Solar cycle in current reanalyses: 
(non)linear attribution study

A. Kuchar, P. Sacha, J. Miksovsky, and P. Pisoft

Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University in Prague, V Holesovickach 2, 180 00 Prague 8, Czech Republic

Accepted: 1 November 2014 – Published: 9 December 2014

Correspondence to: A. Kuchar (kuchar@mbox.troja.mff.cuni.cz)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

This study focusses on the variability of temperature, ozone and circulation characteristics in the stratosphere and lower mesosphere with regard to the influence of the 11 year solar cycle. It is based on attribution analysis using multiple nonlinear techniques (Support Vector Regression, Neural Networks) besides the traditional linear approach. The analysis was applied to several current reanalysis datasets for the 1979–2013 period, including MERRA, ERA-Interim and JRA-55, with the aim to compare how this type of data resolves especially the double-peaked solar response in temperature and ozone variables and the consequent changes induced by these anomalies. Equatorial temperature signals in the lower and upper stratosphere were found to be sufficiently robust and in qualitative agreement with previous observational studies. The analysis also pointed to the solar signal in the ozone datasets (i.e. MERRA and ERA-Interim) not being consistent with the observed double-peaked ozone anomaly extracted from satellite measurements. Consequently the results obtained by linear regression were confirmed by the nonlinear approach through all datasets, suggesting that linear regression is a relevant tool to sufficiently resolve the solar signal in the middle atmosphere. Furthermore, the seasonal dependence of the solar response was also discussed, mainly as a source of dynamical causalities in the wave propagation characteristics in the zonal wind and the induced meridional circulation in the winter hemispheres. The hypothetical mechanism of a weaker Brewer Dobson circulation was reviewed together with discussion of polar vortex stability.

1 Introduction

The Sun is a prime driver of various processes in the climate system. From observations of the Sun’s variability on decadal or centennial time scales, it is possible to identify temporal patterns and trends in solar activity, and consequently to derive the related mechanisms of the solar influence on the Earth’s climate (e.g. Gray et al., 2010).
Of the semi-regular solar cycles, the most prominent is the, approximate 11 year periodicity which manifests in the solar magnetic field or through fluctuations of sunspot number, but also in the total solar irradiance (TSI) or solar wind properties. For the dynamics of the middle atmosphere, where ozone production and destruction occurs, the changes in the spectral irradiance are the most influential, since the TSI as the integral over all wavelengths exhibits variations of orders lower than the ultraviolet part of the spectrum (Lean, 2001). This fact was supported by original studies (e.g. Labitzke, 1987; Haigh, 1994) that suggested the solar cycle influence on the variability of the stratosphere. Gray et al. (2009) have shown, with the fixed dynamical heating model, that the response of temperature in the photochemically controlled region of the upper stratosphere is approximately given 60% by direct solar heating and 40% due to indirect effect by the ozone changes.

Numerous observational studies have identified temperature and ozone changes linked to the 11 year cycle by multiple linear regression. The use of ERA-40 reanalysis Frame and Gray (2010) pointed to a manifestation of annually averaged solar signal in temperature, exhibit predominantly around the equator with amplitudes up to 2 K around the stratopause and with a secondary amplitude maximum of up to 1 K in the lower stratosphere. Soukharev and Hood (2006), Hood et al. (2010) and Randel and Wu (2007) have used satellite ozone data sets to characterize statistically significant responses in the upper and lower stratosphere. The observed double-peaked ozone anomaly in the vertical profile around the equator was confirmed by the simulations of coupled chemistry climate models (Austin et al., 2008).

Statistical studies (e.g. Labitzke et al., 2006; Camp and Tung, 2007) have also focussed on the lower stratospheric solar signal in the polar regions and have revealed modulation by the Quasi-biennial oscillation (QBO), or the well known Holton–Tan relationship (Holton and Tan, 1980) modulated by the solar cycle. Proposed mechanisms suggested that the solar signal induced during early winter in the upper equatorial stratosphere propagates poleward and downward when the stratosphere transits from a radiatively controlled state to a dynamically controlled state involving planetary wave.
propagation (Kodera and Kuroda, 2002). The mechanism of the solar cycle and QBO interaction, which stems from reinforcing each other or canceling each other out (Gray et al., 2004) has been verified by recent model simulations (Matthes et al., 2013). These proved the independence of the solar response in the tropical upper stratosphere from the response dependent on the presence of the QBO in lower altitudes.

The ozone and temperature perturbations associated with the solar cycle have an impact on the middle atmospheric circulation. They produce a zonal wind anomaly around the stratopause (faster subtropical jet) during solar maxima through the enhanced meridional temperature gradient. Since planetary wave propagation is affected by the zonal mean flow (Andrews and McIntyre, 1987), we can suppose that a stronger subtropical jet can deflect planetary waves propagating from higher latitudes. Reduced wave forcing can lead to decreasing/increasing and or upwelling/downwelling motions in the equatorial or higher latitudes respectively (Kodera and Kuroda, 2002). The Brewer–Dobson circulation (BDC) is weaker during solar maxima (Gray et al., 2010) although this appears to be sensitive to the state of the polar winter. Observational studies, together with model experiments (e.g. Matthes et al., 2006) suggest a so-called “Top-Down” mechanism where the solar signal is transferred from the upper to lower stratosphere, and even to tropospheric altitudes.

Observational and modeling studies over the past two decades have fundamentally changed our understanding of wave processes and the coupling between the middle atmosphere and tropospheric conditions (Gerber et al., 2012). It has been shown that the stratosphere plays a significant and active role in tropospheric circulation on various time scales (Baldwin and Dunkerton, 1999; Lu et al., 2013; Solomon et al., 2010). A deeper understanding of the mechanisms of communication between the middle atmosphere and troposphere contributes to better climate change predictions. However, a number of questions about the coupling processes with regard to solar signal perturbation have to be answered. It has been shown that difficulties in the state-of-the-art climate models arise when reproducing the solar signal influence on winter polar circulation, especially in less active sun periods (Ineson et al., 2011). The hypothesis is that
solar UV forcing is too weak in the models. Satellite measurements indicate that variations in the solar UV irradiance may be larger than previously thought (Harder et al., 2009).

At the Earth’s surface, the detection of the solar cycle influence is problematic since there are other significant forcing factors, i.e. greenhouse gases, volcanoes and aerosol changes (Gray et al., 2010), as well as substantial variability attributable to internal climate dynamics. However several studies (van Loon et al., 2007; van Loon and Meehl, 2008; Hood and Soukharev, 2012; Hood et al., 2013; Gray et al., 2013) detected the solar signal in sea level pressure or sea surface temperature which supports the hypothesis of a troposphere-ocean response to the solar cycle. The studies (e.g. Hood and Soukharev, 2012) suggest a so-called “Bottom-Up” solar forcing mechanism. That contributes to the lower ozone and temperature anomaly in connection with the lower stratosphere deceleration of the BDC.

Several past studies (e.g. Soukharev and Hood, 2006; Frame and Gray, 2010; Gray et al., 2013) used multiple linear regression to extract the solar signal and separate other climate phenomena like the QBO, the effect of aerosols, NAO, ENSO or trend variability. Apart from this conventional method, it is possible to use alternative approaches to isolate and examine particular signal components, such as wavelet analysis (Pisoft et al., 2012, 2013) or empirical mode decomposition (Coughlin and Tung, 2004). The nonlinear character of the climate system also suggests potential benefits from the application of alternative, full nonlinear attribution techniques to study of properties and interactions in the atmosphere. However, such nonlinear techniques have been used rather sporadically in the atmospheric sciences (e.g. Walter and Schönwiese, 2003; Pasini et al., 2006; Blume and Matthes, 2012), mainly due to their several disadvantages such as the lack of explanatory power (Olden and Jackson, 2002).

To examine middle atmospheric conditions, it is necessary to study reliable and sufficiently vertically resolved data. Systematic and global observations of the middle atmosphere only began during the International Geophysical Year (1957–1958) and were later expanded through the development of satellite measurements (Andrews
and McIntyre, 1987). Supplementary data come from balloon and rocket soundings, though these are limited by their vertical range (only the lower stratosphere in the case of radiosondes) and the fact that the in situ observations measure local profiles only. By assimilation of these irregularly distributed data and discontinuous measurements of particular satellite missions into an atmospheric/climatic model, we have modern basic datasets available for climate research, so called reanalyses. These types of data are relatively long, globally gridded with a vertical range extending to the upper stratosphere or the lower mesosphere and thus suitable for 11 year solar cycle research. In spite of their known limitations (discontinuities in ERA reanalysis McLandress et al., 2013), they are considered an extremely valuable research tool (Rienecker et al., 2011). Coordinated intercomparison has been initiated by the SPARC community to understand current reanalysis products, and to contribute to future reanalysis improvements (Fujiwara et al., 2012).

2 Datasets

Our analysis was applied to the last generation of three reanalysed datasets: MERRA (Modern Era Reanalysis for Research and Applications, developed by NASA) (Rienecker et al., 2011), ERA-Interim (ECMWF Interim Reanalysis) (Dee et al., 2011) and JRA-55 (Japanese 55 year Reanalysis) (Ebita et al., 2011). We have studied the series for the period 1979–2013. All of the datasets were analysed on a monthly basis. The Eliassen–Palm (EP) flux diagnostics (described below) was analysed on the daily basis and subsequently monthly means were produced. The vertical range extends to the lower mesosphere (0.1 hPa) for MERRA, and to 1 hPa for the remaining ones. The horizontal resolution of the gridded datasets was 1.25° × 1.25° for MERRA and JRA-55 and 1.5° × 1.5° for ERA-Interim respectively.

In comparison with previous generations of reanalysis, it is possible to observe a better representation of stratospheric conditions. This improvement is considered to be connected with increasing the height of the upper boundary of the model domain (Rie-
The Brewer–Dobson circulation was markedly overestimated by ERA-40, an improvement was achieved in ERA-Interim, but the upward transport remains faster than observations indicate (Dee et al., 2011). Interim results of JRA-55 suggest a less biased reanalysed temperature in the lower stratosphere relative to JRA-25 (Ebita et al., 2011).

Except for the standard variables provided in reanalysis, i.e. air temperature, ozone mixing ratio and circulation characteristics – zonal, meridional or omega velocity, we have also analysed other dynamical variables. Of particular interest were the EP flux diagnostics – a theoretical framework to study interactions between planetary waves and the zonal mean flow (Andrews and McIntyre, 1987). Furthermore, this framework allows the study of the wave propagation characteristics in the zonal wind and the induced (large scale) meridional circulation as well. For this purpose the quasi-geostrophic approximation of Transformed Eulerian Mean (TEM) equations was used, in the form employed by (Edmon Jr et al., 1980).

3 Methods

To detect variability and changes due to external climate factors, such as the 11 year solar cycle, we have applied an attribution analysis based on Multiple Linear Regression (MLR) and two nonlinear techniques. The regression model separates the effects of climate phenomena that are supposed to have an impact on middle atmospheric conditions. Our regression model of a particular variable \( X \) is described by the following equation:

\[
X(t, z, \phi, \lambda) = \sum_{i=1}^{12} \alpha_i(z, \phi, \lambda) + \beta(z, \phi, \lambda) t + \gamma(z, \phi, \lambda) \text{SOLAR}(t) + \delta_1(z, \phi, \lambda) \text{QBO}_1(t) \\
+ \delta_2(z, \phi, \lambda) \text{QBO}_2(t) + \delta_3(z, \phi, \lambda) \text{QBO}_3(t) + \epsilon(z, \phi, \lambda) \text{ENSO}(t) \\
+ \zeta(z, \phi, \lambda) \text{SAOD}(t) + \eta(z, \phi, \lambda) \text{NAO}(t) + \epsilon(t, z, \phi, \lambda). \tag{1}
\]
After deseasonalizing which can be represented by $\alpha_i$ indices we have applied a trend regressor $t$ either in linear form or including the Equivalent Effective Stratospheric Chlorine (EESC) index (this should be employed due to the ozone turnover trend around the middle of the 90s). The solar cycle is represented by the 10.7 cm radio flux as a proxy which correlates well with sunspot number variation (the data were acquired from Dominion Radio Astrophysical Observatory (DRAO) in Penticton, Canada).

We have included the quasi-biennial proxies as another stratosphere-related predictor. Similar studies have represented the QBO in multiple regression methods in several ways. Our approach involves three separate QBO indices extracted from the MERRA reanalysis. These three indices are the first three principal components of the residuals of our linear regression model Eq. (1) excluding QBO predictors applied to the equatorial zonal wind. The approach follows the paper by Frame and Gray (2010), or the study by Crooks and Gray (2005). The three principal components explain 49, 47 and 3% of the total variance. The extraction of the first two components reveals a 28 month periodicity and an out-of-phase relationship between the upper and lower stratosphere. The out-of-phase relationship or orthogonality manifests approximately in a quarter period shift of these components. The deviation from the QBO quasi-regular period represented by the first two dominant components is contained in the residual variance of 4%. Linear regression analysis of the zonal wind with the inclusion of the first two principal components reveals a statistically significant linkage between the third principal component and the residuals of this analysis. Furthermore, the regression coefficient of this QBO proxy was statistically significant for all variables tested for a $p$ value < 0.05 (see below for details about statistical significance techniques). Wavelet analysis demonstrates three statistically significant but non-stationary periods exceeding the level of the white noise wavelet spectrum (not shown): an approximate annual cycle (a peak period of 1 year and 2 months), a cycle with a peak period of 3 years and 3 months and a long-period cycle (a peak period between 10 and 15 years). Those interferences can be attributed to the possible non-linear interactions between...
the QBO itself and other signals like the annual cycle or long-period cycle such as the
11 year solar cycle at the equatorial stratosphere.

The El Niño Southern Oscillation (ENSO) is represented by the Multivariate ENSO
index (MEI) which is computed as the first principal component of the six main ob-
erved variables over the Pacific Ocean: sea level pressure, zonal and meridional wind,
sea surface temperature, surface air temperature and total cloudiness fraction of the
sky (NCAR, 2013). The effect of volcanic eruptions is represented by the Stratospheric
Aerosol Optical Depth (SAOD). The time series was derived from the optical extinction
data (Sato et al., 1993). We have used globally averaged time series in our regression
model. The North Atlantic Oscillation (NAO) has also been included in the respective in-
dex derived by a rotated principal component analysis technique applied to the monthly
standardized 500-hPa height anomalies obtained from the Climate Data Assimilation
System (CDAS) in the Atlantic region between 20–90° N (NWS, 2013).

The multiple regression model via Eq. (1) has been used for the attribution analysis,
and supplemented by two nonlinear techniques. The MLR coefficients were estimated
by the least squares method. To avoid the effect of autocorrelation of residuals and
to obtain the Best Linear Unbiased Estimate (BLUE) according to the Gauss–Markov
theorem (Thejll, 2005), we have used an iterative algorithm to model the residuals as
a second-order autoregressive process. The Durbin–Watson statistic has been used
to detect the autocorrelation of the error terms from the regression model. As a result
of the uncorrelated residuals, we can suppose the SDs of the estimated regression
coefficients not to be diminished (Neter et al., 2004). The statistical significance of
the regression coefficients was computed with a $t$ test and verified by a bootstrap
significance test.

The nonlinear approach, in our case, consisted of Multi Layer Perceptron (MLP) and
the relatively novel Support Vector Regression (SVR) technique in our case. The MLP
as a technique inspired by the human brain is highly complex and capable of capturing
non-linear interactions between inputs (regressors) and output (modelled data) (e.g.
Haykin, 2009). The nonlinear approach is achieved by transferring the input signals
through a sigmoid function in a particular neuron and within a hidden layer propagating to the output (a so called feedforward propagation). The standard error backpropagation iterative algorithm to minimize the global error has been used.

The Support Vector Regression technique belongs to the category of kernel methods. Input variables were nonlinearly transformed to a high-dimensional space by a radial basis (Gaussian) kernel, where a linear classification (regression) can be constructed (Cortes and Vapnik, 1995). However, cross-validation must be used to establish a kernel parameter and cost function. We have used 5-fold cross-validation to optimize the SVR model selection for every point in the dataset as a trade-off between the recommended number of folds (Kohavi et al., 1995) and computational time. The MLP model was validated by the holdout method since this method is more expensive in order of magnitude compared to computational time. The datasets were separated into a training set (75% of the whole dataset) and a testing set (25% of the whole dataset).

The earlier mentioned lack of explanatory power of the nonlinear techniques mainly comes from nonlinear interactions during signal propagation and the impossibility to directly monitor the influence of the input variables. In contrast to the linear regression approach, the understanding of relationships between variables is quite problematic. For this reason, the responses of our variables have been modelled by a technique originating from sensitivity analysis studies and also used by e.g. Blume and Matthes (2012). The relative impact $RI$ of each variable was computed as

$$RI = \frac{l_k}{\sum l_k},$$

where $l_k = \sigma(\hat{y} - \hat{y}_k)$. $\sigma(\hat{y} - \hat{y}_k)$ is a difference between the original model output $\hat{y}$ and the model output $\hat{y}_k$ when the $k$-input variable was held at its constant level. There are many possibilities with regard to which constant level to choose. It is possible to choose several levels and then to observe the sensitivity of model outputs varying for example on minimum, median and maximum levels. Our sensitivity measure (relative impact)
was based on the median level. The primary reason comes from pure practical considerations – to compute our results fast enough as another weakness of the nonlinear techniques lies in the larger requirement of computational capacity. In general, this approach was chosen because of their relative simplicity for comparing all techniques to each other and to be able to interpret them too. The contribution of variables in neural network models has already been studied and Gevrey et al. (2003) produced a review and comparison of these methods.

4 Results

4.1 Annual response (MERRA)

Figure 1 shows the annually averaged solar signal in the zonal and altitudinal means of temperature, zonal wind, geopotential height and ozone mixing ratio. The signal is expressed as the average difference between the solar maxima and minima in the period 1979–2013. Statistically significant responses detected by the linear regression in the temperature series (see Fig. 1a) are positive and are located around the equator in the lower stratosphere with values of about 0.5°C. The temperature response increases to 1°C in the upper stratosphere at the equator and up to 2°C at the poles. The significant solar signal anomalies are more variable around the stratopause and not limited to the equatorial regions. Hemispheric asymmetry of the statistical significance can be observed in the lower mesosphere. From a relative impact point of view (in Fig. 1b–d marked as RI), it is difficult to detect a signal with an impact larger than 20% in the lower stratosphere where the volcanic and QBO impacts dominate. In the upper layers (where the solar signal expressed by the regression coefficient is continuous across the equator) we have detected relatively isolated signals (over 20%) around ±15° using the relative impact method. The hemispheric asymmetry also manifests in the relative impact field, especially in the SVR field in the mesosphere.
The annually averaged solar signal in the zonal-mean of zonal wind (Fig. 1e–h) dominates around the stratopause as an enhanced subtropical westerly jet. The zonal wind variability due to the solar cycle corresponds with the temperature variability due to the change of the meridional temperature gradient and via the thermal wind equation. The largest positive anomaly in the Northern Hemisphere reaches 4 m s$^{-1}$ around 60 km (Fig. 1e). In the Southern Hemisphere, the anomaly is smaller and not statistically significant. There is a significant negative signal in the southern polar region and also at the equator especially in the mesosphere. The negative anomalies correspond with a weakening of the westerlies or an amplification of the easterlies. The relative impact of the solar cycle is similarly located zonally even for both nonlinear techniques (Fig. 1f–h). The equatorial region across all the stratospheric layers is dominantly influenced by the QBO (expressed by all 3 QBO regressors) and for this reason the solar impact is minimized around the equator.

The pattern of the solar response in geopotential height (Fig. 1i–l) shows positive values in the upper stratosphere and lower mesosphere. This is also consistent with the zonal wind field thorough thermal wind balance. In the geopotential field, the solar cycle influences the most extensive area among all regressors. The impact area includes almost the whole mesosphere and the upper stratosphere.

The last row of Fig. 1 also shows the annual mean solar signal in the zonal mean of the ozone mixing ratio (expressed as a percent change from the solar maximum to the solar minimum). Using the model with EESC instead of a linear trend over the whole period, we tried to capture the ozone trend change around the year 1996. Another possibility was to use our model over two individual periods, e.g. 1979–1995 and 1996–2013, but the results were quantitatively similar. The main common feature of other results is the positive ozone response in the lower stratosphere, ranging from a 1 to 3 % change. The majority of results share the positive ozone response. In the equatorial upper stratosphere, no other relevant solar signal was detected compared to the study based on satellite measurement (Soukharev and Hood, 2006). By the relative impact method (Fig. 1n–p), we have obtained results comparable with linear regression.
coefficients, but especially around the stratopause the impact suggested by nonlinear techniques does not reach the values achieved by linear regression.

4.1.1 Annual response (comparison with JRA-55, ERA-Interim)

Comparison of the results for the MERRA, ERA-Interim and JRA-55 temperature, zonal wind and geopotential height shows that the annual responses to the solar signal are in qualitative agreement (compare Figs. 1, 2 and 3). The zonal wind and geopotential response seems to be consistent in all presented methods and datasets. The largest discrepancies can be seen in the upper stratosphere and especially in the temperature field (the first row in these figures). The upper stratospheric equatorial anomaly was not detected by any of the regression techniques in the case of the JRA-55 reanalysis. On the other hand, the anomaly in the ERA-Interim temperature in Fig. 2 almost reaches the same value as in the MERRA series.

The variability of the solar signal in the MERRA stratospheric ozone series was compared with the ERA-Interim results. The analysis points to large differences in the ozone response to the solar cycle between the reanalyses and even in comparison with satellite measurements by Soukharev and Hood (2006). In comparison with the satellite measurements, no relevant solar signal was detected in the upper stratosphere in the MERRA series. The signal seems to be shifted above the stratopause (confirmed by all techniques, shown in Figs. 1 and 2m–p). Regarding the ERA-Interim, there is an ozone response to the solar cycle in the upper stratosphere. This statistically significant response indicates negative anomalies with values reaching up to 2% above the equator and up to 5% in the polar regions of both hemispheres. The negative response can be connected with a higher destruction of ozone during the solar maximum period and consequent heating of the region. The lower stratospheric solar signal in the ERA-interim is not limited to the equatorial belt ±30° up to 20 hPa, as in the case of the MERRA reanalysis, and the statistical significance of this signal is rather reduced. The solar signal is detected higher and extends from the subtropical areas to the polar regions. The results suggest that the solar response in the MERRA series is more similar
to the results from satellite measurements (Soukharev and Hood, 2006). Nevertheless, further comparison with independent data sets is needed to assess the data quality in detail.

4.1.2 Comparison of the linear and nonlinear approaches (MLR vs. SVR and MLP)

In this paper, we have applied and compared one linear (MLR) and two nonlinear attribution (SVR and MLP) techniques. The response of the studied variables to the solar signal and other forcings was studied using the sensitivity analysis approach (Blume and Matthes, 2012). This approach does not recognize a positive or negative response as the linear regression does. For this reason, the relative impact results are compared to the regression’s coefficients. Using linear regression, it would be possible to analyse the statistical significance of the regression’s coefficients and a particular level of the relative impact. Due to a higher variance, the significance levels of the relative impact are not estimated. A comparison between the linear and nonlinear approaches by the relative impact fields shows qualitative and in most regions also quantitative agreement. The most pronounced agreement is observed in the zonal wind (Figs. 1, 2 and 3f–h) and geopotential height fields (Figs. 1, 2 and 3j–l). On the other hand the worst agreement is captured in the ozone field where nonlinear techniques have a problem identifying the upper stratospheric ozone anomaly detected by linear regression, although the lower stratospheric ozone anomaly is represented similarly by all techniques. In the temperature field the upper stratospheric solar signal reaches values over 20 %, some individual signals in the Northern Hemisphere even reach 40 %. However, using the relative impact approach, the lower stratospheric solar signal in the temperature field (which is well established by the regression coefficient) does not even reach 20 % because of the dominance the QBO and volcanic effects. These facts emphasize that nonlinear techniques contribute to the robustness of attribution analysis since the linear regression results were plausibly confirmed by the SVR and MLP techniques.
However, the statistical significance of individual responses could have been estimated by the bootstrap technique, which is quite expensive for computational time, and for this reason was not applied. The comparison of various statistical approaches (MLR, SVR and MLP) should actually contribute to the robustness of the attribution analysis including the statistically assessed uncertainties. These uncertainties could partially stem from the fact that the SVR and Neural network techniques are dependent on an optimal model setting which is based on a rigorous cross-validation process, which places a high demand on computing time.

The major differences between the techniques can be seen in how they can simulate the original time series, i.e. coefficient of determination. For instance, the differences of the explained variance reach up to 10% between linear and nonlinear techniques, although the structure of the coefficient of determination is almost the same. To conclude, nonlinear techniques show an ability to simulate the middle atmosphere variability with a higher accuracy than cross-validated linear regression.

4.2 Monthly response (MERRA)

As was pointed out by Frame and Gray (2010), it is necessary to examine the solar signal in individual months because of variable solar impact throughout the year. For example, the amplitude of the lower stratospheric solar signal in the northern polar latitudes in February exceeds the annual response since the solar cycle influence on vortex stability is most pronounced in February. Besides the radiative influences of the solar cycle, we discuss the dynamical response throughout the polar winter (Kodera and Kuroda, 2002).

Statistically significant upper stratospheric equatorial anomalies in the temperature series (winter months in Figs. 4 and 5a–d) are expressed in almost all months. Their amplitude and statistical significance vary throughout the year. The variation between the solar maxima and minima could be up to 1°C in some months. Outside the equatorial regions, the fluctuation could reach several °C. The lower stratospheric equatorial anomaly strengthens during winter. This could be an indication of dynamical changes,
i.e. alternation of the residual circulation between the equator and polar regions (for details, please see the discussion). Aside from the radiative forcing by direct or ozone heating, other factors are linked to the anomalies in the upper levels of the middle atmosphere (Haigh, 1994; Gray et al., 2009). It is necessary to take into consideration the dynamical coupling with the mesosphere through changes of the residual circulation (see the below dynamical effects discussion). That can be illustrated by the positive anomaly around the stratopause in February (up to 4°C around 0.5 hPa). This anomaly propagates downward and, together with spring radiative forcing, affects the stability of the equatorial stratopause. Hemispheric asymmetry in the temperature response above the stratopause probably originates from the hemispheric differences, i.e. different wave activity. These statistically significant and positive temperature anomalies across the subtropical stratopause begin to descend and move to higher latitudes in the beginning of the northern winter. The anomalies manifest fully in February in the region between 60–90° N and below 10 hPa and reach tropospheric levels – contrary to the results for the Southern Hemisphere. The southern hemispheric temperature anomaly is persistent above the stratopause and the solar cycle influence on the vortex stability differs from those in the Northern Hemisphere.

The above described monthly anomalies of temperature correspond with the zonal wind anomalies throughout the year (Figs. 4 and 5e–h). The strengthening of the subtropical jets around the stratopause is most apparent during the winter in both hemispheres. This positive zonal wind anomaly gradually descends and moves poleward similar to Frame and Gray (2010) analysis based on ERA-40 data. In February, the intensive stratospheric warming and mesospheric cooling is associated with a more pronounced transition from winter to summer circulation attributed to the solar cycle (in relative impact methodology up to 30 %). In the Southern Hemisphere, this poleward motion of the positive zonal wind anomaly halts approximately at 60° S. For example in August, we can observe a well-marked latitudinal zonal wind gradient (Fig. 4g). Positive anomalies in the geopotential height field correspond with the easterly zonal wind anomalies. The polar circulation reversal is associated with intrusion of ozone from the
lower latitudes as it is apparent, e.g., in August in the Southern Hemisphere and in February in the Northern Hemisphere (last rows of Figs. 4 and 5).

When comparing the results of the MERRA and ERA-40 series studied by Frame and Gray (2010), distinct differences were found (Fig. 4e–f) in the equatorial region of the lower mesosphere in October and November, while, in the MERRA reanalysis we have detected an easterly anomaly above 1 hPa in both months, a westerly anomaly was identified in the ERA-40 series. Further distinct differences in the zonal mean temperature and zonal wind anomalies were not found.

5 Dynamical effects discussion

In this section, we discuss the dynamical impact of the solar cycle and its influence on middle atmospheric winter conditions. Linear regression was applied to the EP diagnostics. Kodera and Kuroda (2002) suggested that the solar signal produced in the upper stratosphere region is transmitted to the lower stratosphere through the modulation of the internal mode of variation in the polar night jet and through a change in the Brewer–Dobson circulation (prominent in the equatorial region in the lower stratosphere). In our analysis, we discussed the evolution of the winter circulation with an emphasis on the vortex itself rather than the behavior of the jets. Further, we try to deduce the possible processes leading to the observed differences in the quantities of state between the solar maximum and minimum period. Because the superposition principle only holds for linear processes, it is impossible to deduce the dynamics merely from the fields of differences. As noted by Kodera and Kuroda (2002), the dynamical response of the winter stratosphere includes highly nonlinear processes, e.g. wave mean flow interactions. Thus, both the anomaly and the total fields, including climatology, must be taken into account.

We start the analysis of solar maximum dynamics with the period of the northern hemispheric winter circulation formation. The anomalies of the ozone, temperature, geopotential and Eliassen–Palm flux divergence support the hypothesis of weaker BDC
during the solar maximum due to the less intensive wave pumping. This is consistent with previous studies (Kodera and Kuroda, 2002; Matthes et al., 2006). The causality is unclear, but the effect is visible in both branches of BDC as is explained by Fig. 4 and summarized schematically in Fig. 6.

During the early NH winter (including November) when westerlies develop in the stratosphere, we can observe a deeper polar vortex and consequent stronger westerly winds both inside and outside the vortex. However, only the westerly anomaly outside the polar region and around 30° N from 10 hPa to the lower mesosphere is statistically significant (see the evolution of zonal wind anomalies in Fig. 4e–h). The slightly different wind field has a direct influence on the vertical propagation of planetary waves. From the Eliassen–Palm flux anomalies and climatology we can see that the waves propagate vertically with increasing poleward instead of equatorward meridional direction with height. This is then reflected in the EP flux divergence field, where the region of maximal convergence is shifted poleward and the anomalous convergence region emerges inside the vortex above approximately 50 hPa (Fig. 4i–l).

The poleward shift of the maximum convergence area further contributes to the reduced BDC. This is again confirmed by the temperature and ozone anomalies. The anomalous convergence inside the vortex induces anomalous residual circulation, the manifestation of which is clearly seen in the quadrupole-like temperature structure (positive and negative anomalies are depicted schematically in Fig. 6 using red and blue boxes respectively). This pattern emerges in November and even more clearly in December. In December, the induced residual circulation leads to an intrusion of the ozone rich air into the vortex at about the 1 hPa level (Fig. 4o). The inhomogeneity in the vertical structure of the vortex is then also pronounced in the geopotential height differences. This corresponds with the temperature analysis in the sense that above and in the region of the colder anomaly there is a negative geopotential anomaly and vice versa. The geopotential height difference has a direct influence on the zonal wind field (via the thermal wind balance). The result is a deceleration of the upper vortex
parts and consequent broadening of the upper parts (due to the conservation of angular momentum).

Considering the zonal wind field, the vortex enters January approximately with its average climatological extent. The wind speeds in its upper parts are slightly higher. This is because of the smaller geopotential values corresponding to the negative temperature anomalies above approximately 1 hPa. This results from the absence of adiabatic heating due to the suppressed BDC, although the differences in the quantities of state (temperature and geopotential height) are small and insignificant (see the temperature anomalies in Fig. 4c). It is important to note that these differences change sign around an altitude of 40 km inside the vortex further accentuating the vertical inhomogeneity of the vortex. This might start balancing processes inside the vortex, which is confirmed by analysis of the dynamical quantities, i.e. EP flux and its divergence (Fig. 4k). A detailed description of these processes is the key to understanding the dynamics and causality of Sudden Stratospheric Warmings (SSWs) taking place in February.

Significant anomalies of the EP flux indicate anomalous vertical wave propagation resulting in the strong anomalous EP flux convergence being significantly pronounced in a horizontally broad region and confined to upper levels (convergence (negative values) drawn by green or blue shades in Fig. 4i–l). This leads to the induction of an anomalous residual circulation starting to gain intensity in January. The situation then results in the disruption of the polar vortex visible in significant anomalies in the quantities of state in February – in contrast to January. Further strong mixing of air is suggested by the ozone fields. The quadrupole-like structure of the temperature is visible across the whole NH middle atmosphere in February (indicated in the lower diagram of Fig. 6), especially in the higher latitudes. This is very significant and well pronounced by the stratospheric warming and mesospheric cooling.

The hemispheric asymmetry of the solar cycle influence can be especially documented in winter conditions as was already suggested in Sect. 4.2. Since the positive zonal wind anomaly halts at approximately 60° S and intensifies over 10 m s⁻¹, one would expect the poleward deflection of the planetary wave propagation to be accord-
ing to NH winter mechanisms discussed above. This is actually observed from June to August when the highest negative anomalies of the latitudinal coordinates of EP flux are located in the upper stratosphere and in the lower mesosphere (Fig. 5j–l). The anomalous divergence of EP flux develops around the stratopause between 30 and 60° S. Like the hypothetical mechanism of weaker BDC described above, we can observe less wave pumping in the stratosphere and consequently assume less upwelling in the equatorial region. However, the anomalies of the residual circulation pointing to a weaker BDC are not so well established as in the case of the NH winter. These mechanisms could lead to an explanation for the more pronounced temperature response to the solar signal in the equatorial region of the lower stratosphere in August for the SH winter (above 1°C) than in December for the NH winter (around 0.5°C). This is in agreement with another observational study (van Loon and Labitzke, 2000). Overall, the lower stratospheric temperature anomaly is more coherent for the SH winter than for the NH winter, where the solar signal is not so well apparent or statistically significant in particular months and reanalysis datasets.

6 Conclusions

We have analysed the changes of air temperature, ozone and circulation characteristics driven by the variability of the 11 year solar cycle’s influence on the stratosphere and lower mesosphere. Attribution analysis was performed on the last generation of reanalysed data, and aimed to compare how these types of datasets resolve the solar variability throughout the levels where the “top-down” mechanism is assumed. Furthermore, the results originated in linear attribution using MLR were compared with other relevant observational studies and supported by nonlinear attribution analysis using SVR and MLP techniques.

The solar signal extracted from the temperature field from MERRA and ERA-Interim reanalysis using linear regression has the amplitudes around 1°C and 0.5°C, in the upper stratospheric and in the lower stratospheric equatorial region, respectively. These
signals, statistically significant at a $p$ value $< 0.01$, can be considered sufficiently robust and they are in qualitative agreement with previous observational studies (e.g. Frame and Gray, 2010) since we have used the last generation of reanalysed datasets extended to 2013. The statistically significant signal was only observed in the lower part of the stratosphere in the JRA-55 reanalysis, however with similar amplitudes as the other datasets.

Similar to the temperature response, the double-peaked solar response in ozone was detected in satellite measurements (e.g. Soukharev and Hood, 2006) and even confirmed by the coupled chemistry climate model simulations (e.g. Austin et al., 2008). However, the exact position and amplitude of both ozone anomalies remain a point of disagreement between models and observations. The results of our attribution analysis point to large differences in the upper stratospheric ozone response to the solar cycle in comparison with the studies mentioned above and even between reanalyses themselves. The upper stratospheric ozone anomaly reaches 2% in the SBUV(/2) satellite measurements (e.g. Soukharev and Hood, 2006, Fig. 5) which were assimilated as the only source of ozone profiles in MERRA reanalysis. This fact is remarkable since the same signal was not detected in the upper stratosphere in the MERRA results. However, the solar signal in the ozone field seems to be shifted above the stratopause where similar and statistically significant solar variability was attributed. Concerning the solar signal in the ERA-Interim, there is a negative ozone response via a regression coefficient in the upper stratosphere although the solar variability expressed as relative impact appears to be in agreement with satellite measurements. Furthermore, the lower stratospheric solar response in the ERA-Interim’s ozone around the equator is reduced in this dataset and shifted to higher latitudes. Another difference was detected in the monthly response of the zonal wind in October and November in the equatorial region of the lower mesosphere between the results for the MERRA series and ERA-40 data studied by Frame and Gray (2010). While in the MERRA reanalysis we have detected an easterly anomaly, a westerly anomaly was identified in the ERA-40 series.
A similar problem with the correct resolving of the double-peaked ozone anomaly was registered in the study of Dhomse et al. (2011) which investigated in solar response in the tropical stratospheric ozone using a 3-D chemical transport model. The upper stratospheric solar signal observed in SBUV/SAGE and SAGE-based data could only be reproduced in model runs with unrealistic dynamics, i.e. with no inter-annual meteorological changes.

The nonlinear approach to attribution analysis, represented by the application of the SVR and MLP, largely confirmed the solar response computed by linear regression. Consequently, these results can be considered quite robust regarding the statistical modeling of the solar variability in the middle atmosphere. This finding indicates that linear regression is a sufficient technique to resolve the basic shape of the solar signal through the middle atmosphere. However, some uncertainties could partially stem from the fact that the SVR and MLP techniques are highly dependent on an optimal model setting that requires a rigorous cross-validation process (which places a high demand on computing time). As a benefit, nonlinear techniques show an ability simulate the middle atmosphere variability with higher accuracy than linear regression.

In the dynamical effects discussion, we described the dynamical impact of the solar cycle on middle atmospheric winter conditions. The main part deals with the solar influence on northern winter conditions, nevertheless, southern winter anomalies were also discussed. The relevant dynamical effects are summarized in schematic diagrams (Fig. 6). Both diagrams depict average conditions and anomalies induced by the solar cycle. The first one summarizes how equatorward wave propagation is influenced by the westerly anomaly around the subtropical stratopause. The quadrupole-like temperature structure is explained by anomalous residual circulation in the higher latitudes together with the anomalous branch heading towards the equatorial region already hypothesized by Kodera and Kuroda (2002). The second diagram concludes the transition time to vortex disruption during February. Again, a very apparent quadrupole-like temperature structure is even more pronounced, especially in the polar region and seems to be more extended to lower latitudes. However, we can strongly assume that the dy-
Dynamical effects are not zonally uniform, as it is supposed and presented here using two-dimensional (2-D) EP diagnostics and TEM equations. So it would be desirable to extend the discussion of dynamical effects for other relevant characteristics, for example, for the analysis of wave propagation and wave-mean flow interaction using the 3-D formulation (Kinoshita and Sato, 2013).

This paper is fully focused on the solar cycle influence, i.e. on decadal changes in the stratosphere and lower mesosphere, although a huge amount of results concerning other forcings was generated by attribution analysis. The QBO phenomenon could be one of them since the solar-QBO interaction and the modulation of Holton–Tan relationship by the solar cycle are regarded as highly challenging, especially in global climate simulations (Matthes et al., 2013).

Acknowledgements. The authors would like to thank to the relevant working teams for the reanalysis datasets: MERRA (obtained from NASA, http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldings.pl), ERA-Interim (obtained from ECMWF, http://apps.ecmwf.int/datasets/) and JRA-55 (obtained from http://jra.kishou.go.jp/JRA-55/index_en.html). We would also like to express our gratitude to C. A. Svoboda (Foreign Language Studies, Faculty of Mathematics and Physics, Charles University in Prague) for the proofreading of our paper. The study was supported by the Charles University in Prague, Grant Agency project No. 1474314, and by the grant No. SVV-2014-26096.

References


Figure 1. The annually averaged response of the solar signal in the MERRA zonal-mean temperature $t$ (a–d), unit: [°C]; zonal wind $u$ (e–h), unit: [ms$^{-1}$]; geopotential height $h$ (i–l), unit: [m]; and ozone mixing ratio $o_3$ (m–p), unit: percentage change per annual mean. The response is expressed as a regression coefficient RC (corresponding units per $S_{\text{max}}$ minus $S_{\text{min}}$) in the left column and relative impact $RI$ approach in the remaining columns. The relative impact was modeled by MLR, SVR and MLP techniques. The black contour levels in the $RI$ plots are 0.2, 0.4, 0.8 and 1.0. The statistical significance of the scalar fields was computed by a $t$ test. Hatches indicate $p$ values < 0.05.
Figure 2. The annually averaged response of the solar signal in the ERA-Interim zonal-mean temperature $t$ (a–d), unit: [$^\circ$C]; zonal wind $u$ (e–h), unit: [m s$^{-1}$]; geopotential height $h$ (i–l), unit: [m]; and ozone mixing ratio $o_3$ (m–p), unit: percentage change per annual mean. The response is expressed as a regression coefficient RC (corresponding units per $S_{\text{max}}$ minus $S_{\text{min}}$) in the left column and relative impact RI approach in the remaining columns. The relative impact was modeled by MLR, SVR and MLP techniques. The black contour levels in the $RI$ plots are 0.2, 0.4, 0.8 and 1.0. The statistical significance of the scalar fields was computed by a $t$ test. Hatches indicate $p$ values < 0.05.
Figure 3. The annually averaged response of the solar signal in the JRA-55 zonal-mean temperature $t$ (a–d), unit: [$^\circ$C]; zonal wind $u$ (e–h), unit: [ms$^{-1}$]; geopotential height $h$ (i–l), unit: [m]; and ozone mixing ratio $\text{O}_3$ (m–p), unit: percentage change per annual mean. The response is expressed as a regression coefficient RC (corresponding units per $S_{\text{max}}$ minus $S_{\text{min}}$) in the left column and relative impact $RI$ approach in the remaining columns. The relative impact was modeled by MLR, SVR and MLP techniques. The black contour levels in the $RI$ plots are 0.2, 0.4, 0.8 and 1.0. The statistical significance of scalar fields was computed by a $t$ test. Hatches indicate $p$ values $< 0.05$. 
Figure 4. The monthly averaged response of the solar signal in the MERRA zonal-mean temperature $t$ (a–d), unit: [°C]; zonal wind $u$ (e–h), unit: [m s$^{-1}$]; EP flux divergence EPfD (i–l), unit: [m s$^{-1}$ day$^{-1}$]; together with EP flux vectors scaled by the inverse of the pressure, unit: [kg s$^{-2}$]; and ozone mixing ratio, unit: percentage change per monthly mean; with residual circulation $o3 + rc$ (m–p), units: [m s$^{-1}$ m s$^{-1}$] during northern hemispheric winter. The response is expressed as a regression coefficient (corresponding units per $S_{max}$ minus $S_{min}$). The statistical significance of the scalar fields was computed by a $t$ test. Hatches in (a–h) and grey contours in (i–p) indicate $p$ values of $< 0.05$ respectively.
Figure 5. The monthly averaged response of the solar signal in the MERRA zonal-mean temperature $t$ (a–d), unit: [°C]; zonal wind $u$ (e–h), unit: [m s$^{-1}$]; EP flux divergence EPfD (i–l), unit: [m s$^{-1}$ day$^{-1}$]; together with EP flux vectors scaled by the inverse of the pressure, unit: [kg s$^{-2}$]; and ozone mixing ratio, unit: percentage change per monthly mean; with residual circulation $o3 + rc$ (m–p), units: [m s$^{-1}$ Pa s$^{-1}$] during southern hemispheric winter. The response is expressed as a regression coefficient (corresponding units per $S_{\text{max}}$ minus $S_{\text{min}}$). The statistical significance of the scalar fields was computed by a $t$ test. Hatches in (a–h) and grey contours in (i–p) indicate $p$ values of < 0.05 respectively.
Figure 6. Solar cycle modulation of the winter circulation: schema of the related mechanisms. The upper and lower figure show early and later winter respectively. The heating and cooling anomalies are drawn with red and blue boxes. The EP flux divergence and convergence are drawn with green and yellow boxes. The wave propagation anomaly is expressed as a wavy red arrow in contrast to the climatological average drawn by a wavy grey arrow. The induced residual circulation according to the quasi-geostrophic approximation is highlighted by the bold black lines.