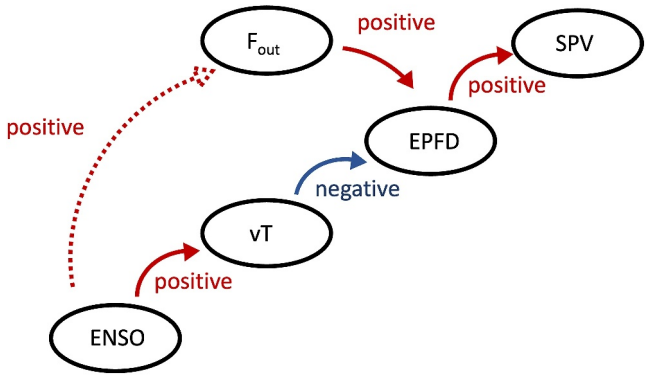

Figure 8. Schematic of how ENSO influences SPV in MIROC6. The solid arrows show correlations that are consistent with observations and theory. The dashed arrow shows a correlation seen only in MIROC6. Red and blue arrows indicate positive and negative correlations, respectively.

H0. A BF greater than one indicates that the relationship found in observations is more likely to happen under the alternative hypothesis H1, and a BF less than one indicates that it is more likely to happen under the null hypothesis H0. A BF between 1/3 and 3 is usually considered as neutral evidence, and only if the BF shows a value greater than 20, or smaller than 1/20, is the evidence in favor of H1 or of H0, respectively, considered to be strong (Kass & Raftery, 1995). For this hypothesis set, the BF is 1.00, which means that the observations are equally likely to have occurred under the two hypotheses. As a further check we applied the leave-one-out method to compute the BF, to which the results were insensitive (Table 1).Therefore, from the observational evidence, we cannot rule out the possibility that this is an additional pathway for ENSO to influence SPV in the real world.

From our earlier analysis, the existence of an additional pathway from ENSO to SPV in MIROC6 can be interpreted as a state-dependence of the [v′T′]-SPV relationship in response to ENSO phase, reflected by the *y*-intercept shift in Figure 4. Hence, we next quantify the extent to which the observations provide evidence for or against the hypothesis that the *y*-intercept changes depending on ENSO phase. Here we consider two hypotheses: the alternative hypothesis H3 is "in the real world, the *y*-intercept changes toward the values in MIROC6 during El Niño and La Niña years", and the null hypothesis H2 is "in the real world, the *y*-intercept remains zero during El Niño and La Niña years", with the detailed forms shown in Table 1. For El Niño years, the BF is 0.76, and for La Niña years, the BF is 0.77, meaning that the data is slightly more likely under H2. But these BFs are still rather neutral, neither providing evidence in favor of or against the hypothesis H3 that the *y*-intercept tends to shift as in the model.

The neutral BFs mean that from a statistical point of view, while we do not have strong evidence to support the hypothesis that the dependence of the [v′T′]-SPV relationship on the state of ENSO as found in MIROC6 is also present in the real world, neither can we rule out this possibility. Since in the observations there is a very limited sample size of ENSO events, the inconsistency between model and observations could be the result of sampling uncertainty (Figure 4h). However, it is just as likely that the state dependence represents a model bias.

6. Making Use of the Model Simulations: A Physically-Based Bias-Adjustment and Its Implications

While a clear discrepancy between MIROC6 and ERA5 exists in terms of the ENSO-SPV statistical relationship, we found that the dynamical mechanisms behind the relationship are actually well captured by the model and that the discrepancy might reflect a state dependence on ENSO of the relationship between [v′T′] and SPV involving the influence of the outflow of planetary waves from the high latitude stratosphere. One possible explanation for the different outflow of planetary waves is that the background waveguide differs from that in observations (Tatebe et al., 2019), which could directly lead to a bias in the wave propagation. However, in this scenario, we would expect a systematic difference in the horizontal wave propagation, unlike the opposite response in El Niño and La Niña years seen in Figures 7a–7c. Potential factors that could lead to a different waveguide in El Niño and La Niña years could be other confounding factors, including the Quasi-Biennial Oscillation (QBO) and the preconditioning of the polar stratosphere (Holton & Tan, 1980; McIntyre, 1982). It is noteworthy that MIROC6

Table 1
Bayes Factors for the Proposed Hypothesis Set

Hypothesis set		Null hypothesis	Alternative hypothesis	Bayes factor (BF)
Existence of the additional pathway		H0: $SPV = \beta \cdot [v'T'] + residual$	H1: $SPV = \alpha \cdot ENSO + \beta \cdot [v'T'] + residual$	1.00 (0.99~1.01)
Y-intercept changes depending on ENSO phase	El Niño	H2: $SPV = \gamma \cdot [v'T'] + residual$	H3: $SPV = \gamma \cdot [v'T'] + \delta + residual$	0.76 (0.45~1.08)
	La Niña			0.77 (0.59~0.95)

Note. The left column shows the aim of the hypothesis set. The equation of each hypothesis is presented with the detailed descriptions provided in the text. The BF is the ratio of the probability of the data occurring under the alternative hypothesis against the probability of the data occurring under the null hypothesis. The range shows the minimum and maximum values of the BF based on the leave-one-out method.

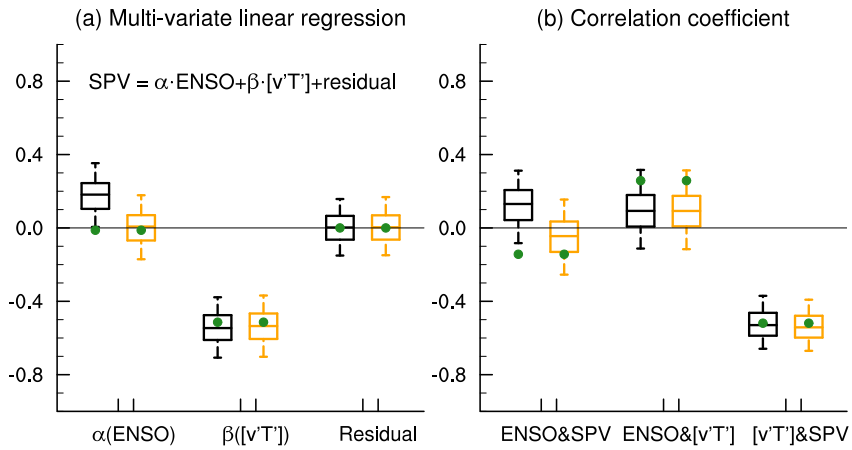

Figure 9. Core relationships before and after the bias adjustment. (a) Multi-variate linear regression model; the indices have been standardized. (b) Correlation coefficients between different variables. Black box is the result before the bias adjustment, while the orange box shows the results after bias adjustment. The green dots indicate the observational values, with the same values shown for the original and bias-adjusted model data. The distributions are obtained from the bootstrapped data with the same sample size as the observations.

can reproduce QBO well, and the relationship between QBO and SPV is similar to that in observations (Rao et al., 2020; Richter et al., 2020). After excluding the linear influence of those two factors on the outflow of planetary waves, the strong positive ENSO-F_{out} relationship still remains (not shown). However, this does not mean that those two factors do not play a role in modulating the wave propagation following ENSO, instead it suggests that the drivers may be complex and involve strong nonlinearities (Ma et al., 2023; Walsh et al., 2022).

As further analysis suggests that the observed evidence cannot rule out the possibility of this state dependence (Table 1), the discrepancy may not be a true model bias arising from deficient dynamical processes; it could instead be a result of the limited sample size of the observations. However, it also worth mentioning that the model produces a weaker tropospheric wave forcing response and a stronger influence of outflow on the EP flux divergence compared with observations (Figures 3b and 6a), which could contribute to the eventual dominance of the F_{out} pathway (Figure 8). This now puts us in a dilemma. Certainly, we cannot be confident enough to use the MIROC6 simulations to investigate scientific questions associated with the ENSO-SPV relationship directly. However, as MIROC6 is one of the largest ensembles to date and one of the few large ensembles which represents the QBO well (Richter et al., 2020; Shiogama et al., 2023), we believe that a better utilization of the model simulations is needed, and that how one uses it should depend on the purpose of the question at hand.

If we believe the state dependence is genuine and is undersampled in the observations, then MIROC6 simulations can be used directly to provide more information about it, and help to better understand how ENSO influences SPV and related aspects of climate variability. On the other hand, if we believe the model is biased and we require an ENSO-SPV relationship that is consistent with observations as the basis for further analysis, then we need to assume that the state dependence of the [v′T′]-SPV relationship seen in MIROC6 is spurious. In this case, the model is bias adjustable in the sense that the relevant physical processes are well captured (Maraun et al., 2017), and bias adjustment methods can be applied to remove the influence of the additional pathway and to retain only the stable ENSO-SPV relationship.

Here we conduct a simple bias adjustment as an example to illustrate how we can use the model output under such an assumption. This approach is based on the MLR in Equation 1, in which we remove the influence of ENSO not mediated by [v′T′], quantified through the parameter α; therefore, the bias-adjusted SPV becomes $SPV\text{-}\alpha\cdot ENSO$. Figure 9a shows the regression coefficients in this MLR before and after the bias adjustment. After the bias adjustment the median of the parameter α is about zero, which is very close to the observational value, implying that we have successfully removed the influence on SPV via other processes. Figure 9b further shows the distribution of the correlation coefficients between SPV and other variables. The overall relationship between ENSO and SPV presents a negative shift compared to the original correlation. The median becomes negative, and the observational value now falls within the IQR. This further underlines that this physically-based bias adjustment is

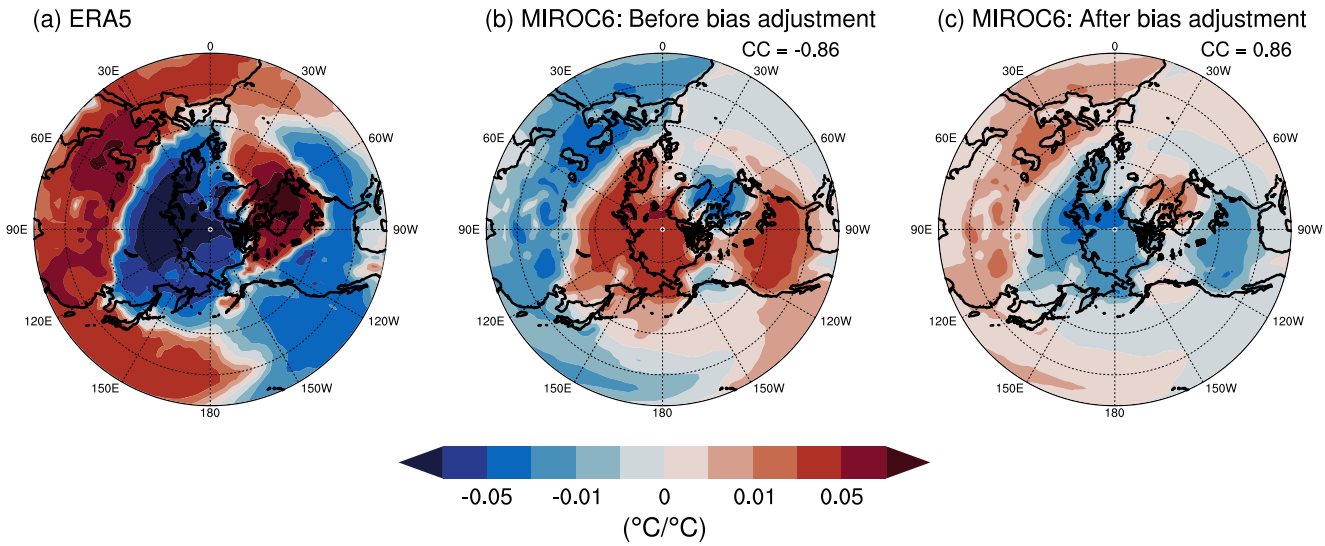

Figure 10. Late winter surface air temperature response to ENSO via the SPV pathway. (a) Regression coefficient in ERA5. (b) and (c) Regression coefficient in MIROC6 before and after bias adjustment, respectively. The pattern correlation coefficient with ERA5 is shown on the upper right. Note that the color bar is nonlinear for visualization purposes.

useful to reproduce the observed ENSO-SPV teleconnection. Meanwhile, this approach does not change the distribution of parameter β and the residual term (Figure 9a), nor the correlation coefficient between [v′T′] and SPV (Figure 9b), implying that the bias adjustment is independent of the well-captured wave-mean flow interaction mechanism.

As the ENSO-SPV relationship is a key teleconnection which not only influences the circulation variability but also has strong impacts on weather and climate, we further test the potential application of this bias adjustment on late winter surface air temperature (SAT). The stratospheric pathway provides the main linkage between ENSO and European climate (Ineson & Scaife, 2009; Jiménez-Esteve & Domeisen, 2019), and is what makes ENSO a useful predictor of European wintertime climate. The combined influence of SPV and ENSO on SAT can be written as:

$$\mathrm{SAT} = \alpha \cdot \mathrm{ENSO} + \beta \cdot \mathrm{SPV} + \mathrm{residual}, \tag{3}$$

where α is the influence of ENSO apart from the stratospheric pathway, and β is the direct influence of SPV on SAT. Hence, the influence of ENSO on SAT via the stratospheric pathway should be computed as the product of the regression coefficient of SPV with respect to ENSO, in the standardized case simply given by the correlation between ENSO and SPV, and SPV's direct influence on SAT denoted by β in Equation 3 (Kretschmer et al., 2021). Figure 10a shows this influence of ENSO via the stratospheric pathway on the late winter SAT, averaged between February and March, in ERA5. While a cold anomaly is seen in the polar regions, the northern part of Eurasia, and the southeastern part of North America, strong warming is seen in midlatitudes over Eurasia. However, before the bias adjustment, MIROC6 shows an opposite SAT response with a −0.86 spatial correlation to the ERA5 response (Figure 10b). This is not to our surprise as the original ENSO-SPV teleconnection is opposite in sign. In fact, this opposite SAT response confirms that the representation of the ENSO-SPV teleconnection has profound impacts, highlighting the necessity to understand it and to decide whether to perform a bias adjustment, based on the scientific question. Under the assumption that the [v′T′]-SPV relationship is independent of ENSO, and after the bias adjustment of the model, the SAT response shows an in-phase structure to the observations (Figure 10c), with the spatial correlation becoming 0.86. In other words, the bias adjustment successfully captures the sign of the observed ENSO influence on SAT, providing the basis for more applications related to the ENSO-SPV teleconnection.

7. Summary

In this work, starting from the opposite ENSO-SPV relationship found in MIROC6 compared to the observations, we explored the physical mechanisms behind the relationship to see if the model is truly deficient or if it captures

the main processes but is biased in a physically understandable way. We found that while the ENSO-SPV correlation is opposite in sign, the model actually produces the correct ENSO-eddy heat flux relationship and the correct eddy heat flux-SPV relationship. This means that the relevant dynamics are well captured by the model, as the wave forcing following ENSO events and the associated wave-mean flow interaction are both similar to what are seen in the observations. While it may seem to be a paradox that the consistent processes eventually lead to the opposite SPV response, we further found that in the model, additional to the influence of wave forcing changes, there is a strong positive relationship between ENSO and SPV. This is reflected in an intensity change of the SPV in ENSO years even without the influence of anomalous wave forcing (Figures 4a–4c), which strongly shapes the causal linkage between ENSO and SPV.

The observed SPV changes in ENSO years apart from wave forcing are similar to what we saw in MIROC6, providing some indication of this state dependence of the [v′T′]-SPV relationship (Figures 4d–4f). From a purely statistical perspective, it is impossible to say whether this difference between model and observations reflects a true physical difference, or undersampling within the observations. The evidence distinguishing between those two hypotheses, quantified in the Bayes factor, is essentially neutral, given the limited sample size of the observations. Hence, we cannot rule out the possibility that the ENSO-SPV relationship in the real world has other linkages with the same sign as in the model, for example, modulated by other confounding factors, a hypothesis for which we can find some support in other studies (Garfinkel et al., 2019; Rao et al., 2019; Xie et al., 2018).

A closer inspection suggested that this opposite SPV response is related to the opposite EP flux divergence. Although the upward-propagating planetary waves from the troposphere are well captured, the wave propagation within the stratosphere is very different between model and observations. This different wave propagation is what leads to the opposite EP flux divergence, and therefore to the opposite SPV response. Therefore, even though the model produces an opposite SPV response to ENSO to that seen in observations, the response is physically reasonable. Depending on the particular scientific question related to the ENSO-SPV relationship that is of interest, different choices can be made to make appropriate use of the MIROC6 large ensemble simulations. Under the assumption that the state dependence is spurious, a physically-based bias adjustment can be applied by only retaining the influence on the SPV via the eddy heat flux. This turns out to produce an ENSO-SPV relationship consistent with observations. Moreover, the bias adjustment will affect any statistical relationship that is influenced by the ENSO-SPV relationship. We applied the bias adjustment to investigate the influence of ENSO on late winter surface air temperature over Northern Hemisphere, and it changed the sign of the SAT response to be consistent with that found in observations.

8. Discussion

While tremendous efforts have been devoted to improve the performance of models, and there are more and more model simulations being produced, there is still an open question to know when a particular model can be safely used for a particular purpose (Deser et al., 2017; Jain et al., 2023). Here we advocate for the use of a forensic approach to understand the discrepancies between climate models and observations (Figure 11). The first step is a physically-based analysis, by which we test if the relevant dynamical processes are well captured and what it is that is leading to the model-observations discrepancy, as in Section 3. If the model-observations discrepancy is physically reasonable (meaning consistent with known physical laws) and the observations do not provide strong evidence against the reason for the discrepancy, then the model can be used to explore aspects and consequences of the teleconnection (as in Sections 4 and 5); otherwise, the model is not appropriate for studying this teleconnection. According to the specific purposes of the research, the model can be used in different ways (as in Section 6). If the study does not require an ENSO-SPV relationship consistent with observations, for example, if it is used to study non-stationarity and state dependence, then the model can be used directly to provide conditional information. On the other hand, if the study requires an ENSO-SPV relationship consistent with observations, then the model should be used only after the application of a physically-based bias adjustment. By considering various explanations for a phenomenon and attempting to disprove them through physical evidence, such a forensic investigation not only gives us a more physically-based picture of model-observations discrepancy, but also provides guidance for a reliable and effective utilization of model simulations. This is particularly important for understanding teleconnections, as they will ultimately affect climate model performance (Wang et al., 2014).

Similar physically-based approaches have been applied to explore the information from climate models. For example, the storyline approach has been widely used to study future projections of regional climate, which can be

Figure 11. A physically-based forensic investigation process as a guide to understand the model-observations discrepancy in teleconnections and to effectively utilize the model outputs. The direction of the arrows indicates the order in which actions are taken. The bubbles enclosed by the black contours indicate the conclusions about whether we can trust the model and how to use the model outputs.

helpful to make sense of the uncertainty and enrich our understanding (Shepherd, 2019). A most evident advantage of a physically-based approach is that it combines the dynamics and the statistics, thus providing a more comprehensive understanding with a clear physical picture (Kretschmer et al., 2020; Shepherd, 2021). Here we show that those approaches also have high potential to help understand model-observations discrepancies, thus helping to provide a clear guide on how to use the model output.

For teleconnections, there is a great need to use large ensemble model datasets to better understand their features, variability, and related aspects. The ENSO-SPV relationship shown in this study is a striking example. The analyses performed here mainly focus on the linear relationships between variables and do not take into account potential nonlinearities. While this is reasonable in the context of teleconnections which are usually studied via linear relationships, for the ENSO-SPV case, its nonlinearity is a topical scientific question (e.g., Hardiman et al., 2019; Jiménez-Esteve & Domeisen, 2019; Rao & Ren, 2016). In the MIROC6 model, there are indeed some clues for possible nonlinearity (Figure 1c). As apparent nonlinearities in the real world remain controversial since they may be a manifestation of internal variability (Deser et al., 2017), large model ensembles could help to better address this question. In addition, the observed ENSO-SPV relationship is not stable across different time periods,

exhibiting a nonstationarity (e.g., Ayarzagüena et al., 2019; Butler & Polvani, 2011; Domeisen et al., 2019). If we look at the ENSO-SPV relationship in MIROC6 across different time periods (Fig. S2 in Supporting Information S1), we see clear differences as well. One other important phenomenon related to the ENSO-SPV relationship is that it behaves differently in early and late winter (Ayarzagüena et al., 2018; Manzini et al., 2006; O'Reilly et al., 2024), which is also seen in this model (Fig. S3 in Supporting Information S1). Similarly, other aspects such as the asymmetry between El Niño and La Niña (e.g., Rao & Ren, 2016) and the relative contribution of different pathways for ENSO to influence the extratropics (e.g., Jiménez-Esteve & Domeisen, 2020), can also be studied. Hence, large ensembles can be used to obtain more information, which can be very helpful to deepen the understanding of the ENSO-SPV relationship. This also applies to other teleconnections, further highlighting the need to understand the large ensembles and use them effectively.

Apart from the teleconnection itself, this physically-based understanding is valuable for further applications. The ENSO-SPV teleconnection, as an example, serves as an important modulator of other relationships. In recent decades, Arctic sea ice has been experiencing a dramatic decrease (Serreze et al., 2007), and is suggested to have strong impacts on midlatitude weather and climate (Cohen et al., 2020; Overland et al., 2016). Among all the pathways, the stratospheric pathway has been a hot topic within the academic community (Kim et al., 2014; Kretschmer et al., 2016, 2020; Zhang et al., 2018). As ENSO can influence the SPV directly, the ENSO state has been suggested to strongly influence the SPV response to the sea ice change (Ma et al., 2022). Therefore, studying the ENSO-SPV linkage is essential not only for analyzing the direct influence of ENSO on the worldwide climate, but also for understanding the variability of the climate system. As the bias adjustment successfully reproduced the observed ENSO-SPV relationship, it provides the potential for using this large ensemble to investigate the influence of ENSO on the SPV response to sea ice change, and the resulting influences on midlatitudes. This example further highlights that a physically-based forensic approach is helpful for understanding both the teleconnections and the variability of the global circulation.

Data Availability Statement

The data used in the present study are openly accessible. MIROC6 data are available from https://esgf-index1.ceda.ac.uk. HadISST data are available from https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html. ERA5 data are available from https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5.

Acknowledgments
This work was supported by the UK NERC Project ArctiCONNECT (NE/V005855/1) and the project XAIDA, funded through the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003469. We thank the Associate Editor as well as Daniela Domeisen and the other two reviewers for their helpful suggestions and comments. We would also like to thank the MIROC modelling group. X. S. would like to thank James Screen at University of Exeter and Blanca Ayarzagüena at Universidad Complutense de Madrid for helpful discussions.

References

Andrews, D. G., Holton, J. R., & Leovy, C. B. (1987). *Middle atmosphere dynamics*. Academic Press Inc.

Ayarzagüena, B., Ineson, S., Dunstone, N. J., Baldwin, M. P., & Scaife, A. A. (2018). Intraseasonal effects of El Niño–Southern oscillation on North Atlantic climate. *Journal of Climate*, *31*(21), 8861–8873. https://doi.org/10.1175/JCLI-D-18-0097.1

Ayarzagüena, B., López-Parages, J., Iza, M., Calvo, N., & Rodríguez-Fonseca, B. (2019). Stratospheric role in interdecadal changes of El Niño impacts over Europe. *Climate Dynamics*, *52*(1–2), 1173–1186. https://doi.org/10.1007/s00382-018-4186-3

Baldwin, M. P., & Thompson, D. W. J. (2009). A critical comparison of stratosphere-troposphere coupling indices. *Quarterly Journal of the Royal Meteorological Society*, *135*(644), 1661–1672. https://doi.org/10.1002/qj.479

Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, *115*(6), 1083–1126. https://doi.org/10.1175/1520-0493(1987)115<1083:Csapol>2.0.Co;2

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, *97*(3), 163–172. https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2

Brönnimann, S. (2007). Impact of El Nino southern oscillation on European climate. *Reviews of Geophysics*, *45*(2), RG3003. https://doi.org/10.1029/2006rg000199

Butler, A. H., & Polvani, L. M. (2011). El Nino, La Nina, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophysical Research Letters*, *38*(13), L13807. https://doi.org/10.1029/2011gl048084

Charlton, A. J., & Polvani, L. M. (2007). A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *Journal of Climate*, *20*(3), 449–469. https://doi.org/10.1175/jcli3996.1

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., et al. (2020). Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nature Climate Change*, *10*(1), 20–29. https://doi.org/10.1038/s41558-019-0662-y

Deser, C., Simpson, I. R., McKinnon, K. A., & Phillips, A. S. (2017). The Northern Hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *Journal of Climate*, *30*(13), 5059–5082. https://doi.org/10.1175/jcli-d-16-0844.1

Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W. A., & Baehr, J. (2015). Seasonal predictability over Europe arising from El Niño and stratospheric variability in the MPI-ESM seasonal prediction system. *Journal of Climate*, *28*(1), 256–271. https://doi.org/10.1175/JCLI-D-14-00207.1

Domeisen, D. I. V., Garfinkel, C. I., & Butler, A. H. (2019). The teleconnection of El Nino southern oscillation to the stratosphere. *Reviews of Geophysics*, *57*(1), 5–47. https://doi.org/10.1029/2018rg000596

Dong, L., & McPhaden, M. J. (2017). Why has the relationship between Indian and Pacific Ocean decadal variability changed in recent decades? *Journal of Climate*, *30*(6), 1971–1983. https://doi.org/10.1175/jcli-d-16-0313.1

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., et al. (2016). Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nature Geoscience*, *9*(11), 809–814. https://doi.org/10.1038/ngeo2824

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman and Hall, 456.

Garfinkel, C. I., & Hartmann, D. L. (2007). Effects of the El Nino-Southern oscillation and the Quasi-biennial oscillation on polar temperatures in the stratosphere. *Journal of Geophysical Research*, *112*(D19), D19112. https://doi.org/10.1029/2007jd008481

Garfinkel, C. I., Schwartz, C., Butler, A. H., Domeisen, D. I. V., Son, S. W., & White, I. P. (2019). Weakening of the teleconnection from El Nino-Southern Oscillation to the Arctic stratosphere over the past few decades: What can be learned from subseasonal forecast models? *Journal of Geophysical Research-Atmospheres*, *124*(14), 7683–7696. https://doi.org/10.1029/2018jd029961

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Ineson, S., Lim, J., & Fereday, D. (2019). The impact of strong El Niño and La Niña events on the North Atlantic. *Geophysical Research Letters*, *46*(5), 2874–2883. https://doi.org/10.1029/2018GL081776

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holton, J. R., & Tan, H. C. (1980). The influence of the equatorial Quasi-Biennial Oscillation on the global circulation at 50mb. *Journal of the Atmospheric Sciences*, *37*(10), 2200–2208. https://doi.org/10.1175/1520-0469(1980)037<2200:Tioteq>2.0.Co;2

Hoskins, B., & Woollings, T. (2015). Persistent extratropical regimes and climate extremes. *Current Climate Change Reports*, *1*(3), 115–124. https://doi.org/10.1007/s40641-015-0020-8

Ineson, S., Dunstone, N. J., Scaife, A. A., Andrews, M. B., Lockwood, J. F., & Pang, B. (2023). Statistics of sudden stratospheric warmings using a large model ensemble. *Atmospheric Science Letters*, *25*(3). https://doi.org/10.1002/asl.1202

Ineson, S., & Scaife, A. A. (2009). The role of the stratosphere in the European climate response to El Nino. *Nature Geoscience*, *2*(1), 32–36. https://doi.org/10.1038/ngeo381

Jain, S., Scaife, A. A., Shepherd, T. G., Deser, C., Dunstone, N., Schmidt, G. A., et al. (2023). Importance of internal variability for climate model assessment. *npj Climate and Atmospheric Science*, *6*(1), 68. https://doi.org/10.1038/s41612-023-00389-0

Jiménez-Esteve, B., & Domeisen, D. I. (2020). Nonlinearity in the tropospheric pathway of ENSO to the North Atlantic. *Weather and Climate Dynamics*, *1*(1), 225–245. https://doi.org/10.5194/wcd-1-225-2020

Jiménez-Esteve, B., & Domeisen, D. I. V. (2019). Nonlinearity in the North Pacific atmospheric response to a linear ENSO forcing. *Geophysical Research Letters*, *46*(4), 2271–2281. https://doi.org/10.1029/2018GL081226

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., et al. (2014). Weakening of the stratospheric polar vortex by Arctic sea-ice loss. *Nature Communications*, *5*(1), 4646. https://doi.org/10.1038/ncomms5646

Kosaka, Y., Nakamura, H., Watanabe, M., & Kimoto, M. (2009). Analysis on the dynamics of a wave-like teleconnection pattern along the summertime Asian jet based on a reanalysis dataset and climate model simulations. *Journal of the Meteorological Society of Japan*, *87*(3), 561–580. https://doi.org/10.2151/jmsj.87.561

Kretschmer, M., Adams, S. V., Arribas, A., Prudden, R., Robinson, N., Saggioro, E., & Shepherd, T. G. (2021). Quantifying causal pathways of teleconnections. *Bulletin of the American Meteorological Society*, *102*(12), E2247–E2263. https://doi.org/10.1175/bams-d-20-0117.1

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *Journal of Climate*, *29*(11), 4069–4081. https://doi.org/10.1175/JCLI-D-15-0654.1

Kretschmer, M., Zappa, G., & Shepherd, T. G. (2020). The role of Barents-Kara sea ice loss in projected polar vortex changes. *Weather Climate Dynamics*, *1*(2), 715–730. https://doi.org/10.5194/wcd-1-715-2020

Kumar, V., Hitchman, M. H., Du, W., Dhaka, S. K., & Yoden, S. (2024). Teleconnection of the Quasi-biennial oscillation with boreal winter surface climate in Eurasia and North America. *Communications Earth & Environment*, *5*(1), 251. https://doi.org/10.1038/s43247-024-01422-7

Kushner, P. J., & Polvani, L. M. (2004). Stratosphere-troposphere coupling in a relatively simple AGCM: The role of eddies. *Journal of Climate*, *17*(3), 629–639. https://doi.org/10.1175/1520-0442(2004)017<0629:Sciars>2.0.Co;2

Ma, T. J., Chen, W., An, X. D., Garfinkel, C. I., & Cai, Q. Y. (2023). Nonlinear effects of the stratospheric Quasi-Biennial Oscillation and ENSO on the North Atlantic winter atmospheric circulation. *Journal of Geophysical Research-Atmospheres*, *128*(17), e2023JD039537. https://doi.org/10.1029/2023jd039537

Ma, X., Wang, L., Smith, D., Hermanson, L., Eade, R., Dunstone, N., et al. (2022). ENSO and QBO modulation of the relationship between Arctic sea ice loss and Eurasian winter climate. *Environmental Research Letters*, *17*(12), 124016. https://doi.org/10.1088/1748-9326/aca4e9

Manzini, E., Ayarzagüena, B., Calvo, N., & Matei, D. (2024). Nonlinearity and asymmetry of the ENSO stratospheric pathway to North Atlantic and Europe, Revisited. *Journal of Geophysical Research: Atmospheres*, *129*(2), e2023JD039992. https://doi.org/10.1029/2023JD039992

Manzini, E., Giorgetta, M. A., Esch, M., Kornblueh, L., & Roeckner, E. (2006). The influence of sea surface temperatures on the northern winter stratosphere: Ensemble simulations with the MAECHAM5 model. *Journal of Climate*, *19*(16), 3863–3881. https://doi.org/10.1175/jcli3826.1

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., et al. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*(11), 764–773. https://doi.org/10.1038/nclimate3418

Matsuno, T. (1971). Dynamical model of stratospheric sudden warming. *Journal of the Atmospheric Sciences*, *28*(8), 1479–1494. https://doi.org/10.1175/1520-0469(1971)028<1479:Admots>2.0.Co;2

McIntyre, M. E. (1982). How well do we understand the dynamics of stratospheric warmings. *Journal of the Meteorological Society of Japan*, *60*(1), 37–65. https://doi.org/10.2151/jmsj1965.60.1_37

Monnin, E., Kretschmer, M., & Polichtchouk, I. (2022). The role of the timing of sudden stratospheric warmings for precipitation and temperature anomalies in Europe. *International Journal of Climatology*, *42*(6), 3448–3462. https://doi.org/10.1002/joc.7426

Mudhar, R., Seviour, W. J. M., Screen, J. A., Geen, R., Lewis, N. T., & Thomson, S. I. (2024). Exploring mechanisms for model-dependency of the stratospheric response to Arctic warming. *Journal of Geophysical Research: Atmospheres*, *129*(10), e2023JD040416. https://doi.org/10.1029/2023JD040416

Notz, D. (2015). How well must climate models agree with observations? *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *373*(2052), 20140164. https://doi.org/10.1098/rsta.2014.0164

O'Reilly, C. H., Drouard, M., Ayarzagüena, B., Ambaum, M. H. P., & Methven, J. (2024). The role of storm-track dynamics in the intraseasonal variability of the winter ENSO teleconnection to the North Atlantic. *Quarterly Journal of the Royal Meteorological Society*, *150*(761), 2069–2086. https://doi.org/10.1002/qj.4691

Overland, J. E., Dethloff, K., Francis, J. A., Hall, R. J., Hanna, E., Kim, S.-J., et al. (2016). Nonlinear response of mid-latitude weather to the changing Arctic. *Nature Climate Change*, *6*(11), 992–999. https://doi.org/10.1038/nclimate3121

Polvani, L. M., & Waugh, D. W. (2004). Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of Climate*, *17*(18), 3548–3554. https://doi.org/10.1175/1520-0442(2004)017<3548:Uwafaa>2.0.Co;2

Qasmi, S., Cassou, C., & Boé, J. (2017). Teleconnection between Atlantic Multidecadal variability and European temperature: Diversity and evaluation of the Coupled model Intercomparison project phase 5 models. *Geophysical Research Letters*, *44*(21), 11140–11149. https://doi.org/10.1002/2017gl074886

Rao, J., Garfinkel, C. I., & Ren, R. C. (2019). Modulation of the northern winter stratospheric El Nino-Southern oscillation teleconnection by the PDO. *Journal of Climate*, *32*(18), 5761–5783. https://doi.org/10.1175/jcli-d-19-0087.1

Rao, J., Garfinkel, C. I., & White, I. P. (2020). Impact of the Quasi-Biennial Oscillation on the Northern winter stratospheric polar vortex in CMIP5/6 models. *Journal of Climate*, *33*(11), 4787–4813. https://doi.org/10.1175/jcli-d-19-0663.1

Rao, J., & Ren, R. (2016). Asymmetry and nonlinearity of the influence of ENSO on the northern winter stratosphere: 1. Observations. *Journal of Geophysical Research-Atmospheres*, *121*(15), 9000–9016. https://doi.org/10.1002/2015JD024520

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020). Progress in simulating the Quasi-biennial oscillation in CMIP models. *Journal of Geophysical Research-Atmospheres*, *125*(8), e2019JD032362. https://doi.org/10.1029/2019jd032362

Runge, J., Petoukhov, V., & Kurths, J. (2014). Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *Journal of Climate*, *27*(2), 720–739. https://doi.org/10.1175/jcli-d-13-00159.1

Serreze, M. C., Holland, M. M., & Stroeve, J. (2007). Perspectives on the Arctic's shrinking sea-ice cover. *Science*, *315*(5818), 1533–1536. https://doi.org/10.1126/science.1139426

Shepherd, T. G. (2019). Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences*, *475*(2225), 20190013. https://doi.org/10.1098/rspa.2019.0013

Shepherd, T. G. (2021). Bringing physical reasoning into statistical practice in climate-change science. *Climatic Change*, *169*(1–2), 2. https://doi.org/10.1007/s10584-021-03226-6

Shiogama, H., Tatebe, H., Hayashi, M., Abe, M., Arai, M., Koyama, H., et al. (2023). MIROC6 large ensemble (MIROC6-LE): Experimental design and initial analyses. *Earth System Dynamics*, *14*(6), 1107–1124. https://doi.org/10.5194/esd-14-1107-2023

Sigmond, M., & Shepherd, T. G. (2014). Compensation between resolved wave driving and parameterized orographic gravity wave driving of the Brewer-Dobson circulation and its response to climate change. *Journal of Climate*, *27*(14), 5601–5610. https://doi.org/10.1175/jcli-d-13-00644.1

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research-Atmospheres*, *118*(4), 1716–1733. https://doi.org/10.1002/jgrd.50203

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Trenberth, K. E. (1997). The definition of El Nino. *Bulletin of the American Meteorological Society*, *78*(12), 2771–2777. https://doi.org/10.1175/1520-0477(1997)078<2771:Tdoeno>2.0.Co;2

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N. C., & Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research*, *103*(C7), 14291–14324. https://doi.org/10.1029/97jc01444

Van Loon, H., & Labitzke, K. (1987). The Southern Oscillation. 5. Th e anomalies in the lower stratosphere of the Northern-Hemisphere in winter and a comparision with the Quasi-Biennial Oscillation. *Monthly Weather Review*, *115*(2), 357–369. https://doi.org/10.1175/1520-0493(1987)115<0357:Tsopvt>2.0.Co;2

Van Loon, H., & Madden, R. A. (1981). The Southern Oscillation. 1. Global associations with pressure and temperature in Northern winter. *Monthly Weather Review*, *109*(6), 1150–1162. https://doi.org/10.1175/1520-0493(1981)109<1150:Tsopig>2.0.Co;2

Wallace, J. M. (2000). North Atlantic Oscillation/annular mode: Two paradigms-one phenomenon. *Quarterly Journal of the Royal Meteorological Society*, *126*(564), 791–805. https://doi.org/10.1256/smsqj.56401

Wallace, J. M., & Gutzler, D. S. (1981). Teleconections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review*, *109*(4), 784–812. https://doi.org/10.1175/1520-0493(1981)109<0784:Titghf>2.0.Co;2

Walsh, A., Screen, J. A., Scaife, A. A., & Smith, D. M. (2022). Non-Linear response of the extratropics to tropical climate variability. *Geophysical Research Letters*, *49*(23). https://doi.org/10.1029/2022gl100416

Wang, C. Z., Zhang, L. P., Lee, S. K., Wu, L. X., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. https://doi.org/10.1038/nclimate2118

Wang, T., Tian, W. S., Lin, Y. F., Gou, X. H., Liu, H. W., Wang, X. J., & Xie, F. (2023). Decadal changes in the relationship between Arctic stratospheric ozone and sea surface temperatures in the North Pacific. *Atmospheric Research*, *292*, 106870. https://doi.org/10.1016/j.atmosres.2023.106870

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, *63*(11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:Scoteo>2.0.Co;2

Xie, F., Zhou, X., Li, J. P., Sun, C., Feng, J., & Ma, X. (2018). The key role of background sea surface temperature over the cold tongue in asymmetric responses of the Arctic stratosphere to El Nino-Southern Oscillation. *Environmental Research Letters*, *13*(11), 114007. https://doi.org/10.1088/1748-9326/aae79b

Yang, P., Bao, M., Ren, X., & Tan, X. (2023). The role of vortex preconditioning in influencing the occurrence of strong and Weak sudden stratospheric warmings in ERA5 reanalysis. *Journal of Climate*, *36*(4), 1197–1212. https://doi.org/10.1175/JCLI-D-22-0319.1

Yessimbet, K., Shepherd, T. G., Ossó, A. C., & Steiner, A. K. (2022). Pathways of influence between northern hemisphere blocking and stratospheric polar vortex variability. *Geophysical Research Letters*, *49*(23), e2022GL100895. https://doi.org/10.1029/2022GL100895

Zhang, P., Wu, Y., Simpson, I. R., Smith, K. L., Zhang, X., De, B., & Callaghan, P. (2018). A stratospheric pathway linking a colder Siberia to Barents-Kara Sea sea ice loss. *Science Advances*, *4*(7), eaat6025. https://doi.org/10.1126/sciadv.aat6025