Earth and Space Science

RESEARCH ARTICLE
10.1029/2021EA001676

Improved Circulation in the Northern Hemisphere by Adjusting Gravity Wave Drag Parameterizations in Seasonal Experiments With ICON-NWP

Raphael Köhler[1,2], Dörthe Handorf[1], Ralf Jaiser[1], Klaus Dethloff[1,2], Günther Zängl[3], Detlev Majewski[3], and Markus Rex[1,2]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany, [2]University of Potsdam, Institute of Physics and Astronomy, Potsdam, Germany, [3]Deutscher Wetterdienst, Offenbach am Main, Germany

Key Points:

- Seasonal simulations with ICON-NWP underestimate stratospheric vortex strength in winter
- Adjusted gravity wave drag schemes lead to reduced biases in Northern Hemisphere circulation
- Stratosphere-troposphere coupling is intensified with a reduced gravity wave drag

Correspondence to:
R. Köhler,
Raphael.Koehler@awi.de

Citation:
Köhler, R., Handorf, D., Jaiser, R., Dethloff, K., Zängl, G., Majewski, D., & Rex, M. (2021). Improved circulation in the Northern Hemisphere by adjusting gravity wave drag parameterizations in seasonal experiments with ICON-NWP. *Earth and Space Science*, *8*, e2021EA001676. https://doi.org/10.1029/2021EA001676

Received 22 JAN 2021
Accepted 5 FEB 2021

Abstract The stratosphere is one of the main potential sources for subseasonal to seasonal predictability in midlatitudes in winter. The ability of an atmospheric model to realistically simulate the stratospheric dynamics is essential in order to move forward in the field of seasonal predictions in midlatitudes. Earlier studies with the ICOsahedral Nonhydrostatic atmospheric model (ICON) point out that stratospheric westerlies in ICON are underestimated. This is the first extensive study on the evaluation of Northern Hemisphere stratospheric winter circulation with ICON in numerical weather prediction (NWP) mode. Seasonal experiments with the default setup are able to reproduce the basic climatology of the stratospheric polar vortex. However, westerlies are too weak and major stratospheric warmings too frequent in ICON. Both a reduction of the nonorographic, and a reduction of the orographic gravity wave and wake drag lead to a strengthening of the stratospheric vortex and a bias reduction, in particular in January. However, the effect of the nonorographic gravity wave drag scheme on the stratosphere is stronger. Stratosphere-troposphere coupling is intensified and more realistic due to a reduced gravity wave drag. Furthermore, an adjustment of the subgrid-scale orographic drag parameterization leads to a significant error reduction in the mean sea level pressure. As a result of these findings, we present our current suggested improved setup for seasonal experiments with ICON-NWP.

Plain Language Summary Although seasonal forecasts for midlatitudes have the potential to be highly beneficial to the public sector, they are still characterized by a large amount of uncertainty. Exact simulations of the circulation in the stratosphere can help to improve tropospheric predictability on seasonal time scales. For this reason, we investigate how well the new German atmospheric model is able to simulate the stratospheric circulation. The model reproduces the basic behavior of the Northern Hemisphere stratospheric polar vortex, but the westerly circulation in winter is underestimated. The stratospheric circulation is influenced by gravity waves that exert drag on the flow. These processes are only partly physically represented in the model, but are very important and are hence parameterized. By adjusting the parameterizations for the gravity wave drag, the stratospheric polar vortex is strengthened, thereby yielding a more realistic stratospheric circulation. In addition, the altered parameterizations improve the simulated surface pressure pattern. Based upon this, we present our current suggested improved model setup for seasonal experiments.

1. Introduction

The ICOsahedral Non-hydrostatic atmospheric model (ICON) has been developed jointly by the German weather service (DWD) and the Max Planck Institute for Meteorology (MPI-M), and is the central piece of the new unified modeling approach in Germany (Bonaventura, 2004; Zängl et al., 2015). ICON currently exists in two main configurations: one for numerical weather predictions by DWD (hereafter ICON-NWP), which has been operational since 2015, and one for climate simulations by MPI-M (ICON-A). Both configurations share the same dynamical core, but differ in their physical packages. The dynamical core has been tested (Zängl et al., 2015), the ICON-NWP forecasts are constantly verified by DWD and ICON-A has been described (Giorgetta et al., 2018) and evaluated (Crueger et al., 2018), but there is no extensive study on the evaluation of Northern Hemisphere (NH) stratospheric winter circulation in ICON.

Due to the fact that the stratospheric polar vortex is highly variable in winter, models often struggle to capture its variability and strength (i.e., Butchart et al., 2011; Charlton-Perez et al., 2013; Seviour et al., 2016). Baldwin and Dunkerton (1999, 2001) showed that Arctic Oscillation anomalies can propagate downward from the stratosphere into the troposphere. Thus, due to this downward coupling, a misrepresentation of stratospheric dynamics in a model may lead to errors in the troposphere. Furthermore, recent studies have suggested that improved stratospheric processes in the models contribute to a more realistic representation of tropospheric circulation patterns (i.e., Overland et al., 2016; Romanowsky et al., 2019; Sigmond et al., 2013). While downward propagation of stratospheric signals only plays a minor role for numerical weather prediction due to the short time scales, the stratosphere is one of the main potential sources for subseasonal to seasonal predictability in midlatitudes (Butler et al., 2019). Based upon this, a realistic representation of stratospheric dynamics is essential and needs to be tested, in order to improve predictability on medium-range and seasonal time scales. Therefore, in the first part of this study, we analyze a seasonal climatology of ICON-NWP simulations focusing on the stratospheric dynamics and variability (cf. Section 3.1).

The strength and variability of the stratospheric polar vortex are determined by the nonlinear interactions of the vortex with the total amount of horizontal momentum transported vertically, predominantly from the troposphere. This momentum is transported to the stratosphere by large-scale Rossby waves and mesoscale gravity waves. Borchert et al. (2019) show that stratospheric westerlies in winter are too weak in ICON, despite using an upper atmosphere extension of ICON. They suggest a retuning of the gravity wave drag parameterizations in ICON. The drag exerted on the flow by gravity waves can be categorized depending on the origin of waves: The drag related to waves generated by flow over orography is described as orographic drag (Nastrom & Fritts, 1992), whereas the drag related to waves originating from all kinds of dynamical motions such as fronts, convection or jet streams, is named nonorographic gravity wave drag (Fritts & Nastrom, 1992). First, parameterizations for the orographic drag were included into atmospheric models to reduce the systematic westerly biases in the Northern Hemisphere wintertime flow (Palmer et al., 1986), and more recently, the nonorographic gravity wave drag was originally implemented to further improve stratospheric processes in the tropics, Southern Hemisphere and the mesosphere (Dunkerton, 1997; Orr et al., 2010). In ICON, the orographic drag is parameterized as part of the subgrid scale orographic (SSO) drag scheme, which also includes a low-level drag component (wake drag). The nonorographic drag in ICON is parameterized separately. Imbalances resulting from atmospheric flow structures can force air parcels to oscillate, and thereby radiate nonorographic gravity waves. As these source mechanisms have a wide spatio-temporal range and are often poorly constrained, the nonorographic gravity wave drag is especially connected to a large range of uncertainty (Fritts & Alexander, 2003). In addition, the effect of the nonorographic gravity wave drag parameterizations on the stratosphere-troposphere coupling is still fairly unexplored. A first study by Polichtchouk, Shepherd, and Byrne (2018) showed that a reduced nonorographic gravity wave drag leads to intensified coupling between stratosphere and troposphere and vice versa. Furthermore, Choi et al. (2018) demonstrate that enhanced nonorographic gravity wave drag parameterizations can improve the seasonal representations of the stratospheric winds. The resolution dependency of the orographic and nonorographic gravity wave drag parameterizations further add to the uncertainty of these parameterizations (i.e., Vosper et al., 2020). How the different reduced gravity wave drag parameterizations affect the stratospheric circulation and the coupling to the troposphere in ICON will be discussed in the second part of this study (cf. Section 3.2). In the third part of this study, we will investigate the surface pressure patterns in the troposphere and how they are impacted by changes to the parameterizations (cf. Section 3.3).

2. Model and Experimental Setup

This study evaluates the atmospheric general circulation model ICON-NWP version 2.1.0 as distributed by DWD in the 2018 ICON tutorial. In our configuration, the model is set up with the horizontal resolution R2B5, which corresponds to a grid mesh of approximately 80 km, and 90 vertical levels up to a height of 75 km. The SSO drag scheme in ICON is based on Lott and Miller (1997), whereas the nonorographic gravity wave drag is based on Warner and McIntyre (1996) and Scinocca (2003), and is described in Orr et al. (2010). All experiments are initialized with ERA-Interim reanalysis data (Dee et al., 2011) on the first of September of the respective year and run for 9 months, thereby including boreal autumn, winter, and spring. Based

Table 1
Summary of the Parameters Used for the Different ICOsahedral Non-Hydrostatic Atmospheric Model Experiments in This Study

Experiment	tune_gfluxlaun	tune_gkwake	tune_gkdrag	tune_gfrcrit
$ICON_{ctl}$	2.5 mPa	1.5	0.075	0.4
$ICON_{nogwd-}$	2.0 mPa	1.5	0.075	0.4
$ICON_{sso-}$	2.5 mPa	1.2	0.060	0.5
$ICON_{gwd-}$	2.0 mPa	1.2	0.060	0.5

The parameter tune_gfluxlaun describes the total launch momentum flux on each azimuth and is based on Warner and McIntyre (1996) and Scinocca (2003), and described in Orr et al. (2010). The parameters tune_gkwake, tune_gkdrag, and tune_gfrcrit describe the low level wake drag constant, the gravity wave drag constant, and the critical Froude number, respectively and are based on the SSO scheme by Lott and Miller (1997). SSO, subgrid scale orographic; ICON, ICOsahedral Non-hydrostatic atmospheric model.

on the Atmosphere Model Intercomparison Project (AMIP) protocol, we prescribe mid-monthly sea surface temperatures and sea ice concentrations produced by the Program for Climate Model Diagnosis and Intercomparison for the AMIP experiments of CMIP6 (Taylor et al., 2000). The main results of this study are not sensitive to the choice of boundary data, as sensitivity experiments with ERA-Interim sea surface temperatures and sea ice concentrations show the same features. Concerning the volume mixing ratios of CO_2, CH_4, N_2O, CFC-11 and CFC-12, we used the historical greenhouse gas concentrations for CMIP6 (Meinshausen et al., 2017), complemented by NOAA/ESRL Global Monitoring Division global and monthly mean data for 2015, 2016, and 2017.

All ICON experiments are generated by simulating periods from September to May for 1979/80–2016/17 (38 years), with each experiment consisting of five ensemble members. Consequently, this leads to a total of 190 single model runs per experiment. The ensemble members were generated by shifting the initialization by ±6 h and ±12 h. The ICON experiments only differ in their gravity wave drag related parameters. Whereas $ICON_{ctl}$ uses the default settings for the parameterizations of the SSO scheme and the nonorographic gravity wave drag, these parameterizations were adjusted in the sensitivity experiments $ICON_{nogwd-}$, $ICON_{sso-}$, and $ICON_{gwd-}$ with the goal of strengthening the stratospheric polar vortex in winter. The nonorographic gravity wave drag is reduced in $ICON_{nogwd-}$ by decreasing the parameter for the total launch momentum flux in each azimuth (tune_gfluxlaun) by 20% from 0.0025 to 0.0020 Pa. Based on the experience of the DWD and small earlier sensitivity experiments, in $ICON_{sso-}$ the strength of the SSO scheme is reduced by decreasing the low level wake drag constant (tune_gkwake) and the gravity wave drag constant (tune_gkdrag) by 20% from 1.5 to 1.2 and from 0.075 to 0.060, respectively. Furthermore, the critical Froude number (tune_gfrcrit) is increased from 0.4 to 0.5. It is the inverse of the nondimensional height, so that a large critical Froude number will increase the frequency of flow going over the mountain and gravity waves are forced by the vertical motion of the fluid, whereas a small critical Froude number will lead to more frequent low level flow blocking by orography. Hence, an increase of the critical Froude number will lead to an increase in gravity wave excitation, which is partly compensated for by the reduced gravity wave drag constant. $ICON_{gwd-}$ combines parameterizations from $ICON_{nogwd-}$ and $ICON_{sso-}$. The parameters for all four experiments are listed in Table 1.

The results from the ICON simulations are evaluated against an ERA-Interim reanalysis climatology, consisting of the months September–May for the years 1979/80–2016/17. Major stratospheric warmings (MSWs) in simulations and reanalysis are identified by a reversal of winds from westerlies to easterlies at 10 hPa and 60°N, and a simultaneous reversal of the zonal-mean temperature gradient between 60°N and 90°N, in order that T(60°N) < T(90°N) (Labitzke, 1981). Two major warmings are separated by at least 20 days of westerlies and, in order to rule out final warmings, events that are not succeeded by at least 10 successive days of westerlies are not taken into account. To enable comparison between the ICON ensembles and the reanalysis, MSWs are given in MSWs per decade, thereby the number of MSWs is computed for each ensemble member separately and then averaged over the five members. The statistical significance in this study is assessed with a two-sided Wilcoxon-test (Bauer, 1972; Hollander et al., 2013). Reference calculations with the Student's t-test showed very comparable results. However, the Wilcoxon-test is non-parametric and therefore is independent of the distribution of the analyzed data. Based on Wilks (2016), *p*-values small enough to satisfy the false discovery rate (FDR) criterion with a test level of $\alpha_{FDR} = 0.1$ are described as statistically significant.

3. Results

3.1. Variability and Strength of the Stratospheric Polar Vortex in ICON-NWP

For the evaluation of variability and strength of the stratospheric polar vortex in ICON-NWP, we analyze the ICON experiment with default gravity wave drag settings ($ICON_{ctl}$) and compare it to the ERA-Interim

climatology. The effect of reduced gravity wave drag on the atmospheric circulation will be discussed in Section 3.2.

The NH wintertime stratospheric polar vortex is centered around 60°N. Therefore, the zonal-mean zonal wind at 10 hPa at this latitude is a good measure of the polar vortex strength in the mid-stratosphere. This is depicted in Figure 1 for ERA-Interim (a) and the four ICON-experiments (b)–(e). The evolution of the wintertime polar vortex can be separated into three phases. The first is the formation of the vortex in autumn with intensifying westerlies and weak variability. In the second phase, there is a peak in vortex strength in early winter which is accompanied by intensified vertical propagation of planetary waves and subsequent wave breaking, leading to a more disturbed and variable polar vortex (Charney & Drazin, 1961; Matsuno, 1971; Polvani & Waugh, 2004). Depending on the intensity of vertical wave propagation, January can be the month with peak zonal-mean zonal winds, but also can be characterized by vortex breakdowns and reversals. The third phase in spring is characterized by declining vortex strength and variability, and, following the final warming, there is a transition to weak summer easterlies accompanied by stratospheric high pressure in summer. ICON is able to reproduce the climatological characteristics of the different vortex phases, additionally the standard deviation and extreme values are in a realistic range. Nevertheless, there are some distinct differences: The mean stratospheric polar vortex strength in $ICON_{ctl}$ is underestimated in winter. This is especially true in January, where the monthly mean zonal-mean zonal wind at 10 hPa and 60°N is considerably weaker than in ERA-Interim. Furthermore, the strength of the polar vortex in October and early November is slightly underestimated, whereas the vortex strength is simulated well in spring. Moreover, extreme values in $ICON_{ctl}$ have a larger range, in particular in autumn. This slightly larger range of extreme values could be connected to the larger amount of realizations, i.e. ensemble members.

The polar cap mean (PCM) zonal wind bias of $ICON_{ctl}$ relative to ERA-Interim is presented in Figure 2a. The PCM is defined as the field mean from 65.25°N to 90°N and is strongly influenced by the strength and position of the polar vortex. The mean stratosphere of $ICON_{ctl}$ is dominated by a negative wind bias in the PCM, and especially in January this is large and highly significant. This negative bias of zonal winds is connected to a weak polar vortex, high stratospheric temperatures and a high frequency of MSWs (cf. Figure 3e). The strong January bias is preceded by a highly significant and downward propagating negative zonal wind bias in October and November. Contrary to the dominant negative bias in the midstratosphere, the upper stratosphere is dominated by significant positive wind bias in November, early December, February, and March. The positive wind bias in November and early December can be explained by the weakened westerlies in the ERA-Interim climatology (cf. green contours in Figure 2a). This feature is caused by an accumulation of minor vortex weakening events in the end of November, an attribute that is not reproduced by the model.

In accordance with the underestimated strength of the stratospheric polar vortex, the frequency of MSWs is overestimated by $ICON_{ctl}$ in all relevant months. However once again, January stands out with on average four MSWs per decade compared to 1.58 for the reanalysis (cf. Figure 3e). Stratospheric warming events are driven by the breaking of planetary waves propagating up from the troposphere. High vortex breakdown frequencies can be an indicator of an overestimated upward propagation of planetary waves and smaller scale gravity waves. In addition, MSWs are the strongest manifestation of the stratosphere-troposphere coupling. Different studies have shown that stratospheric warmings precede increased geopotential in the Arctic and negative anomalies in the midlatitudes, a signal associated with the negative phase of the Arctic oscillation (i.e., Charlton & Polvani, 2007; Cohen & Jones, 2011; Sigmond et al., 2013). How strong this connection is in ICON and how the surface pressure systems are influenced by changes in the stratosphere, will be investigated in Section 3.3.

3.2. Sensitivity of the Polar Vortex to the Gravity Wave Drag Related Parameterizations

An earlier study has described that the stratospheric polar vortex in ICON-A shows similar features to those described in the previous section: Borchert et al. (2019) state that stratospheric winter westerlies are too weak in ICON and suggest a retuning of orographic and nonorographic gravity wave parameters. In this section, we will discuss the adjustment of the mentioned parameters in ICON-NWP.

Polichtchouk, Shepherd, and Hogan (2018) and Polichtchouk, Shepherd, and Byrne (2018) suggest that the nonorographic gravity wave drag is a tuneable parameter for obtaining a more realistic

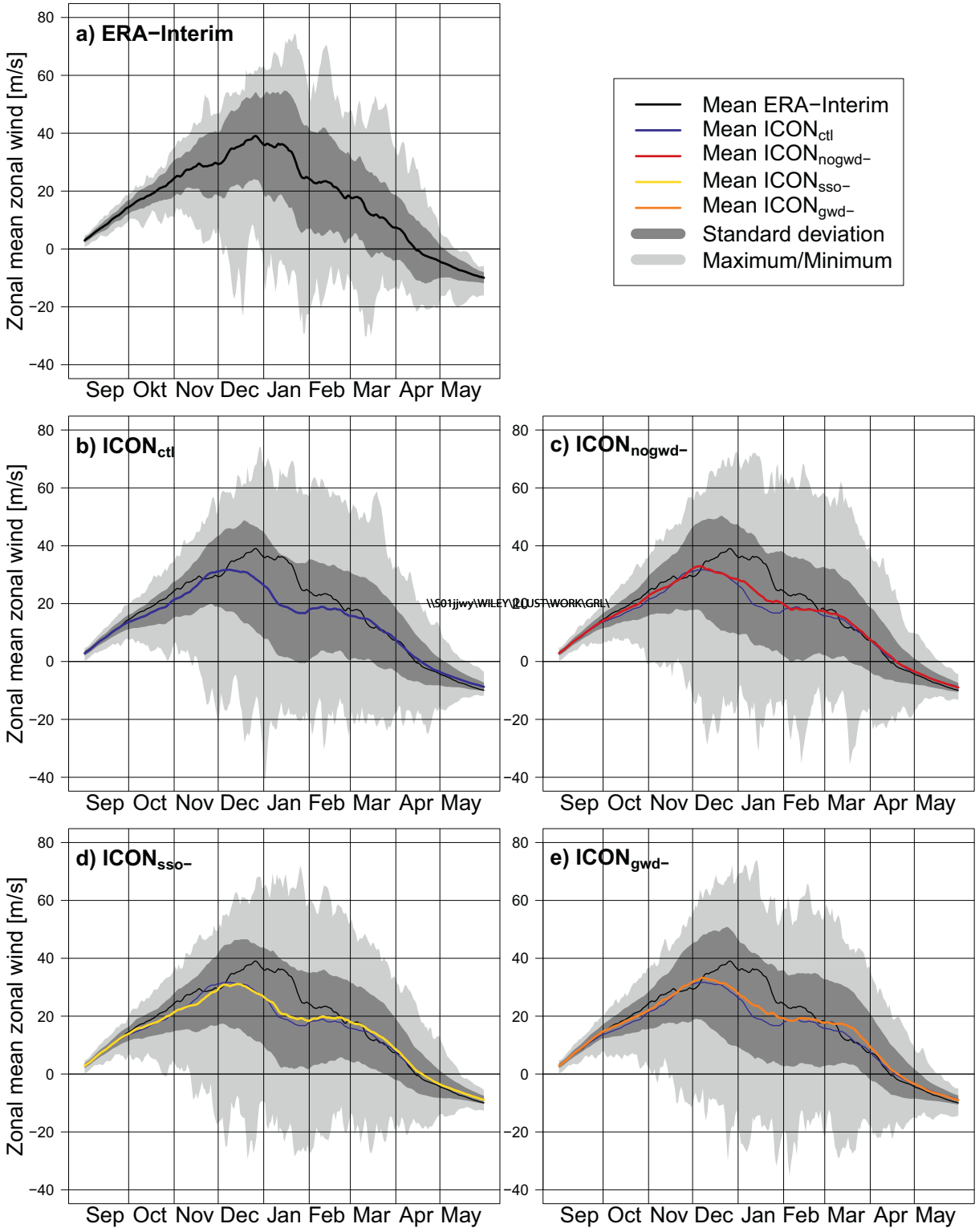

Figure 1. Zonal-mean zonal wind at 10 hPa and 60°N (lines), standard deviation (dark gray shading) and extrema (light gray shading) for 1979/80–2016/17 period of ERA-Interim (black), ICON$_{ctl}$ (blue), ICON$_{nogwd-}$ (red), ICON$_{sso-}$ (yellow), and ICON$_{gwd-}$ (orange). Positive (negative) values indicate westerly (easterly) circulation. ICON, ICOsahedral nonhydrostatic atmospheric model.

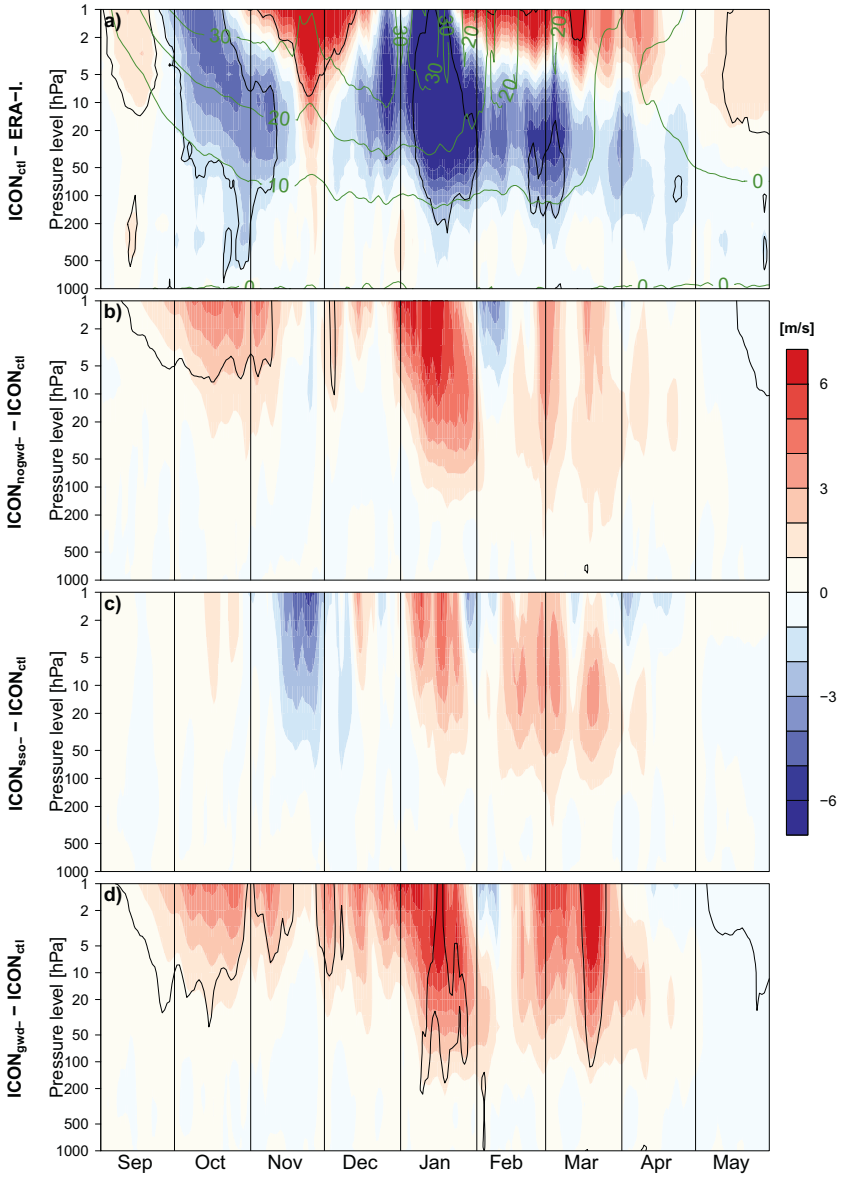

Figure 2. Time-height cross section of climatological mean zonal wind differences averaged over 65.25°N–90°N. $ICON_{ctl}$ biases relative to ERA-Interim (a), $ICON_{nogwd-}$ (b), $ICON_{sso-}$ (c), and $ICON_{gwd-}$ (d) differences to $ICON_{ctl}$. Black lines indicate statistical significance with p-values small enough to satisfy the FDR criterion with $\alpha_{FDR} = 0.1$. The p-values are based on a two-sided Wilcoxon-test. Climatological mean values of ERA-Interim data are given by the green contours in (a). FDR, false discovery rate; ICON, ICOsahedral nonhydrostatic atmospheric model.

stratosphere-troposphere coupling behavior in models. Moreover, a reduced nonorographic gravity wave drag leads to a reduction in the MSW frequency. Taking into account the results from Section 3.1, we tested these hypotheses with the goal of strengthening the wintertime stratospheric vortex in ICON-NWP. Therefore, the nonorographic gravity wave drag was reduced by 20% in $ICON_{nogwd-}$. The parameters controlling the strength of the orographic gravity wave drag and orographic wake drag are tuneable parameters as part of the SSO scheme. With the goal of strengthening the stratospheric polar vortex, we also reduced the orographic gravity wave drag constant by 20% in a further experiment named $ICON_{sso-}$. Initial seasonal sensitivity experiments in which only the orographic gravity wave drag constant (tune_gkdrag) was reduced, exhibited only small effects on the stratospheric circulation, and furthermore indicated a mean sea level pressure which can be described as "highs too low, lows too high." Williams et al. (2020) found the polar high pressure bias in the Met Office Unified Model to be strongly related to the orographic drag

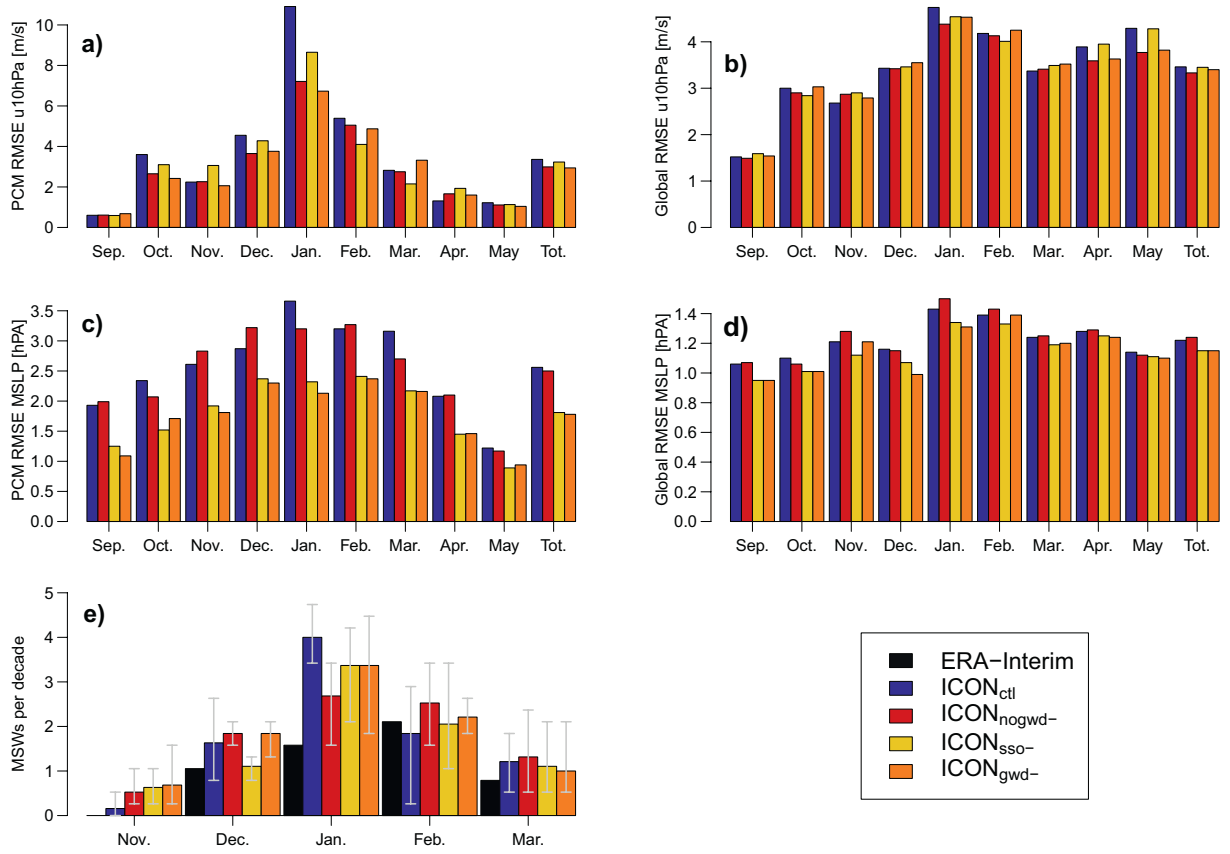

Figure 3. PCM (65.25°N–90°N) and global mean RMSE of zonal wind at 10 hPa (a), (b) and mean sea level pressure (c), (d) and frequency of MSWs per decade (e). Total in the last column denotes the 9-month average RMSE and the total frequency of MSWs per decade. The MSW frequency ensemble range of the different ICON experiments is given by the gray error bars. ICON, ICOsahedral nonhydrostatic atmospheric model; MSWs, major stratospheric warmings; PCM, polar cap mean; RMSE, root-mean-square error.

parameterizations. As a result, parameters controlling the low level wake drag and the partitioning between gravity wave drag and wake drag (cf. Table 1) were also adjusted in $ICON_{sso-}$ based on the operational experience of the DWD. The sensitivity experiment $ICON_{gwd-}$ then combines the adjustments made to both the orographic and nonorographic gravity wave drag schemes. All sensitivity experiments are compared to $ICON_{ctl}$ as well as the ERA-Interim reanalysis.

Figures 1b–1e demonstrate the effects of the adjusted gravity wave drag parameterizations on the strength of the stratospheric polar vortex. Compared to $ICON_{ctl}$, the sensitivity experiment $ICON_{nogwd-}$ with reduced nonorographic gravity wave drag exhibits a strengthened polar vortex, in particular in January. The impact of reduced SSO forcing ($ICON_{sso-}$) is less pronounced in the stratosphere, but also leads to a moderate strengthening of the polar vortex in January. The experiment with the combined adjusted parameterizations $ICON_{gwd-}$ strongly resembles $ICON_{nogwd-}$, suggesting a stronger impact of the nonorographic gravity wave drag parameterization on the stratosphere. Nevertheless, all experiments demonstrate that a reduction of gravity wave drag leads to a strengthening of the stratospheric polar vortex in winter. Changes in the drag parameterizations have only marginal effects on the variability and extrema of the stratospheric polar vortex. Since the drag reduction leads to a vortex strengthening in winter, the sensitivity experiments match the ERA-Interim reanalysis better than $ICON_{ctl}$.

The direct effect of the altered gravity wave drag parameterizations on the PCM zonal wind is shown in Figures 2b–2d by exhibiting differences between the sensitivity experiments and the control run. Reduced gravity wave drag leads to an intensification of the westerly circumpolar circulation. This effect is strongest in January in all three sensitivity experiments. However, not all signals are statistically significant. Significant changes in $ICON_{nogwd-}$ are confined to the upper stratosphere in autumn, where a strengthening of

the westerly circulation leads to a bias reduction, in particular in October. Due to the high variability of the polar vortex in winter, the large differences in January do not satisfy the FDR criterion. The stratospheric effects of reduced SSO forcing is less pronounced and not statistically significant. An intensification of PCM westerlies is visible in the mid and upper stratosphere from January to March. The combination of both adjusted gravity wave drag parameterizations has the strongest effect on the polar circulation. Westerlies are significantly strengthened in the upper stratosphere from September to December, and in the whole stratosphere in January and March. Thereby, biases are significantly reduced, especially in January. Nevertheless, the intensification of the zonal wind has negative effects on the bias of the upper stratosphere in November, December, and March. Having said this, in total the bias reductions in the mid and lower stratosphere outweigh the bias increase in the upper stratosphere. The positive effect of reduced gravity wave drag on the stratospheric circulation of the high latitudes can be quantified by calculating the PCM root-mean-square error (RMSE) of monthly mean zonal wind data at 10 hPa for the different experiments (cf. Figure 3a). Compared to $ICON_{ctl}$ the PCM error averaged over the whole simulation period is reduced by 18% in $ICON_{nogwd-}$, 11% in $ICON_{sso-}$, and 19% $ICON_{gwd-}$. In January this effect is strongest, and errors are reduced by 34%, 21%, and 38% respectively. Largest effects of the adjusted parameterizations are visible in winter in the high latitudes of the NH. The stratospheric zonal winds in the Tropics and the Southern Hemisphere are not largely influenced. Hence, the effects on the global mean RMSE are small (cf. Figure 3b). In autumn, the sensitivity experiments exhibit a slightly larger global mean RMSE. This increase is caused by an increase of error in the high latitudes of the Southern Hemisphere. In winter, the global signal is dominated by the improvements in the high latitudes in the Northern Hemisphere and the RMSE reductions in spring are mainly based on improvements in lower latitude due to reduced nonorographic drag. The global mean RMSE averaged over the whole simulation period is reduced by 3.8% in $ICON_{nogwd-}$, 0.3% in $ICON_{sso-}$, and 1.7% $ICON_{gwd-}$.

This improvement in the model circulation is manifested in Figure 4, which illustrates the zonal mean RMSE of 10 hPa zonal wind (a–f) and MSLP (m–r) as a function of latitude and the 50°N–70°N spatial RMSE of zonal wind as a function of altitude (g–i). Whereas the MSLP errors will be discussed in the following section, in this part we focus on the latitudinal and height distribution of zonal wind errors of the different experiments. The months September, April, and May are not shown, due to the fact that the errors, along with the impact of the adjusted gravity wave drag parameterizations are small in these months. Largest impacts of the adjusted parameterizations on the zonal mean zonal wind at 10 hPa are visible in latitudes north of 50°N. In October, December, and February all gravity wave drag reductions lead to small RMSE reductions. The 10 hPa errors in November and March are small, and there is no RMSE reduction in the sensitivity experiments. Once again, January stands out: It is the month with the largest errors in $ICON_{ctl}$, but also the month with the largest error reductions in the sensitivity experiments. The strong RMSE reduction in the sensitivity experiments is caused by a strengthening of the stratospheric polar vortex as a result of reduced gravity wave drag. The improvement in the NH stratosphere due to the nonorographic gravity wave drag is larger than by the SSO scheme, which includes the orographic drag. This error reduction is clearly present in all vertical levels of the stratosphere and also in the upper troposphere (cf. Figure 4j). It seems to originate in the upper stratosphere in December and propagates downward in the course of winter, thereby peaking in the mid-stratosphere in January. The upper stratospheric increase of the RMSE in the sensitivity experiments in March does not affect the layers below, due to the fact that the polar vortex is in the declining phase transitioning to the stable summer regime.

In contrast with the results from Polichtchouk, Shepherd, and Hogan (2018), the frequency of MSWs in ICON-NWP does not seem to be affected by changes in the gravity wave drag parameterizations (cf. Figure 3e). Having said this, the modification of the nonorographic gravity wave drag used by Polichtchouk, Shepherd, and Hogan (2018) is a lot larger than in our study. Nevertheless, the nonorographic gravity wave drag in particular, seems to have an impact on the timing of MSWs. Whereas most MSWs occur in February, followed by January, December, and March in ERA-Interim, $ICON_{ctl}$ and $ICON_{sso-}$ exhibit a strong peak in January. In particular in $ICON_{nogwd-}$ the distribution is more realistic, as less MSWs occur in January. On the other hand, MSWs become more frequent in November, an event which has not been observed in the ERA-Interim data. However, keeping in mind, that the statistic of ERA-Interim is based solely on 21 MSWs and therefore needs to be treated with caution. Thus, it makes sense to compare the reanalysis to the ICON ensemble spread. Independent of the experiment, most ensemble members overestimate the MSW

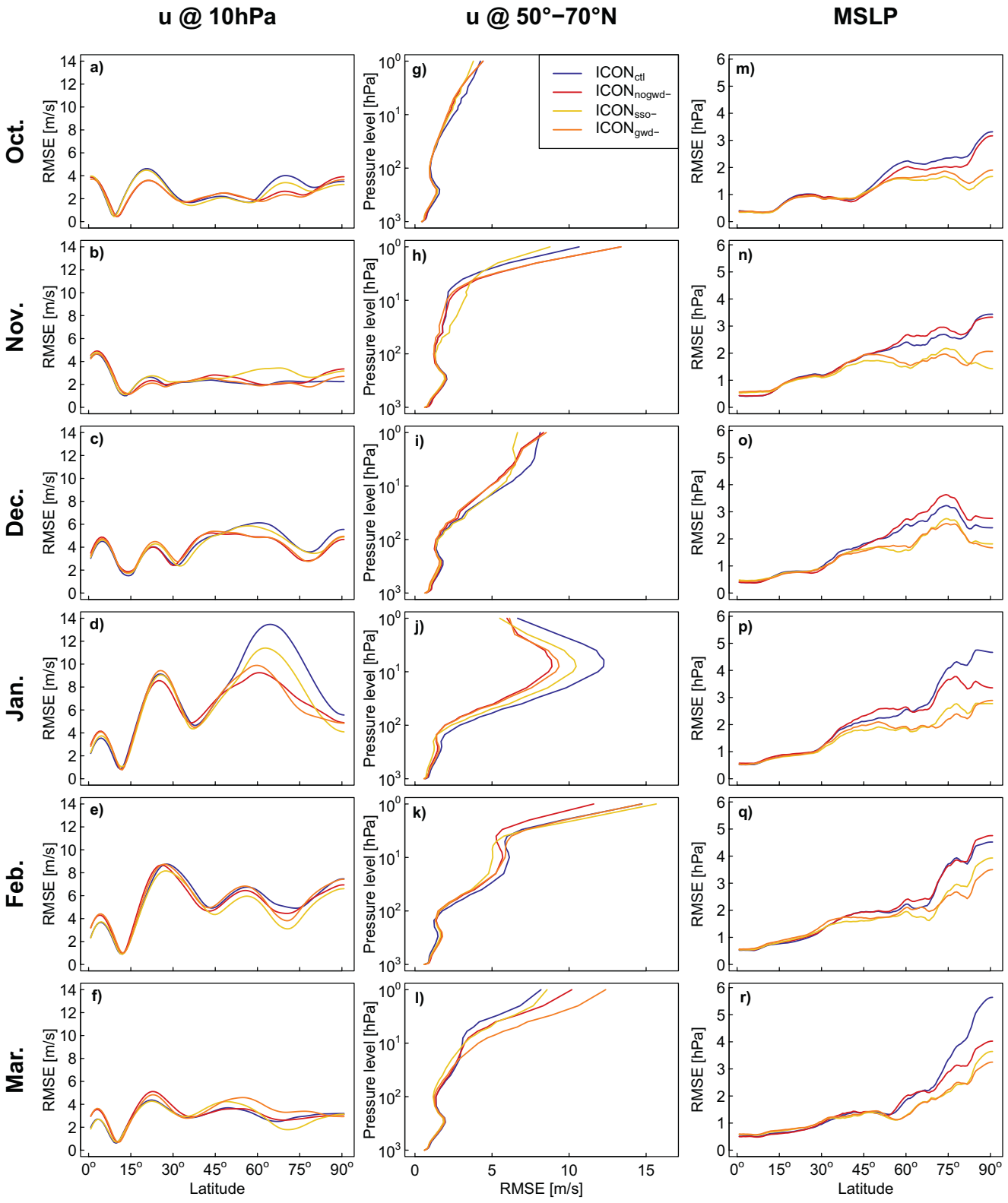

Figure 4. Zonal mean RMSE of 10 hPa zonal wind (a–f) and MSLP (m–r) as a function of latitude and the 50°N–70°N spatial RMSE of zonal wind as a function of altitude (g–i). The months from October to March are given in the rows. The different experiments are denoted by the colors blue ($ICON_{ctl}$), red ($ICON_{nogwd–}$), yellow ($ICON_{sso–}$) and orange ($ICON_{gwd–}$). RMSE, root-mean-square error.

frequency from November to January, whereas the ERA-Interim frequency is met by the ensemble spread of the different ICON experiments in February and March. In January, the ensemble spread shifts toward the more realistic lower frequencies in all sensitivity experiments.

3.3. Impact on the Tropospheric Circulation

Based on Baldwin and Dunkerton (2001), we investigate the downward coupling of the stratosphere to the troposphere, by analyzing the Northern Annular Mode (NAM), which is also known as the Arctic Oscillation in the troposphere. To achieve this, we calculated an index, based on the empirical orthogonal function (EOF) of daily zonal mean geopotential at each pressure level. The NAM is represented by the first principle component of this EOF. This method was suggested by Baldwin and Thompson (2009), as it shows the daily evolution of stratosphere-troposphere coupling most clearly and is robust. In contrast to Baldwin and Dunkerton (2001), weak stratospheric vortex events are not defined by the crossing of a threshold in the 10 hPa annular mode, but by the onset of MSWs in winter, a method also used by Polichtchouk, Shepherd, and Byrne (2018). MSWs are the strongest manifestation of stratosphere-troposphere coupling and represent a weak stratospheric vortex state. As they are furthermore defined at the same altitude as the threshold criteria, both approaches deliver comparable results. This was tested for the reanalysis as well as the ICON simulations. Furthermore, the analysis was carried out for strong stratospheric vortex events. The key findings on the downward coupling from the stratosphere to the troposphere stay the same.

A composite of the NAM index for MSW events is shown in Figure 5 for ERA-Interim (a) and the different ICON experiments (b)–(e). MSWs lead to a strong weakening of the NAM in the stratosphere. After the onset in the upper stratosphere, the signal propagates downward and remains present in the lower stratosphere with a lag of up to 60 days. Some of the weak NAM signals also propagate to the troposphere leading to a weakening of the Arctic Oscillation. The described characteristic pattern is reproduced in all ICON experiments - yet there are some distinct differences between the gravity wave drag experiments. The downward propagation of weak NAM signals in $ICON_{ctl}$ is rather fast with a stratospheric lag of only up to 50 days. In addition, the troposphere is significantly affected only in the first 30 days after the onset of the MSW. A decrease of gravity wave drag leads to a more persistent negative NAM signal in the lower stratosphere in all three sensitivity experiments. Furthermore, the influence on the troposphere is also increased, demonstrated by significant weak tropospheric NAM signals with a lag of up to 60 days. Although the reduction in the nonorographic gravity wave drag is not as strong as in Polichtchouk, Shepherd, and Byrne (2018), their finding of an intensified coupling between stratosphere and troposphere, due to reduced gravity wave drag, is reproduced with ICON. In addition, Figure 5d exemplifies that the same is true for a reduction of orographic gravity wave drag. All sensitivity experiments show that a reduction of gravity wave drag lead to a more persistent and realistic signal in the lower stratosphere and the troposphere. The coupling is strengthened not only for weak vortex events (MSWs), but also for strong stratospheric vortex events (not shown). The time-height development of the NAM is suitable to investigate the downward propagation of extreme events, but drawing conclusions on the mean state is more ambiguous. To understand how the mean tropospheric state is influenced by the adjusted parameterizations, we investigate MSLP biases and RMSEs.

Figure 6a illustrates the bias of the $ICON_{ctl}$ MSLP in January. Before we discuss the effects of the different gravity wave drag parameterization, we describe the pressure patterns in $ICON_{ctl}$: The weak polar vortex and the high sudden stratospheric warming frequency in $ICON_{ctl}$ is connected to a tropospheric signal resembling the negative phase of the Arctic Oscillation. The MSLP is significantly too high in the Arctic and too low in the mid-latitude Atlantic and east Pacific, which are typical features of the negative phase of the Arctic Oscillation. This pattern is visible in ICON from November to March, but biases in January are particularly large. These results agree with an earlier study evaluating ICON-A: Crueger et al. (2018) illustrate positive winter MSLP biases over the Arctic and negative winter biases over North Atlantic and Northwest Pacific.

In contrast to the results from previous section, the effect of the adjusted SSO scheme dominates in the troposphere, as the adjusted wake drag is acting directly within the troposphere. The two experiments with the adjusted SSO scheme show strongly reduced MSLP RMSEs in all months (cf. Figures 4m–4r and Figure 3c). This effect can be attributed to more suitable values for the low level wake drag constant and the critical

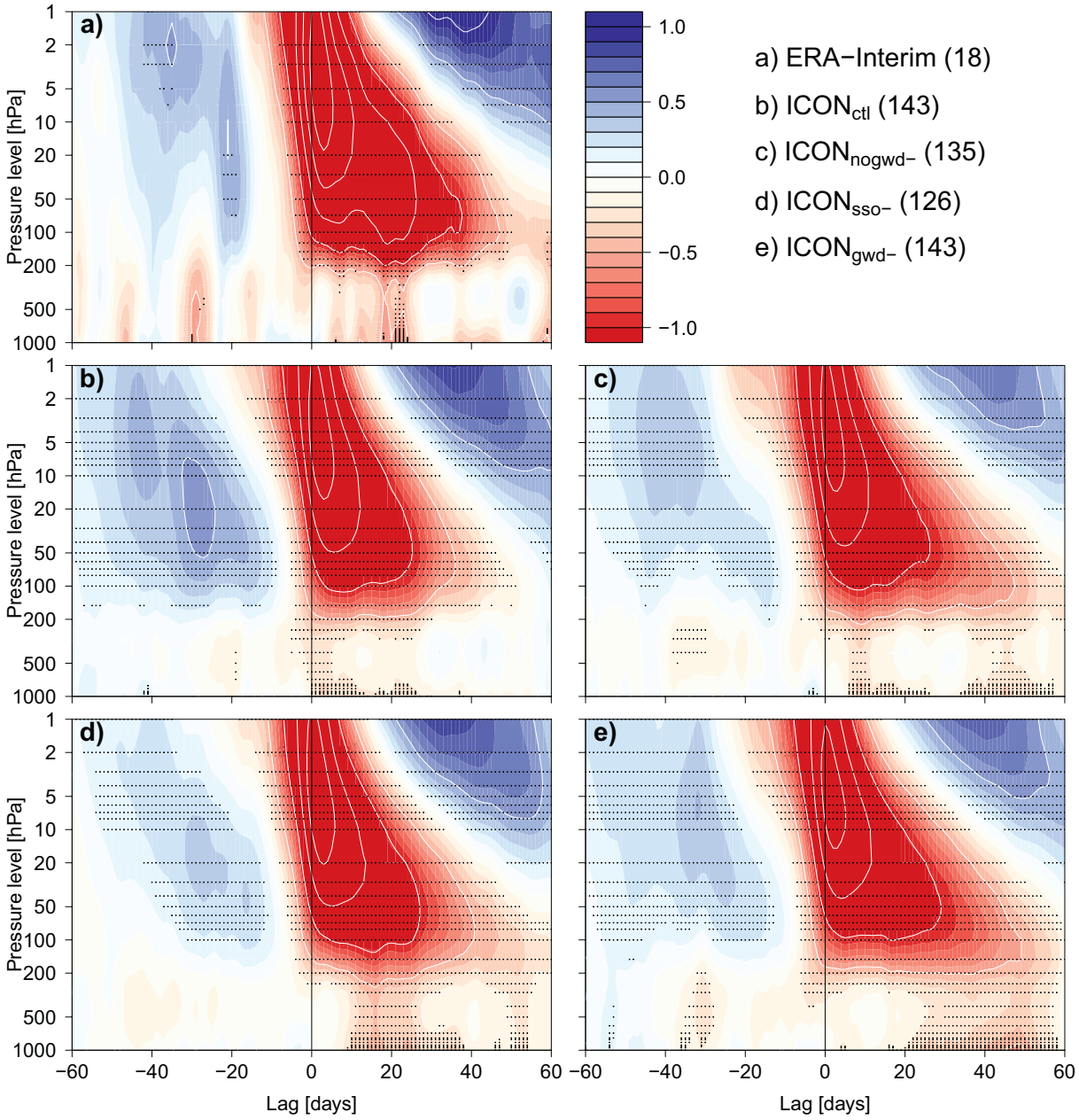

Figure 5. Composites of time-height development of the NAM for MSWs in winter. The composites are shown for ERA-Interim (a), ICON$_{ctl}$ (b), ICON$_{nogwd-}$ (c), ICON$_{sso-}$ (d), and ICON$_{gwd-}$ (e), respectively. The vertical line marks the day of onset of the MSWs at 10 hPa and the numbers in brackets indicate the total amount of MSWs. The NAM index is nondimensional. The contour interval for the white contours is 0.5 and stippling indicates difference from 0 at the 95% significance level. Note that due to common practice negative values (weak NAM) are red and positive values (strong NAM) are blue. MSWs, major stratospheric warmings; NAM, northern annular mode.

Froude number. Furthermore, the improved stratospheric winter circulation seems to impact the surface pressure patterns by stratosphere-troposphere coupling. However, this impact is very small compared to the direct effect by the wake drag. Indicators for this impact are the error reduction in ICON$_{nogwd-}$ compared to ICON$_{ctl}$ in addition to the improvements of ICON$_{gwd-}$ compared to ICON$_{sso-}$. In comparison to the control experiment the PCM RMSE average over the whole period is reduced by 2% in ICON$_{nogwd-}$, 29% in ICON$_{sso-}$ and 30% ICON$_{gwd-}$. Once more, the effects of the adjusted parameterizations are the largest in the high latitudes of the NH, so that the global mean RMSE differences are small. The adjustment of the SSO scheme is connected to a global error reduction in all months. The global mean RMSE averaged over the whole

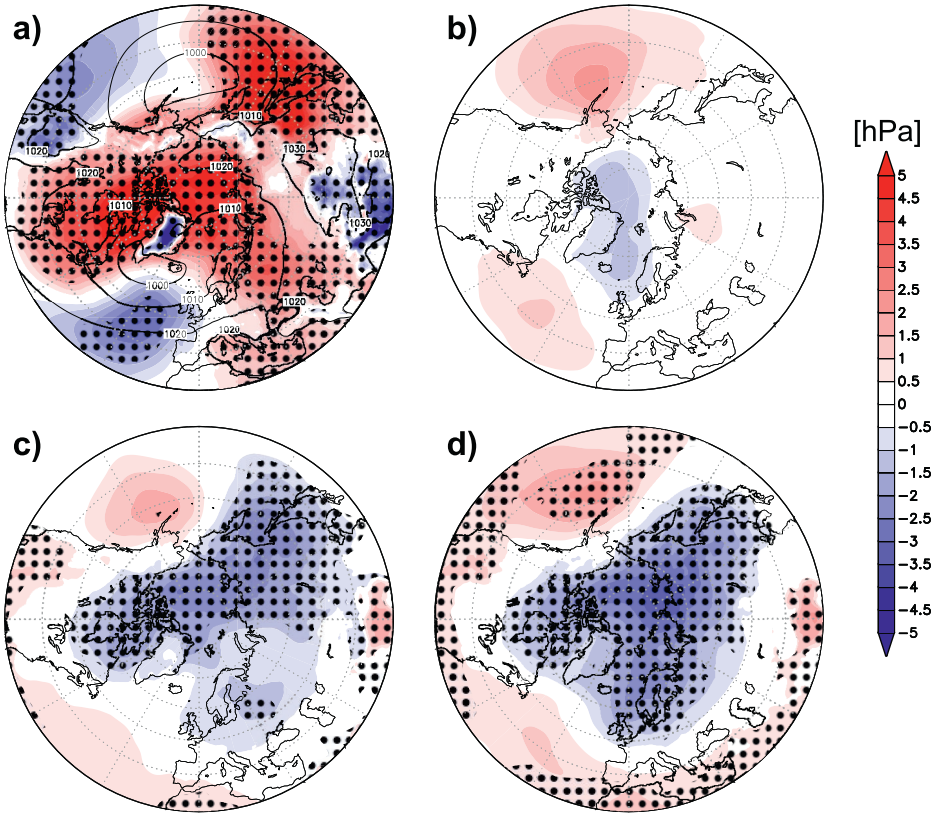

Figure 6. January mean MSLP bias relative to ERA-Interim for $ICON_{ctl}$ (a). $ICON_{nogwd-}$ (b), $ICON_{sso-}$ (c), and $ICON_{gwd-}$ (d) differences to $ICON_{ctl}$. Stippling indicates statistical significance with p-values small enough to satisfy the FDR criterion with $\alpha_{FDR} = 0.1$. The *p*-values are based on a two-sided Wilcoxon-test. The ERA-Interim climatological mean is depicted by black contours in (a). FDR, false discovery rate.

simulation period is increased by 1.6% in $ICON_{nogwd-}$ and decreased by 5.7% in $ICON_{sso-}$ and $ICON_{gwd-}$ (cf. Figure 3d).

The tropospheric RMSE improvements of the sensitivity experiments are also most evident in January. This is when the positive effect of the improved stratospheric circulation on the MSLP patterns becomes particularly apparent. The downward propagation of stratospheric signals leads to a reduction of the PCM RMSE of the MSLP by 13% in $ICON_{nogwd-}$ (cf. Figure 4p and Figure 3 c). Adjusting the SSO scheme produces an error reduction of 37% and the combined effect leads to an improvement of 42% in January. Figures 6b and 6c display the MSLP differences of the three ICON sensitivity experiments compared with $ICON_{ctl}$ in January. All sensitivity experiments exhibit lower pressure in the Arctic region and higher pressure in the Pacific and Atlantic region, leading to a reduction in bias magnitude and significance. Largest effects are visible in the central Arctic region in the experiments with adjusted SSO scheme. Due to reduced MSLP, the Arctic bias is small and less significant. Smaller bias reductions are also visible in the mid-latitudes. The direct effect from the adjusted wake drag in the SSO scheme is statistically significant and clearly stronger than the nonorographic gravity wave drag effect via the stratosphere. Largest and most significant effects are visible in $ICON_{gwd-}$.

4. Conclusions and Discussion

With the default settings for the gravity wave drag parameterizations, ICON-NWP is able to capture the basic behavior and variability of the stratospheric polar vortex. Nevertheless, our analysis shows that the vortex strength is underestimated in winter. On average, this leads to a too warm polar stratosphere. Largest

biases relative to the reanalysis are visible in January. This finding corresponds very well with the results from Borchert et al. (2019): They state that NH stratospheric winter westerlies are too weak in ICON, including simulations using an upper-atmosphere extension of ICON. We also detect a higher frequency of MSWs and weak downward coupling between stratosphere and troposphere in ICON. The weak stratospheric vortex is furthermore connected to positive MSLP bias in the Arctic, and negative bias in the mid and lower latitudes. This is a known issue in ICON-A (Crueger et al., 2018). They also see an easterly bias in annual stratospheric zonal winds in the high latitudes.

With the goal of strengthening the stratospheric westerly circulation, we follow Polichtchouk, Shepherd, and Byrne (2018) by reducing the nonorographic gravity wave drag by 20%. Particularly in January, the stratospheric polar vortex is strengthened and, thus, midstratospheric biases relative to the reanalysis are reduced. What is more, stratosphere-troposphere coupling is amplified, which is consistent with Polichtchouk, Shepherd, and Byrne (2018). Although the tropospheric signal from reduced nonorographic gravity wave drag is weak, there is a small reduction in the MSLP bias over the pole in January. A second experiment with a 60% reduction of nonorographic gravity wave drag is not shown, as this experiment exhibited larger biases in the upper stratosphere.

The results from the experiment with adjusted SSO scheme are not as easily interpreted, due to the fact that three parameters were altered simultaneously: Apart from decreasing the orographic gravity wave drag constant by 20%, we also altered the low level wake drag constant, which acts to reduce the near-surface drag, and the critical Froude number, which acts to decrease blocking by orography. These further adjustments were undertaken, as a first small sensitivity experiment with a solely adjusted orographic gravity wave drag constant showed only small effects. The adjustment of the latter two parameters was successful as they influence the MSLP patterns directly and lead to a strong and significant reduction of errors. The stratospheric vortex in winter is strengthened as a result of the reduced orographic gravity wave drag constant, but the effect is less strong than the effect from the nonorographic drag. However, keeping in mind, that the interplay between the changes caused by the altered orographic gravity wave forcing and the reduced wake drag may also induce changes in the stratospheric circulation. Furthermore, the stratosphere-troposphere coupling is strengthened and more realistic.

Combining the reduced nonorographic drag with the adjusted SSO scheme and thereby reducing the total amount of gravity wave drag, leads to a significant bias reduction in the troposphere and mid-stratosphere. Again, a reduction of gravity wave drag leads to intensified stratosphere-troposphere coupling. As a result of this sensitivity study, this experiment has been adopted as our current improved setup for seasonal experiments with ICON-NWP. Bearing in mind that the gravity wave drag is resolution dependent and potentially needs to be adjusted for differing resolutions.

Holton and Tan (1980) show that the stratospheric polar vortex is influenced by the Quasi-Biennial Oscillation (QBO). Depending on the year, our simulations with ICON are initialized with different QBO phases in September. Nevertheless, by January they all drift toward a weak easterly phase of the QBO. Numerous studies show that the easterly phase of the QBO is connected to a weak stratospheric polar vortex in winter (i.e., Anstey & Shepherd, 2014; Holton & Tan, 1980; Watson & Gray, 2014). Independent of the experiment, ICON does not produce a QBO and favors the easterly phase of the QBO. The gravity wave drag adjustment has no effect on this behavior. The dominance of the easterly phase in ICON could have a weakening effect on the stratospheric polar vortex. To test this, nudging in the tropical stratosphere would be highly favorable. At the time of performing this study, this kind of nudging was not implemented in ICON-NWP.

The focus of this study is on the stratospheric circulation of the Northern Hemisphere, as we see potential for improved seasonal predictions based on an accurate simulation of the highly variable northern polar vortex (Domeisen et al., 2019; Nie et al., 2019). Nevertheless, we acknowledge that the influence of gravity wave drag on the circulation of the Southern Hemisphere is of interest for future studies. The presented adjustment of the gravity wave drag parameterizations does not lead to a significant bias reduction in the Southern Hemisphere, however, no major negative impacts on the circulation of the Southern Hemisphere are identified.

Data Availability Statement

The ERA-interim data sets are available from ECMWF at http://apps.ecmwf.int/datasets/data/interim-full-daily. Access requires a registration/login as ECMWF user, which includes a specification of your name and email as well as an acceptance of ECMWF terms and conditions. The data used for this study are made available at https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=DKRZ\text{_}LTA\text{_}238\text{_}ds00001. The ICON model is distributed under an institutional license. A license can be obtained in contact to DWD via contacticon@dwd.de or by following the information on the public ICON website https://code.mpimet.mpg.de/projects/iconpublic.

Acknowledgments
The authors would like to thank Sabine Erxleben for technical support. R. Köhler is supported by the project QUARCCS funded by the German Federal Ministry for Education and Research (Grant/Award Number: 03F0777A) and the project POLEX funded by Helmholtz Association of German Research Centers (Grant Number: HRSF-0036). D. Handorf and K. Dethloff gratefully acknowledge the funding by the German Research Foundation for the project "ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)^{3}" (268020496-TRR 172). D. Handorf is also supported by the project ClimXtreme funded by the German Federal Ministry for Education and Research (Grant/Award Number: 01LP1901D). R. Jaiser is supported by the Helmholtz Climate Initiative REKLIM and the project SynopSys funded by the German Federal Ministry for Education and Research (Grant/Award Number: 03F0872A). We acknowledge support by the Open Access Publication Funds of Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung.

References

Anstey, J. A., & Shepherd, T. G. (2014). High-latitude influence of the quasi-biennial oscillation. *Quarterly Journal of the Royal Meteorological Society*, *140*, 1–21. https://doi.org/10.1002/qj.2132

Baldwin, M. P., & Dunkerton, T. J. (1999). Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *Journal of Geophysical ResearchAtmosphere*, *104*, 30937–30946. https://doi.org/10.1029/1999JD900445

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*, 581–584. https://doi.org/10.1126/science.1063315

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

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, *67*, 687–690.

Bonaventura, L. (2004). The ICON project: Development of a unified model using triangular geodesic grid. *Proceedings of the ECMWF annual seminar on developments in numerical methods for atmospheric and ocean modelling* (pp. 75–86). ECMWF. https://www.ecmwf.int/node/8263

Borchert, S., Zhou, G., Baldauf, M., Schmidt, H., Zängl, G., & Reinert, D. (2019). The upper-atmosphere extension of the ICON general circulation model (version: ua-icon-1.0). *Geoscientific Model Development*, *12*, 3541–3569. https://doi.org/10.5194/gmd-12-3541-2019

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S., Haynes, P., Krüger, K., et al.. (2011). Multimodel climate and variability of the stratosphere. *Journal of Geophysical ResearchAtmosphere*, *116*. https://doi.org/10.1029/2010JD014995

Butler, A., Charlton-Perez, A., Domeisen, D. I., Garfinkel, C., Gerber, E. P., Hitchcock, P., et al.. (2019). Sub-seasonal predictability and the stratosphere. In *Sub-seasonal to seasonal prediction* (pp. 223–241). Elsevier. https://doi.org/10.1016/B978-0-12-811714-9.00011-5

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

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., et al.. (2013). On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical ResearchAtmosphere*, *118*, 2494–2505. https://doi.org/10.1002/jgrd.50125

Charney, J. G., & Drazin, P. G. (1961). Propagation of planetary scale waves from the lower atmosphere to the upper atmosphere. *Journal of Geophysical Research*, *66*, 83–109. https://doi.org/10.1029/JZ066i001p00083

Choi, H.-J., Han, J.-Y., Koo, M.-S., Chun, H.-Y., Kim, Y.-H., & Hong, S.-Y. (2018). Effects of non-orographic gravity wave drag on seasonal and medium-range predictions in a global forecast model. *Asia-Pacific Journal of Atmospheric Sciences*, *54*, 385–402. https://doi.org/10.1007/s13143-018-0023-1

Cohen, J., & Jones, J. (2011). Tropospheric precursors and stratospheric warmings. *Journal of Climate*, *24*, 6562–6572. https://doi.org/10.1175/2011JCLI4160.1

Crueger, T., Giorgetta, M. A., Brokopf, R., Esch, M., Fiedler, S., Hohenegger, C., et al. (2018). ICON-A, The atmosphere component of the ICON earth system model: II. Model evaluation. *Journal of Advances in Modeling Earth Systems*, *10*, 1638–1662. https://doi.org/10.1029/2017MS001233

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al.. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., & Dunn-Sigouin, E., et al.. (2019). The role of the stratosphere in subseasonal to seasonal prediction: 2. predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmosphere*, *125*. e2019JD030923. https://doi.org/10.1029/2019JD030923

Dunkerton, T. J. (1997). The role of gravity waves in the quasi-biennial oscillation. *Journal of Geophysical ResearchAtmosphere*, *102*, 26053–26076. https://doi.org/10.1029/96JD02999

Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Reviews of Geophysics*, *41*. https://doi.org/10.1029/2001RG000106

Fritts, D. C., & Nastrom, G. D. (1992). Sources of mesoscale variability of gravity waves. Part II: Frontal, convective, and jet stream excitation. *Journal of the Atmospheric Sciences*, *49*, 111–127. https://doi.org/10.1175/1520-0469(1992)049⟨0111:SOMVOG⟩2.0.CO;2

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., et al. (2018). ICON-A, the atmosphere component of the ICON earth system model: I. Model description. *Journal of Advances in Modeling Earth Systems*, *10*, 1613–1637. https://doi.org/10.1029/2017MS001242

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). Nonparametric Statistical Methods (751, p. 848). John Wiley & Sons.

Holton, J. R., & Tan, H.-C. (1980). The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, *37*, 2200–2208. https://doi.org/10.1175/1520-0469(1980)037⟨2200:TIOTEQ⟩2.0.CO;2

Labitzke, K. (1981). Stratospheric-mesospheric midwinter disturbances: A summary of observed characteristics. *Journal of Geophysical Research: Oceans*, *86*, 9665–9678. https://doi.org/10.1029/JC086iC10p09665

Lott, F., & Miller, M. J. (1997). A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, *123*, 101–127. https://doi.org/10.1002/qj.49712353704

Matsuno, T. (1971). A dynamical model of the stratospheric sudden warming. *Journal of the Atmospheric Sciences*, *28*, 1479–1494. https://doi.org/10.1175/1520-0469(1971)028%3C1479:ADMOTS%3E2.0.CO;2

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al.. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*, 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Nastrom, G. D., & Fritts, D. C. (1992). Sources of mesoscale variability of gravity waves. Part I: Topographic excitation. *Journal of the Atmospheric Sciences*, *49*, 101–110. https://doi.org/10.1175/1520-0469(1992)049⟨0111:SOMVOG⟩2.0.CO;2

Nie, Y., Scaife, A. A., Ren, H.-L., Comer, R. E., Andrews, M. B., Davis, P., & Martin, N. (2019). Stratospheric initial conditions provide seasonal predictability of the north atlantic and arctic oscillations. *Environmental Research Letters*, *14*, 034006. https://doi.org/10.1088/1748-9326/ab0385

Orr, A., Bechtold, P., Scinocca, J., Ern, M., & Janiskova, M. (2010). Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, *23*, 5905–5926. https://doi.org/10.1175/2010JCLI3490.1

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*, 992. https://doi.org/10.1038/nclimate3121

Palmer, T., Shutts, G., & Swinbank, R. (1986). Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, *112*, 1001–1039. https://doi.org/10.1002/qj.49711247406

Polichtchouk, I., Shepherd, T. G., & Byrne, N. J. (2018b). Impact of parametrized nonorographic gravity wave drag on stratosphere-troposphere coupling in the northern and southern hemispheres. *Geophysical Research Letters*, *45*, 8612–8618. https://doi.org/10.1029/2018GL078981

Polichtchouk, I., Shepherd, T. G., Hogan, R., & Bechtold, P. (2018a). Sensitivity of the brewer–Dobson circulation and polar vortex variability to parameterized nonorographic gravity wave drag in a high-resolution atmospheric model. *Journal of the Atmospheric Sciences*, *75*, 1525–1543. https://doi.org/10.1175/JAS-D-17-0304.1

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*, 3548–3554. https://doi.org/10.1175/1520-0442(2004)017%3C3548:UWAFAA%3E2.0.CO;2

Romanowsky, E., Handorf, D., Jaiser, R., Wohltmann, I., Dorn, W., Ukita, J., et al.(2019). The role of stratospheric ozone for Arctic-midlatitude linkages. *Scientific Reports*, *9*, 7962. https://doi.org/10.1038/s41598-019-43823-1

Scinocca, J. F. (2003). An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *Journal of the Atmospheric Sciences*, *60*(4), 667–682. https://doi.org/10.1175/1520-0469(2003)060⟨0667:AASNGW⟩2.0.CO;2

Seviour, W. J., Gray, L. J., & Mitchell, D. M. (2016). Stratospheric polar vortex splits and displacements in the high-top CMIP5 climate models. *Journal of Geophysical Research: Atmosphere*, *121*, 1400–1413. https://doi.org/10.1002/2015JD024178

Sigmond, M., Scinocca, J., Kharin, V., & Shepherd, T. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*, 98. https://doi.org/10.1038/ngeo1698

Taylor, K. E., Williamson, D., & Zwiers, F. (2000). *The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations. Program for Climate Model Diagnosis and Intercomparison* (pp. 1–24). Lawrence Livermore National Laboratory, University of California.

Vosper, S. B., van Niekerk, A., Elvidge, A., Sandu, I., & Beljaars, A. (2020). What can we learn about orographic drag parametrization from high-resolution models? a case study over the rocky mountains. *Quarterly Journal of the Royal Meteorological Society*, *146*, 979–995. https://doi.org/10.1002/qj.3720

Warner, C., & McIntyre, M. (1996). On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *Journal of the Atmospheric Sciences*, *53*, 3213–3235. https://doi.org/10.1175/1520-0469(1996)053⟨3213:OTPADO⟩2.0.CO;2

Watson, P. A., & Gray, L. J. (2014). How does the quasi-biennial oscillation affect the stratospheric polar vortex? *Journal of the Atmospheric Sciences*, *71*, 391–409. https://doi.org/10.1175/JAS-D-13-096.1

Wilks, D. (2016). the stippling shows statistically significant grid points: How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, *97*, 2263–2273. https://doi.org/10.1175/BAMS-D-15-00267.1

Williams, K., van Niekerk, A., Best, M., Lock, A., Brooke, J., Carvalho, M., et al.. (2020). Addressing the causes of large-scale circulation error in the met office unified model. *Quarterly Journal of the Royal Meteorological Society*, *146*, 2597–2613. https://doi.org/10.1002/qj.3807

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*, 563–579. https://doi.org/10.1002/qj.2378