Comparing the effects of solar-related and terrestrial drivers on the northern polar vortex

Antti Salminen1,*, Timo Asikainen1, Ville Maliniemi2, and Kalevi Mursula1

1 Space Physics and Astronomy Research Unit, University of Oulu, 90014 Oulu, Finland
2 Birkeland Centre for Space Science, Department of Physics and Technology, University of Bergen, 5007 Bergen, Norway

Received 30 April 2020 / Accepted 7 October 2020

Abstract – Northern polar vortex experiences significant variability during Arctic winter. Solar activity contributes to this variability via solar irradiance and energetic particle precipitation. Recent studies have found that energetic electron precipitation (EEP) affects the polar vortex by forming ozone depleting NOx compounds. However, it is still unknown how the EEP effect compares to variabilities caused by, e.g., solar irradiance or terrestrial drivers. In this study we examine the effects of EEP, solar irradiance, El-Niño-Southern Oscillation (ENSO), volcanic aerosols and quasi-biennial oscillation (QBO) on the northern wintertime atmosphere. We use geomagnetic Ap-index to quantify EEP activity, sunspot numbers to quantify solar irradiance, Niño 3.4 index for ENSO and aerosol optical depth for the amount of volcanic aerosols. We use a new composite dataset including ERA-40 and ERA-Interim reanalysis of zonal wind and temperature and multilinear regression analysis to estimate atmospheric responses to the above mentioned explaining variables in winter months of 1957–2017. We confirm the earlier results showing that EEP and QBO strengthen the polar vortex. We find here that the EEP effect on polar vortex is stronger and more significant than the effects of the other drivers in almost all winter months in most conditions. During 1957–2017 the considered drivers together explain about 25–35% of polar vortex variability while the EEP effect alone explains about 10–20% of it. Thus, a major part of variability is not due to the linear effect by the studied explaining variables. The positive EEP effect is particularly strong if QBO-wind at 30 hPa has been easterly during the preceding summer, while for a westerly QBO the EEP effect is weaker and less significant.

Keywords: energetic electron precipitation / solar irradiance / ENSO / QBO / polar vortex

1 Introduction

During winter the dark polar stratosphere becomes colder than the sunlit mid- and low-latitudes. According to thermal wind shear balance this temperature difference is accompanied by a strong westerly thermal wind, the so-called polar vortex, surrounding the polar region. In addition to the meridional temperature gradient, the strength of polar vortex is affected by planetary waves, dynamical disturbances originating mostly in the subtropical and midlatitude troposphere (Charney & Drazin, 1961). In regions where these waves break or converge they decelerate the westerly winds and drive meridional residual circulation in the stratosphere, the so-called Brewer–Dobson circulation (BDC), which also transports trace gases, e.g., ozone from the equator to the high-latitudes (Butchart, 2014). The BDC is completed by upwelling in the tropics and downwelling in high latitudes, which leads to adiabatic cooling and warming respectively. Due to larger land-sea contrasts and mountain regions, planetary wave activity is stronger in the northern than in the southern hemisphere (van Loon et al., 1973). Therefore, the northern polar vortex exhibits significant variability. Variations of the polar vortex can be transmitted to the troposphere where they are seen as variations in the surface climate modes of the Northern Annular Mode (NAM) and its Atlantic part the North Atlantic Oscillation (NAO) (Baldwin & Dunkerton, 2001), which are the dominant modes of climate variability affecting air pressure, temperature and wind in the Northern Hemisphere during winter. On average a strengthening (weakening) of the polar vortex is followed approximately a week later by a positive (negative) response in the NAM/NAO (Baldwin & Dunkerton, 2001).

Earlier studies have shown that the Sun affects the polar vortex via two factors: varying solar irradiance (e.g., Labitzke & Van Loon, 1988; Kodera & Kuroda, 2002; Gray et al., 2010)
and energetic electron precipitation (EEP) (e.g., Rozanov et al., 2005; Seppälä et al., 2013; Arsenovic et al., 2016; Salminen et al., 2019). EEP forms reactive odd nitrogen (NOx) and hydrogen (HOx) oxides which deplete ozone catalytically (Crutzen et al., 1975). EEP affects directly the lower thermosphere and upper mesosphere, but during winter the NOx formed by EEP (EEP–NOx) can survive in polar darkness and descend from the mesosphere and lower thermosphere down to the stratosphere (Randall et al., 2007; Funke et al., 2014). Ozone loss in the stratosphere leads to a radiative warming in mid-winter darkness and to a radiative cooling after polar sunrise in spring (Sinnhuber et al., 2018). Several modelling studies (e.g., Rozanov et al., 2005; Baumgaertner et al., 2011; Arsenovic et al., 2016) and observational studies (e.g., Lu et al., 2008b; Seppälä et al., 2013; Salminen et al., 2019) have shown that increased EEP in winter is associated with reduced ozone in the polar stratosphere and strengthened polar vortex. Several studies have suggested that the EEP effect on mesosphere and upper stratosphere is due to ozone depletion by the descending EEP–NOx while the lower stratosphere and polar vortex are dynamically altered after the initial effect (e.g., Baumgaertner et al., 2011; Salminen et al., 2019; Asikainen et al., 2020). Polar vortex variations, also those initiated by EEP, propagate to the troposphere and surface (Baldwin & Dunkerton, 2001), which leads to significant correlation between EEP and NAO/NAM (Palamara & Bryant, 2004; Maliniemi et al., 2013, 2016).

Varying solar irradiance affects the stratosphere directly via ozone which absorbs UV radiation. Higher UV-irradiance increases ozone production and temperature especially in the tropical upper stratosphere (Soukharev & Hood, 2006; Frame & Gray, 2010), while the lower stratosphere is mostly affected by UV-induced changes in circulation (Kodera & Kuroda, 2002; Salby & Callaghan, 2004). Kodera & Kuroda (2002) showed that, in winter, the increased UV enhances the westerly wind in the subtropical upper stratosphere (near 30° latitude), and that this positive zonal wind anomaly moves downward and poleward in late winter. However, Camp & Tung (2007) found a positive correlation between the sunspot number (a proxy for solar radiation) and wintertime temperature in the polar stratosphere, which suggests that increased UV radiation weakens the polar vortex.

Although both EEP and solar radiation vary with the solar cycle, they correlate only weakly with each other. Total and spectral solar irradiance follow the 11-year sunspot cycle and, e.g., in the 150–250 nm wavelength range (UV-irradiation near the O2 photolysis region) irradiance is 4–8% larger in the sunspot maximum phase than in minimum phase (Floyd et al., 2003). The occurrence of EEP over the sunspot cycle is more complicated. The acceleration and precipitation of magnetospheric energetic electrons into the atmosphere is dominantly driven by high-speed solar wind streams originating from solar coronal holes (e.g., Meredith et al., 2011; Asikainen & Ruopola, 2016). Coronal holes tend to extend to lower solar latitudes a few years after the solar maximum during the declining phase of the sunspot cycle (Bame et al., 1976; Mursula et al., 2015), which is why EEP activity typically peaks during the declining phase.

In addition to the solar-related effects, the polar vortex is affected by internal atmospheric drivers. One of these drivers is the quasi-biennial oscillation (QBO), a mode of alternating zonal winds in the equatorial stratosphere with a period of 28–34 months (Baldwin et al., 2001). The QBO phase in the tropical stratosphere determines the latitude of the critical line where zonal wind changes from westerly (present in the winter hemisphere) to easterly (present in the summer hemisphere). This critical line acts as a boundary for planetary waves which can propagate only in the westerly background wind (Charney & Drazin, 1961). If the QBO wind is easterly, the critical line is located at higher latitudes and planetary waves are guided more poleward. If the QBO wind is westerly, the critical line is at a lower latitude and planetary waves can also propagate in the tropical stratosphere. As a result, the polar vortex is less disturbed in the westerly QBO phase than in the easterly phase. This is the so-called Holton–Tan effect (Holton & Tan, 1980). The QBO is also connected to meridional circulation which modulates the BDC so that the rate of upwelling in the equatorial stratosphere is higher during the easterly phase (Flury et al., 2013). By continuity also the downwelling at polar stratosphere is enhanced in QBO-E.

QBO phase also modulates how some other drivers affect the polar vortex. Palamara & Bryant (2004) and Maliniemi et al. (2013) found that the EEP effect on the NAO is only visible in the QBO-E phase. Salminen et al. (2019) found that the EEP effect on the northern polar vortex is stronger and more significant in December–March winter months in the QBO-E phase than in the QBO-W phase. Brewer–Dobson circulation and the downwelling in the polar stratosphere are known to be stronger in the QBO-E phase (Salby, 2008). Salminen et al. (2019) suggested that the EEP effect on the lower polar stratosphere is stronger in QBO-E because planetary wave activity and meridional circulation in the polar stratosphere are enhanced in QBO-E. The effect of solar UV radiation is also modulated by QBO. Labitzke & Van Loon (1988) showed that the solar lower stratosphere is warmer (corresponding to weaker polar vortex) in solar minimum than in maximum years if the QBO is easterly, but in westerly QBO phase the situation is reversed. Camp & Tung (2007) found the same positive correlation between polar stratospheric temperature and solar radiation in westerly QBO phase but in easterly QBO phase the correlation did not exist.

Another influence to the polar vortex variability is exerted by the El-Niño-Southern Oscillation (ENSO). This weather mode dominates the variability of, e.g., sea surface temperatures, winds, and circulation cells (Hadley & Walker cells) in the Pacific. Thus, ENSO affects the formation of planetary waves (Honda et al., 2001; Manzini et al., 2006; Garfinkel & Hartmann, 2008) and upwelling in the equatorial lower stratosphere (Randel et al., 2009). Positive ENSO (El-Niño) winters are associated with an accelerated Hadley cell circulation (Seager et al., 2003), increased planetary wave activity (e.g., Manzini et al., 2006) and warmed polar stratosphere (e.g., Garfinkel & Hartmann, 2007). Pozo-Vázquez et al. (2001) found that in negative ENSO (La Niña) winters the NAO mode is more positive, corresponding to a colder and stronger polar vortex. Similarly as for EEP and solar irradiance effects, the QBO also seems to modulate the effect of ENSO on the polar vortex. Garfinkel & Hartmann (2007) concluded that when the QBO is easterly, the polar vortex effect of ENSO is reduced. Volcanic eruptions can also affect the polar vortex by releasing massive amounts of sulfuric compounds which form
aerosols in the stratosphere. Aerosols absorb incoming solar radiation, thereby warming the stratosphere and cooling the troposphere (Robock, 2000). Volcanic aerosols can strengthen the polar vortex by warming the stratosphere at low latitudes, thereby increasing the thermal gradient between the polar and mid-latitude regions (Van Loon & Labitzke, 1987). This effect tends to enhance the NAO towards its positive phase during a year or two after the eruption, depending on the latitude of the erupted volcano (Robock & Mao, 1992). However, the temperature and wind anomalies caused by volcanic aerosols affect the propagation of planetary waves so that the Brewer–Dobson circulation strengths (Graf et al., 2007; Toohey et al., 2014) and can therefore counteract the initial vortex strengthening.

The combined effects of all these solar-related (EEP, solar irradiance) and terrestrial (QBO, ENSO, volcanic aerosols) drivers and how they relate to each other are still partly open issues. Many studies have considered only one or two of these variables at a time. In this study we examine the atmospheric effects of all these five drivers in the Northern Hemisphere winter to clarify the interannual variability in this region. We use the multiple linear regression analysis to estimate responses of zonal wind and temperature to the mentioned variables. In order to take the possible QBO modulation of other drivers into account, we also perform the analysis in the two QBO phases separately. Section 2 describes the data and methods. Section 3 presents the results of multiple linear regression without QBO phase separation. Section 4 compares the results of multiple linear regression in the two QBO phases. In Section 5 we give our conclusions.

2 Data and methods

To study atmospheric variables we use the ERA-40 (Uppala et al., 2005) and ERA-Interim (Dee et al., 2011) reanalysis datasets provided by European Centre for Medium-Range Weather Forecasts (ECMWF) (https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets). ERA-40 data cover the period from September 1957 to August 2002 and ERA-Interim data are available from January 1979 onwards. ERA-40 and ERA-Interim data resemble each other in the overlapping time period from January 1979 to August 2002, but there are some differences, e.g., in upper stratospheric temperatures in the polar region. We use here a composite data series of ERA-40 and ERA-Interim data which is formed by scaling the ERA-40 data to ERA-Interim level with a novel methodology developed recently by Asikainen et al. (2020). The composite dataset consists of scaled ERA-40 data for August 1957 to December 1978 and ERA-Interim data for 1979–2017. In this study we examine monthly and zonally averaged values of temperature and zonal wind at latitudes 0°–90° N (2.5° resolution) and at 23 pressure levels from the surface (1000 hPa) to upper stratosphere (1 hPa).

We use geomagnetic Ap-index (http://isi.unistra.fr) as a proxy to quantify energetic electron precipitation. Ap-index correlates well with precipitating electron fluxes measured by NOAA/POES satellites (in monthly scale $c_c = 0.86$, $p < 10^{-11}$ and is used as the basis of empirical models of EEP-NOx (Funke et al., 2016). As a proxy index for solar irradiance we use the international sunspot number (version 1; SSN) provided by WDC-SILSO of Royal Observatory of Belgium (http://www.sidc.be/silso/versionarchive). Since the version 1 SSN data has not been updated after May 2015, we extend the version 1 data with the version 2 data and scale it by factor 0.7 for June 2015 – December 2017. As an ENSO index we use the Niño 3.4 index which is the sea surface temperature averaged over 5° S–5° N latitudes and 120°–170° W longitudes in the Pacific (https://www.esrl.noaa.gov/psd/data/timeseries/monthly/NINO34/). To quantify volcanic activity we use stratospheric aerosol optical depth (AOD) at 550 nm averaged over the Northern Hemisphere (Sato et al., 1993) (https://data.giss.nasa.gov/modelforce/strataer/). The AOD time series does not cover times after 2012. Since there was only weak volcanic activity after 2012, we extend the time series to 2017 with zero values. The results of our analysis remain almost exactly the same if the AOD time series is extended with low non-zero values (e.g., 0.0025). We determine the QBO index by averaging the zonal wind at 30 hPa over latitudes 10°–10° N and all longitudes. We studied winters 1957/1958 – 2016/2017, but we excluded winters 1984/1985 and 2003/2004 since unusually early and long-lasting stratospheric sudden warmings occurred during these winters (Manney et al., 2005), and they have been found to be notable outliers in regression analyses relating polar vortex to EEP (Salminen et al., 2019). There are also other winters with rather strong SSW events, e.g., winters 2008/2009 and 2012/2013, but wind reversal periods were shorter and reversal times later in winter than in the excluded winters, and they did not appear as similar, clear outliers in our analysis.

We calculate multiple linear regressions (MLR) in which the response variable is either zonal wind or temperature and the explaining variables are $A_p$, SSN, Nino3.4, AOD and QBO. Both the response and the explaining variables are first detrended by subtracting from the data a smoothly changing trend estimated with the LOWESS (LOcally WEighted Scatterplot Smoothing) method (Cleveland, 1979) using a 31-year window. Such smooth local trends model the long-term (over 10 years) variation in the different variables better than a linear trend. Due to this detrending our results represent the interannual variability of the studied parameters. On the other hand, the long-term variations, e.g., in ozone, which decreased until 1990–2000 due to increased stratospheric chlorine loading, are removed by this detrending and do not affect the results. As an example of the explaining variables Figure 1 shows detrended and standardized time series of February $A_p$, SSN, Nino3.4, AOD and QBO. The regressions are then calculated separately at each latitude-pressure level grid point for each winter month (December–March). The regression model is

$$ Y_t = x + \sum_{k=1}^{m} (\beta_k X_{k,t}) + \epsilon_t, $$

in which $Y$ is the response variable, $x$ is constant, $m$ is the number of explaining variables, $X_k$ is one of the explaining variables and $\beta_k$ is the corresponding regression coefficient ($k = 1, \ldots, m$), and $\epsilon$ is the residual term. To calculate regressions we use the Cochrane–Orcutt method (Cochrane & Orcutt, 1949) in which the residual term is modelled as an autoregressive AR(1) process so that $\epsilon_t = \rho \epsilon_{t-1} + e_t$, where $\rho$ is the lag-1 autocorrelation coefficient and $e_t$ is white noise. The AR(1) residual is used here to model both stochastic
noise and any unaccounted externally driven or internal variability. The regression coefficients $\beta_k$ and the noise autocorrelation coefficient $\rho$ are both iteratively solved from the modified regression model equation

$$Y_t - \rho Y_{t-1} = \alpha + \sum_{k=1}^{K} \beta_k (X_{k,t} - \rho X_{k,t-1}) + e_t.$$  

(2)

Significances of the regression parameters obtained from equation (2) can be estimated with the Student’s $t$-test, because in equation (2) the remaining noise term is uncorrelated white noise. With this method we can estimate the significance of obtained results more accurately since autocorrelation in the residuals may distort the significance estimations in the case of normal multiple linear regression analysis of equation (1). We have also used a Monte-Carlo simulation to estimate the significances in order to confirm the correctness of the regression approach. We first generated a new dataset for each explaining variable by randomly mixing the phases of the Fourier components of the original time series. This surrogate random dataset has the same mean, variance and autocorrelation function as the corresponding original dataset. Then we use the generated datasets to calculate the similar MLR analysis as earlier. By repeating this procedure 1000 times, we get a distribution of values for each regression parameter. The $p$-value is derived as a fraction of values from the Monte Carlo simulations which deviates more from zero than the value obtained with the original real data. We found that the significances estimated with the Monte Carlo method are almost the same as the significances calculated with the Student’s $t$-test, which validates the use of the Cochrane–Orcutt method. Finally we multiply each regression parameter $\beta_k$ with the standard deviation of the corresponding explaining variable $X_k$. As a result we obtain a change in response variable $Y$ which corresponds to an increase of one standard deviation in the explaining variable $X_k$. Standard deviations are calculated from the whole time series of explaining variables to preserve comparability between results of different months. The regression model used here is described in more detail in the earlier studies by Maliniemi et al. (2018) and Salminen et al. (2019).

To study how well the computed regression models explain the variability in a response variable we calculate the coefficient of determination ($R^2$) for each regression model. $R^2$ is calculated as the square of Pearson’s correlation coefficient between the observed and fitted values of the response variable. The $R^2$ indicates how large a fraction of the variability in the response variable is explained by the regression model. We also study how much each explaining variable contributes to this explained variability by calculating $\Delta R^2$, which for explaining variable $X_k$ is defined here as

$$\Delta R^2 = R^2_{\text{complete}} - R^2_{\text{reduced}},$$  

(3)

in which $R^2_{\text{complete}} = R^2$ is the coefficient of determination of the complete model ($X_k$ included) and $R^2_{\text{reduced}}$ is the coefficient of determination of the model without $X_k$ as an explaining variable. We use the $F$-test for each explaining variable in each regression model to test whether adding an explaining variable to the regression model significantly improves the model. In this case the test statistic $F$ is

\[ F = \frac{(R^2_{\text{complete}} - R^2_{\text{reduced}})/(K_2 - K_1)}{(1 - R^2_{\text{complete}})/(N - K_2)} \]
where $\text{SSE}_{\text{reduced}}$ and $\text{SSE}_{\text{complete}}$ are sums of squares of residuals for reduced and complete model, respectively, $n$ is the number of observations and $m$ is the total number of explaining variables. The test statistic $F$ follows an $F$-distribution with degrees of freedom of 1 and $n - m - 1$. We also test the goodness of the complete model by calculating $F$-test with a test statistic

$$ F = \frac{\text{SSR}_{\text{complete}}}{\text{SSE}_{\text{complete}} / (n - m - 1)}, $$

in which $\text{SSR}_{\text{complete}}$ is the sum of squares due to regression. In this case $F$ follows a $F$-distribution with degrees of freedom $m$ and $n - m - 1$. We also tested collinearities between the explaining variables by calculating variance inflation factor (VIF) for each variable in each setting, and in all cases VIF values remained lower than 2 indicating that multicollinearity between the explaining variables is not a problem.

We also calculate regressions separately in the easterly (QBO-E), and westerly (QBO-W) QBO phase. In these regressions QBO is not used as an explaining variable. Individual months are classified into either QBO phase separately and, therefore, months of the same winter are not necessarily in the same QBO phase. We also tested to separate whole winters into the two QBO phases and the results remained essentially the same. When sectioning the data into two groups based on the QBO phase the regression model needs to be modified to allow for different regression parameter values in the two groups. This can be done by introducing a binary indicator variable $I$, which has the value of 1 for QBO-E and 0 for QBO-W. The modified regression model is now

$$ Y_i = \alpha + aI_i + \sum_{k=1}^{m} (\beta_k X_{i,k} + b_k I_i X_{i,k}) + \epsilon_i, $$

The response in $Y$ to $X_{i}$ in QBO-W ($I = 0$) is now described by the regression parameter $\beta_k$ and in QBO-E ($I = 1$) the response is obtained as a sum $\beta_k + b_k$. With the autoregressive error term $\epsilon_i$ the regression equation can be written as,

$$ Y_{i} - \rho Y_{i-1} = \alpha (1 - \rho) + a(I_i - \rho I_{i-1}) + \sum_{k=1}^{m} \beta_k (X_{i,k} - \rho X_{i,k-1}) $$

$$ + \sum_{k=1}^{m} b_k (I_i X_{i,k} - \rho I_{i-1} X_{i,k-1}) + \epsilon_i. $$

The regression parameters are then solved iteratively as described by Malinieri et al. (2018) and Salminen et al. (2019). In computing the significances of the responses in QBO-E and QBO-W we need the variances of the regression parameters. For QBO-W ($I = 0$) the variances are directly obtained as the variance of the $\beta_k$ parameters, but for QBO-E ($I = 1$) the variances are obtained as a sum

$$ \text{var}(\beta_k + b_k) = \text{var}(\beta_k) + \text{var}(b_k) + 2 \text{cov}(\beta_k, b_k), $$

where the variances of $\beta_k$ and $b_k$ and their covariances are obtained from the least squares fitting procedure.

### 3 MLR responses of zonal wind and temperature

Figure 2 presents the MLR responses in zonally averaged zonal wind to $Ap$, SSN, Nino3.4, AOD and QBO (rows 1–5) separately for December–March (columns 1–4) in the Northern Hemisphere (latitudes $0^\circ$–$90^\circ$ N). In February and March we used a one-month lead time for $Ap$ in order to take into account the descent time of NO$_x$ from upper atmosphere (Funke et al., 2014). The responses were similar but slightly weaker in all winter months for 0-month and 2-month lead times. We used here a 6-month lead time for the QBO which is roughly the transport time from equator to pole associated to Brewer–Dobson circulation, and which has been found to maximize the difference in the EEP response between the two QBO phases (Salminen et al., 2019). We also tested different lead times for QBO and the results were essentially the same but somewhat weaker using the lead times of 0–5 or 7–8 months. We also calculated the regression analysis by including QBO at 50 hPa as an additional explaining variable. In this case the responses to $Ap$, SSN, Nino3.4 and AOD remained the same but the responses to the two QBO indices were weaker and less significant than the responses to one QBO index in Figure 2.

Figure 2 (first row) shows that increased $Ap$ is associated with a significant strengthening of westerly winds (the polar vortex) in the stratosphere at latitudes $50^\circ$–$90^\circ$ N in every winter month, with the strongest response in January and February. Corresponding responses are also seen in the troposphere. Zonal wind responses to SSN (Fig. 2, second row) are not significant at high latitudes except for the small decreases seen in the troposphere in January and March and in the upper stratosphere in February. The most persistent response to SSN is the increase of equatorial upper tropospheric winds, which is seen in December–February. ENSO responses in zonal wind (Fig. 2, third row) are positive and significant in the low-latitude ($<40^\circ$) troposphere in all winter months, while at higher latitudes an increased Nino3.4 index is associated with a significant weakening of zonal wind in the troposphere in January and in upper stratosphere in February. Figure 2 (fourth row) shows that AOD enhances the polar zonal wind in the stratosphere and troposphere in December while in January the positive response at high latitudes is significant only in the troposphere. AOD also causes a negative response at $30^\circ$–$50^\circ$ N in December–February. In March AOD strengthens the zonal wind in the troposphere at $20^\circ$–$40^\circ$ N and weakens it at $40^\circ$–$60^\circ$ N. The bottom row of Figure 2 shows the QBO signal in the equatorial stratosphere with a positive response at 30–60 hPa and a negative response at 7–10 hPa. QBO also strengthens the zonal wind in the stratosphere at $40^\circ$–$70^\circ$ N in December and January while in February and March this response is weaker.

Figure 3 shows the MLR responses in temperature, similarly to the zonal wind responses in Figure 2. An increase in $Ap$ (Fig. 3, first row) is related to a cooling in the polar lower stratosphere in January–March and warming in the polar upper stratosphere in January and March. (The upper polar stratosphere is warming also in February but does not quite reach statistical significance). Increased SSN (Fig. 3, second row) is associated with a cooling in the polar upper stratosphere in January and a warming at 8–10 hPa poleward of $70^\circ$ in February. The most significant temperature responses to ENSO (Fig. 3, third row)
are, as expected, seen in the tropics where positive ENSO is associated with tropospheric warming and cooling of tropopause/lower stratosphere (occasionally even wider in stratosphere). ENSO also seems to warm the polar upper stratosphere in December. AOD (Fig. 3, fourth row) cools the troposphere and warms the tropopause/lower stratosphere at low latitudes, which is opposite to the ENSO effect. AOD does not cause significant temperature variations in the polar troposphere or stratosphere. QBO (Fig. 3, fifth row) is associated with a pattern of equatorial temperature responses, in which the temperature is increased (decreased) below (just above) the positive zonal wind response at 30–60 hPa and above (below) the negative zonal wind response at 7–10 hPa. An almost reversed temperature pattern is located at subtropical latitudes (20°–40° N). In the polar region the positive QBO is associated with a cooling in the lower stratosphere in December–March and warming in the upper stratosphere in February and March.

The above described effects of Ap on zonal wind and temperature in the polar stratosphere are similar to those found in earlier, less extensive studies (e.g., Baumgaertner et al., 2011; Seppälä et al., 2013; Salminen et al., 2019). Warming in the upper stratosphere in January and February is likely due to decreased ozone which acts as a radiative cooler in darkness. Several studies (e.g., Langematz et al., 2003; Baumgaertner et al., 2011; Salminen et al., 2019) have suggested that the strengthening of polar vortex and cooling in the polar lower stratosphere are dynamical responses to the initial warming in the upper stratosphere (and mesosphere). This is most likely due to the wave–mean flow interaction in which planetary waves and the background wind interact with each other in a positive feedback loop (Andrews et al., 1987). Salminen et al. (2019) suggested that warming in the upper stratosphere changes the propagation of planetary waves so that less waves are converged in the lower stratosphere, which leads to a stronger polar vortex, weakened Brewer–Dobson circulation and anomalous cooling in the polar lower stratosphere.

Figure 2 (second row) shows that increased solar UV radiation (increased sunspot number) does not cause a significant effect on the polar vortex. This agrees, e.g., with Labitzke & Van Loon (1988) who did not find a sunspot cycle effect on polar vortex without separating the data according to the QBO phase. The warming of polar stratosphere as a response to SSN in February (see Fig. 3, second row) was also found by Lu et al. (2017) who suggested that this response is due to
increased breaking of planetary waves in the polar upper stratosphere. The negative response of the polar vortex to Nino3.4 seen in Figure 2 and the related warming seen in Figure 3 are in agreement with the results by Garfinkel & Hartmann (2007) who linked El Niño winters (positive Nino3.4 index) to warmer polar stratosphere (weaker polar vortex). Garfinkel & Hartmann (2008) suggested that this ENSO effect is due to increased planetary wave activity in El Niño winters. Temperature and zonal wind responses to volcanic aerosols in December and January agree with earlier results, e.g., by Van Loon & Labitzke (1987). Aerosols radiatively warm the lower stratosphere at sunlit latitudes, which increases the temperature gradient between the pole and lower latitudes and, thereby, enhances the polar vortex. Note, however, that only two large volcanic eruptions occurred during the studied period (see the AOD time series in Fig. 1), which may affect the amplitudes and significances of AOD responses. The polar vortex responses to QBO support the Holton–Tan mechanism (Holton & Tan, 1980), in which polar vortex is stronger in the QBO westerly phase than in the easterly phase since planetary waves are directed more poleward in the QBO-E phase. We find that this response is strong and significant in early winter (December–January) while in late winter it is weaker and only marginally significant. This is in agreement with earlier studies by Lu et al. (2008a) and Maliniemi et al. (2016).

Figure 4 shows $\Delta R^2$ at each latitude-pressure level grid box for each explaining variable (rows 1–5) in zonal wind response in December–March (columns 1–4), corresponding to the regression results shown in Figure 2. The total coefficient of determination at each grid box is shown in the bottom row of Figure 4. It is interesting to see that the Ap (Fig. 4, first row) explains the largest amount of variance in polar vortex in January and February (about 10%–20%) and a bit less (about 5–10%) in December and March. In December the QBO explains a larger fraction of the variance of high-latitude zonal wind (about 10–25%) than EEP. The total coefficient of determination (Fig. 4, sixth row) is 25–35% at best in the high-latitude stratosphere and troposphere, which indicates that a major part of the variance during 1957–2017 is not explained by the linear model including the studied variables. The unexplained part of the variance may be, e.g., due to an unaccounted or unknown variable, due to internal variability of the polar vortex or due to more complicated non-linear relations between explaining variables and polar vortex. One unaccounted factor which probably has a significant effect on the polar vortex is planetary wave activity. However, the effect of planetary waves is difficult...
to estimate and interpret since their activity and propagation in the stratosphere is determined not only by the tropospheric source but also by the state of the stratosphere (Scott & Polvani, 2004). Moreover, the effects of other drivers may depend on the properties of the planetary waves.

4 MLR responses of zonal wind in the two QBO phases

Figures 5 and 6 show the responses of zonal wind to Ap, SSN, Nino3.4 and AOD separately in the QBO-E and QBO-W phases respectively. We use here the same QBO index to determine the phase which was used as an explaining variable in Section 3 (QBO determined at 30 hPa and with a 6-month lead time). Figure 5 (first row) shows that an increase in Ap leads to the strengthening of the polar vortex in December–March in the QBO-E phase. These responses are somewhat stronger and closely similar to the responses obtained without QBO separation (Fig. 2, first row). In the QBO-W phase (Fig. 6, first row) a significant positive response of polar vortex to Ap exists only in January, but even this response is only weakly (90% confidence level) significant. In QBO-W the Ap weakens the polar vortex in the upper stratosphere in March contrary to the response in QBO-E. In the troposphere the Ap is associated with a positive response at 40–50° N and a negative response at 20–30° N in December in QBO-W. These responses are not seen in QBO-E or without QBO separation. The response of the polar vortex to Ap in QBO-E phase is larger and more significant in all winter months than the responses to any of the other variables, while in QBO-W phase the Ap effect
on polar vortex is not very significant. These findings agree with the earlier studies by Palamara & Bryant (2004), Maliniemi et al. (2013, 2016) and Salminen et al. (2019), which conclude that the EEP effect on polar vortex and NAO is stronger in QBO-E than in QBO-W phase.

The zonal wind responses to SSN in QBO-E (Fig. 5, second row) and QBO-W (Fig. 6, second row) are mostly weak and insignificant, as without QBO separation (Fig. 2, second row). In QBO-W the weakening of stratospheric polar vortex is weakly significant in January and slightly more intense in January and February than in QBO-E. Even though there is only a minor difference in the SSN effect on polar vortex between the QBO phases, the observed stronger negative effect in QBO-W is in accordance with earlier results (Labitzke & Van Loon, 1988; Camp & Tung, 2007). Camp & Tung (2007) did not find significant correlations between polar vortex and sunspot cycle in QBO-E, which agrees with our results. The negative response in polar troposphere is significant in January in QBO-E and in February and March in QBO-W. In QBO-E there is a weakly significant positive response in December in the upper stratosphere at 20°–40° N which is not significant without QBO separation or in QBO-W.

The positive response to the ENSO seen in Figure 2 (third row) around 20° latitude in the troposphere in all winter months remains roughly the same both in QBO-E (Fig. 5, third row) and in QBO-W (Fig. 6, third row). In QBO-E ENSO strengthens the polar vortex in March, but this response is only weakly significant and it is not seen without QBO separation or in QBO-W. In the QBO-E ENSO also causes a positive response near the stratosphere and negative response in middle stratosphere at low latitudes in March. In QBO-W the ENSO produces a weakly significant negative response in the polar stratosphere in March. In QBO-W the ENSO also causes a negative response in the upper stratosphere at 20°–40° N in December which is reversed to positive in January. The differences in ENSO effect on polar vortex between the QBO phases agree with the earlier results by Garfinkel & Hartmann (2007) who found that the negative effect of ENSO on polar vortex is weakened in QBO-E phase.

Responses to AOD in QBO-E (Fig. 5, fourth row) are quite similar to the responses without QBO separation (Fig. 2, fourth row). However, at high latitudes the positive response to AOD in December is less significant and the negative response in March is more significant in QBO-E than without QBO separation. In QBO-W the AOD (Fig. 6, fourth row) produces only weak responses. There is a significant response in the low-latitude tropopause in December, but no other significant responses, contrary to the QBO-E. In the QBO-W phase, there
is a weakly positive response to AOD in the polar vortex in March, which is opposite to the QBO-E phase. These weak responses in QBO-W are probably due to weak volcanic activity in QBO-W winter months. Two largest volcanic eruptions of the studied time period (in 1982 and 1991) affected mostly in QBO-E winters (1982/1983 and 1991/1992).

5 Discussion and conclusions

We have studied here the effects of solar-related (Ap/EEP and SSN/solar irradiance) and terrestrial drivers (ENSO, volcanic activity and QBO) on the northern polar vortex. We found that the polar vortex is significantly affected by energetic electron precipitation, as increased EEP is associated with a stronger polar vortex, cooler polar lower stratosphere and warmer polar upper stratosphere. Similar EEP responses have been found in earlier studies (Rozanov et al., 2005; Baumgaertner et al., 2011; Seppälä et al., 2013; Arsenovic et al., 2016). However, here we have demonstrated that the EEP effect on the polar vortex dominates over all the other studied effects in all winter months, except in December, when the QBO-related variability is comparable to the EEP effect.

In particular, our results indicate that the EEP is the dominant solar-related driver affecting the polar vortex, while the role of the varying solar irradiance is considerably weaker and mostly insignificant. Solar irradiance weakens the polar vortex in the upper stratosphere in February. In troposphere the solar irradiance is related to a negative zonal wind response in the polar region in January and March, and to a positive response in the equator in December–February. As to polar stratospheric temperature our results agree with Labitzke & Van Loon (1988) and Camp & Tung (2007) who did not find a clear correlation between solar irradiance and northern polar stratospheric temperatures unless winters are separated according to the QBO phase. Our results are partly in contradiction with the top–down effect of solar irradiance suggested by Kodera & Kuroda (2002). They used composite analysis and found that the polar vortex in December and January is enhanced in high SSN years compared to low SSN years. This difference in results may be due to different time periods between our study (1957–2017) and that of Kodera & Kuroda (2002) (1979–1998) or due to different analysis methods (multilinear regression vs. composite analysis).

ENSO weakens the polar vortex in January and February, while volcanic activity strengthens the polar vortex in December and January. These effects are in accordance with earlier studies on ENSO (e.g., Garfinkel & Hartmann, 2007) and volcanic activity (e.g., Van Loon & Labitzke, 1987). QBO (here determined at 30 hPa and with a 6-month lead time) strengthens

Fig. 6. Same as Figure 5 but for QBO-W phase.
the polar vortex significantly in December and January while in February and March the positive response is weaker but still marginally significant. These results agree with earlier studies (e.g., Holton & Tan, 1980), which suggest that QBO modifies the propagation of planetary waves so that they are directed more poleward in the easterly QBO phase. Our results also confirm the finding of earlier studies (Lu et al., 2008a; Maliniemi et al., 2016) that the QBO effect on polar vortex is stronger in early winter than in late winter.

The EEP forms ozone-depleting NOx compounds in the thermosphere and upper mesosphere that descend down to the lower mesosphere and upper stratosphere during the winter (Funke et al., 2005). Ozone depletion warms the polar mesosphere and upper stratosphere in midwinter since ozone act as a radiative cooler in polar darkness (Sinnhuber et al., 2018). Earlier studies (Langematz et al., 2003; Baumgaertner et al., 2011; Arsenovic et al., 2016; Salminen et al., 2019) have suggested that ozone depletion and temperature changes in the upper stratosphere strengthen the polar vortex and weaken the Brewer–Dobson circulation leading to a cooling and decreased ozone in the lower polar stratosphere. However, mechanisms for these dynamical responses are still partly unclear and unconfirmed. EEP affects the wintertime stratosphere in the polar region while the varying solar irradiance mainly affects the low- and mid-latitudes during winter, which may explain the larger role of EEP on polar vortex. ENSO is suggested to affect the polar vortex by altering the formation of planetary waves at surface level (Honda et al., 2001; Garfinkel & Hartmann, 2008) or by altering the upwelling in the tropical lower stratosphere (Randel et al., 2009). Volcanic eruptions release aerosols to the stratosphere and, thereby, warm the stratosphere at sunlit latitudes (Robock, 2000). This warming increases the temperature difference between low/middle and polar latitudes, which leads to a stronger polar vortex. However, volcanic eruptions are also suggested to increase planetary wave activity and Brewer–Dobson circulation (Toohey et al., 2014) which would have a negative effect on the polar vortex.

We have also studied here the effect of the solar-related and terrestrial drivers on the polar vortex separately in the two QBO phases. We have shown that the EEP effect on the polar vortex in the easterly QBO phase is stronger and more significant than the effects of solar irradiance, ENSO and volcanic activity, while in the westerly QBO phase the EEP effect is mostly insignificant and does not stand out against the effects of other drivers. This QBO phase relationship agrees with earlier studies, e.g., by Palamaru & Bryant (2004), Maliniemi et al. (2013, 2016) and Salminen et al. (2019). More importantly, our results confirm that the EEP effect and its QBO modulation are not artifacts caused by some other drivers of the polar stratosphere. Salminen et al. (2019) suggested that the QBO modulation could be due to stronger Brewer–Dobson circulation in the QBO-E. Asikainen et al. (2020) showed that the EEP effect on the polar vortex is stronger in the months preceding a sudden stratospheric warming. They suggested that increased planetary wave activity amplifies the dynamical effect of the EEP. Increased planetary wave activity also causes sudden stratospheric warmings and enhances the Brewer–Dobson circulation, leading to a relation with the stronger EEP effect.

We found that the effect of solar irradiance/sunspot number on the polar vortex is not very systematic or significant in either QBO-E or QBO-W phase. Solar irradiance weakens the polar vortex in QBO-W January but this response is only marginally significant. In the polar troposphere SSN weakens the zonal wind in January in QBO-E and in February and March in QBO-W. We found that in March the ENSO weakens the stratospheric polar vortex in QBO-W, but strengthens it in QBO-E. However, these responses are only marginally significant. Garfinkel & Hartmann (2007) found that the negative ENSO effect on polar vortex is weakened in QBO-E phase but not reverse. We found the effect of volcanic activity to be quite different in the two QBO phases. Volcanic activity is associated with a marginally significant strengthening of polar vortex in December and January and weakening in March in QBO-E phase while in QBO-W volcanic activity leads only to a marginally significant positive response in March. A notable lack of significant tropospheric responses was found in QBO-W, contrary to QBO-E. The small number of winters with high volcanic activity (only two major eruptions occurred during the studied time period, both in the QBO-E phase) may partly affect the observed difference in AOD effect between the QBO phases.

It is clear that QBO significantly modulates the effect of the different drivers on the northern polar vortex. QBO is known to affect the planetary wave propagation (Holton & Tan, 1980) and Brewer–Dobson circulation (Flury et al., 2013; Salminen et al., 2019), which are possible reasons for the QBO modulation of the different effects on polar vortex. EEP (e.g., Baumgaertner et al., 2011; Salminen et al., 2019), solar irradiance (Kodera & Kuroda, 2002), ENSO (Honda et al., 2001; Garfinkel & Hartmann, 2008) and volcanic aerosols (Toohey et al., 2014) have been reported to affect the Brewer–Dobson circulation and, thereby, the polar vortex. The influence of planetary wave activity on EEP effect is supported by the recent study of Asikainen et al. (2020) who show that the EEP effect is stronger in times before sudden stratospheric warmings when planetary waves more strongly converge in the polar stratosphere. Increased planetary wave activity and Brewer–Dobson circulation may enhance the effect of EEP, which originates in the polar mesosphere and upper stratosphere, and weaken the effect of the other drivers, like the ENSO, which originates in the low-latitude surface, or the solar irradiance, which affects directly at the low-latitude stratosphere. The detailed mechanisms responsible for the QBO modulation of the effects of these different drivers require still additional studies. We also note that during 1957–2017 the linear model including the explaining variables studied here explains approximately 25–35% of variance in the northern high-latitude wintertime atmosphere, implying that a major part of variance remains unexplained. A part of the unexplained variance is probably due to internal variability, but non-linear effects of the drivers considered here and possibly some other drivers not considered here like, e.g., tropospheric wave activity may affect the polar vortex dynamics and should be examined in future studies.

Acknowledgements. We acknowledge the financial support by the Academy of Finland to the ReSoLVE Center of Excellence (project no. 307411) and to the PROSPECT (project no. 321440), and by the Kvantum institute of the University of Oulu (CAESAR project). V.M. was funded by the Norwegian Research Council under contract 223252/F50. We thank
ECMWF for providing the ERA-40 and ERA-Interim reanalysis datasets (https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets), International service of geomagnetic indices for Ap index (http://isgi.unistra.fr), WDC—SILSO, Royal Observatory of Belgium for the interplanetary sunspot number (http://www.sidc.be/silso/versionarchive), NOAA/ESRL Physical Sciences Division for Nino 3.4 index (https://www.esrl.noaa.gov/psd/data/timeseries/monthly/NINO34/) and NASA for the stratospheric aerosol optical depth data (https://data.giss.nasa.gov/modelforce/strataer/). The editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References


