



Understanding and forecasting polar stratospheric variability with statistical models

C. Blume^{1,2,*} and K. Matthes^{1,2,*}

¹Helmholtz Centre Potsdam, German Research Centre for Geosciences (GFZ), Potsdam, Germany

²Institute for Meteorology, Free University of Berlin (FUB), Berlin, Germany

* now at: Helmholtz Centre for Ocean Research Kiel (GEOMAR), Kiel, Germany

Correspondence to: C. Blume (christian.blume@met.fu-berlin.de)

Received: 27 December 2011 – Published in Atmos. Chem. Phys. Discuss.: 22 February 2012

Revised: 27 May 2012 – Accepted: 6 June 2012 – Published: 2 July 2012

Abstract. The variability of the north-polar stratospheric vortex is a prominent aspect of the middle atmosphere. This work investigates a wide class of statistical models with respect to their ability to model geopotential and temperature anomalies, representing variability in the polar stratosphere. Four partly nonstationary, nonlinear models are assessed: linear discriminant analysis (LDA); a cluster method based on finite elements (FEM-VARX); a neural network, namely the multi-layer perceptron (MLP); and support vector regression (SVR). These methods model time series by incorporating all significant external factors simultaneously, including ENSO, QBO, the solar cycle, volcanoes, to then quantify their statistical importance. We show that variability in reanalysis data from 1980 to 2005 is successfully modeled. The period from 2005 to 2011 can be hindcasted to a certain extent, where MLP performs significantly better than the remaining models. However, variability remains that cannot be statistically hindcasted within the current framework, such as the unexpected major warming in January 2009. Finally, the statistical model with the best generalization performance is used to predict a winter 2011/12 with warm and weak vortex conditions. A vortex breakdown is predicted for late January, early February 2012.

1 Introduction

The variability of the north-polar stratospheric vortex (Baldwin and Holton, 1988) is a key dynamical feature of the middle atmosphere. Every two years on average it breaks down during winter, resulting in a sudden stratospheric warming (Labitzke and Naujokat, 2000) with greatly increasing

temperatures within a few days. A large number of these extreme warming events in the stratosphere have been found to descend downward to the troposphere, influencing the weather (Baldwin and Dunkerton, 2001).

There are external variability factors that influence the dynamics of the polar stratosphere. These factors include the Quasi-Biennial Oscillation (QBO) (e.g. Baldwin et al., 2001), the El Niño-Southern Oscillation (ENSO) (e.g. Manzini et al., 2006), the 11-yr solar cycle (e.g. Gray et al., 2010), and high impact volcanic eruptions (Robock, 2000). These external factors (or forcings) also interact with each other, resulting in a complex and nonlinear dynamical response (e.g. Calvo et al., 2009; Richter et al., 2011).

Previous efforts investigated the impact of these factors. Holton and Tan (1980, 1982) showed that the QBO east phase leads to a generally warmer, more disturbed polar vortex and vice versa for QBO west. This so-called Holton-Tan relationship was later shown to be present during solar minimum but significantly weaker during solar maximum (e.g. Labitzke, 1987; Labitzke and van Loon, 1988). They showed that sudden stratospheric warmings are most likely to happen during solar maximum (minimum) and QBO west (east) phase. Accordingly, a study by Camp and Tung (2007a) found the least-perturbed vortex state to take place during solar minimum and QBO west conditions. Recent studies have shown that positive ENSO phases (El Niño) lead to a more disturbed polar vortex, as opposed to negative ENSO phases (La Niña) where the vortex is less disturbed (Camp and Tung, 2007b; Mitchell et al., 2011). The least understood forcing factor is aerosols injected into the stratosphere by very strong but rare volcanic eruptions (Robock, 2000), leading to

nonlinear feedbacks with other forcings (e.g. Garfinkel and Hartmann, 2007).

When making polar stratospheric forecasts, general circulation model runs consisting of multiple observation constrained ensemble members are performed. These forecasts are reliable on a daily scale but on a seasonal scale they quickly become computationally very expensive and lose their forecast skill (Gerber et al., 2009; Kuroda, 2010). This work investigates statistical models that are mathematically much simpler and demand significantly less computer power, and even though they do not simulate physical processes explicitly, one can learn about underlying relationships. A wide class of statistical models are considered with respect to their ability to model and seasonally forecast geopotential and temperature anomalies representing variability in the polar middle stratosphere from 10 hPa to 30 hPa.

Common statistical methods analyzing polar stratospheric variability are linear (e.g. Camp and Tung, 2007b; Crooks and Gray, 2005) and do not consider more than a few atmospheric forcing factors at the same time. In this work, the statistical methods used are partly nonstationary and nonlinear, incorporating all external forcings simultaneously. The statistical models are trained for the time period from 1 July 1980 through 31 June 2005, where training denotes the process of minimizing a method specific cost function that quantifies the deviation from the truth. Once trained, the statistical methods are used for hindcasting the period from 1 July 2005 to 30 April 2011. The impact of each forcing factor on the statistical response will be quantified. While making reasonable a priori assumptions about the external factors, the statistical model with the best generalization performance is used to forecast the winter 2011/12.

2 Data

This work makes use of two reanalysis data sets resolving the stratosphere, available up to 0.1 hPa from 1979 to present. These data sets are the ERA-Interim (hereafter ERA) reanalysis (Simmons et al., 2006) and the MERRA reanalysis (Rienecker et al., 2011), both considered from 1 July 1980 through 31 June 2011. ERA will be used to train the statistical models on the period from 7 July 1980 through 31 June 2005 and to test the models in a hindcast experiment from 1 July 2005 through 31 June 2011. MERRA is used to validate the final results but not further utilized.

Two daily target geopotential and temperature time series are computed so as to represent the variability in the polar middle stratosphere. Anomalies of the area-weighted average on the polar cap (60–90° N) are computed at 10, 20, and 30 hPa. A subsequent principal component analysis (Jolliffe, 2002) of the three time series reveals that the first principal component (P1) explains more than 90% of the overall variance in both ERA and MERRA. Therefore, only P1 was retained for both geopotential (P1Z) and temperature

(P1T). P1Z and P1T are both positive for weak and warm vortex events and negative for strong and cold vortex conditions. P1T was recently used in Blume et al. (2012) to classify sudden stratospheric warmings events while incorporating important external forcings.

A polar cap average of geopotential anomalies is equivalent to the Northern Annual Mode (NAM) (Baldwin and Dunkerton, 2001), only reversed in sign. The NAM is a popular scalar index to measure polar stratospheric variability (e.g. Thompson, 2003; Baldwin and Thompson, 2009). The NAM is the leading principal component of geopotential anomalies north of 20° N. We decided to use the polar cap method as it is simpler and the resulting time series for geopotential and temperature are positively correlated ($R = 0.8$) (Baldwin and Thompson, 2009), pointing in the same direction during extreme vortex events. P1Z and P1T are physically closely correlated. A lead-lag correlation analysis between P1Z and P1T reveals that there is a correlation of approx. 0.7 when lagging P1T with 10 days, whereas lagging P1Z leads to only 0.3. For instance, during a sudden stratospheric warming, the temperature anomaly usually appears first and the actual vortex breakdown a few days later. In addition, P1T reflects the strong stratospheric cooling (overturning) proceeding most major warmings. For simplicity, P1Z is referred to hereafter as *geopotential* and P1T as *temperature*.

This analysis makes use of nine physical *external factors* which describe large-scale phenomena important for the polar stratosphere. Their purpose is to improve model variability and to obtain insight into relationships and impacts of the various forcings. The factors representing variability in sea surface temperatures (SSTs) are the El Niño-Southern Oscillation (ENSO) (Trenberth, 1997), the Pacific Decadal Oscillation (PDO) (MacDonald and Case, 2005), and the Atlantic Multidecadal Oscillation (AMO) (Delworth and Mann, 2000). Deser et al. (2010) reviews variabilities in sea surface temperature and describes how corresponding indices can be computed. In this work, they have been calculated with an EOF analysis of detrended SST anomalies from 60° S to 60° N with ENSO being the leading EOF (equivalent to the Niño3.4 index).

Furthermore, the first two principal components of equatorial stratospheric zonal wind anomalies (QBO1 and QBO2) (Wallace et al., 1993) are included. Factors representing tropospheric high-latitude blockings (Martius et al., 2009; Woollings and Hoskins, 2008) are the first two principal components of geopotential anomalies between 35° N and 85° N at 500 hPa (BLOC1 and BLOC2). BLOC1, as the leading principal component, is equivalent to the NAM in 500 hPa and represents blockings in both the Atlantic and Pacific sectors simultaneously. Moreover, the F10.7 cm radio flux representing solar variability (SFL; ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA) and the aerosol optical depth (AOD; <http://data.giss.nasa.gov/modelforce/strataer>) representing volcanic eruptions are included. Additionally, three

baseline factors representing the seasonal cycle (one sine and one cosine with a period of one year) and a linear trend term are included. Since the different external factors have different magnitudes and physical units, they are normalized on the full period from 1980 to 2011, such that the minimum is at -1 and the maximum at $+1$. To reduce short-term fluctuations and extreme values, the daily external factors are low-pass filtered using a 5-day running mean.

In order to obtain an idea on how the different factors vary with time, their time series are visualized in Fig. 1. The time series of the sine, the cosine, and the trend term were omitted for simplicity. It is noted that the factors vary on very different timescales from days (e.g. BLOC1) to decades (e.g. AMO). Others vary on both scales such as SFL and factors such as AOD represent only singular events.

In order to improve regression results, optimal time lags of each of the nine physical external factors were calculated using a lead-lag correlation analysis. Every external factor was correlated with geopotential (temperature) for different lags from 0 to 365 days. The largest statistically significant correlation on this period indicates the optimal lag. We obtain zero lags for PDO, BLOC1, BLOC2, and AOD for both geopotential and temperature. For ENSO we computed 96 (82) days; for AMO 8 (185) days; for QBO1 173 (137) days; for QBO2 0 (264) days; and for SFL 0 (50) days, for geopotential and temperature respectively. Please note that we have compiled a list with acronyms and their meanings in Table 1.

3 Statistical models

This work assesses and compares four statistical models with respect to their ability to model geopotential and temperature. These models have been chosen because they make different assumptions about the underlying data. The model intercomparison is similar to that in Blume et al. (2012), in which three different learning approaches were compared with respect to their ability to classify SSWs in major, minor, and final warmings by incorporating polar-cap temperature anomalies along with external factors. The main differences to Blume et al. (2012) are that here we solve a regression problem and use the trained models for forecasting. In addition, an advanced nonstationary method is included, namely FEM-VARX. A statistical method is nonstationary if its response depends on the event number and on time when dealing with time series. The four models are:

1. Linear discriminant analysis (LDA) (Montgomery et al., 2001), also known as multiple linear regression, which is a linear and stationary model. LDA is one of the most common statistical tools to analyze stratospheric variability (e.g. SPARC CCMVal, 2010; Randel et al., 2009; Crooks and Gray, 2005). LDA models data by finding the coefficients to a linear combination of external factors using the method of least-squares. The simplicity

Table 1. The acronyms and their meanings as used in this study.

AMO	Atlantic Multidecadal Oscillation
AOD	Aerosol Optical Depth
BLOC1	First PC of 500 hPa geopotential anomalies from 35° N to 85° N
BLOC2	Second PC after BLOC1
COS	Cosine term with one year period
ENSO	El Niño-Southern Oscillation
EOF	Empirical Orthogonal Function
FEM-VARX	Finite element method plus vector autoregression with external factors
LDA	Linear Discriminant Analysis
MLP	Multi-Layer Perceptron
PC	Principal Component
PDO	Pacific Decadal Oscillation
QBO	Quasi-Biennial Oscillation
QBO1	First PC of equatorial stratospheric zonal mean zonal wind anomalies
QBO2	Second PC after QBO1
SFL	11-yr solar cycle
SIN	Sine term with one year period
SVR	Support Vector Regression
TREND	Linear trend term

and robustness of LDA make it a popular method. LDA does not have free tuning parameters.

2. A cluster method based on finite elements (FEM-VARX) (Horenko, 2010, 2011), which is locally linear in the determined clusters but nonstationary since the switching process between clusters is time dependent. FEM-VARX models data by (a) finding persistent clusters in the time series given the external factors using a finite element method (FEM); and (b) linearly modeling the data corresponding to each cluster by incorporating the external factors using a VARX (vector autoregression with external factors) approach. In this work, autoregressive processes are not considered for simplicity so that VARX boils down to a simple linear combination. FEM-VARX has two tuning parameters in this work: The number of clusters K and the persistency threshold C which limits the maximum transitions from a given cluster to any other cluster.
3. The multi-layer perceptron (MLP) (Zhang et al., 1998), a fully-connected, feed-forward neural network which is a stationary but nonlinear model. Approximation and generalization performance of the MLP stem from the nonlinear transfer function (sigmoid) and the numerous connections within the hidden layers (Bishop, 1995). This analysis is restricted to two hidden layers since it was shown that an MLP with two hidden layers and sigmoidal transfer function can approximate any continuous function (Kurkova, 1992). MLP is trained using back-propagation (Avriel, 2003) which iteratively

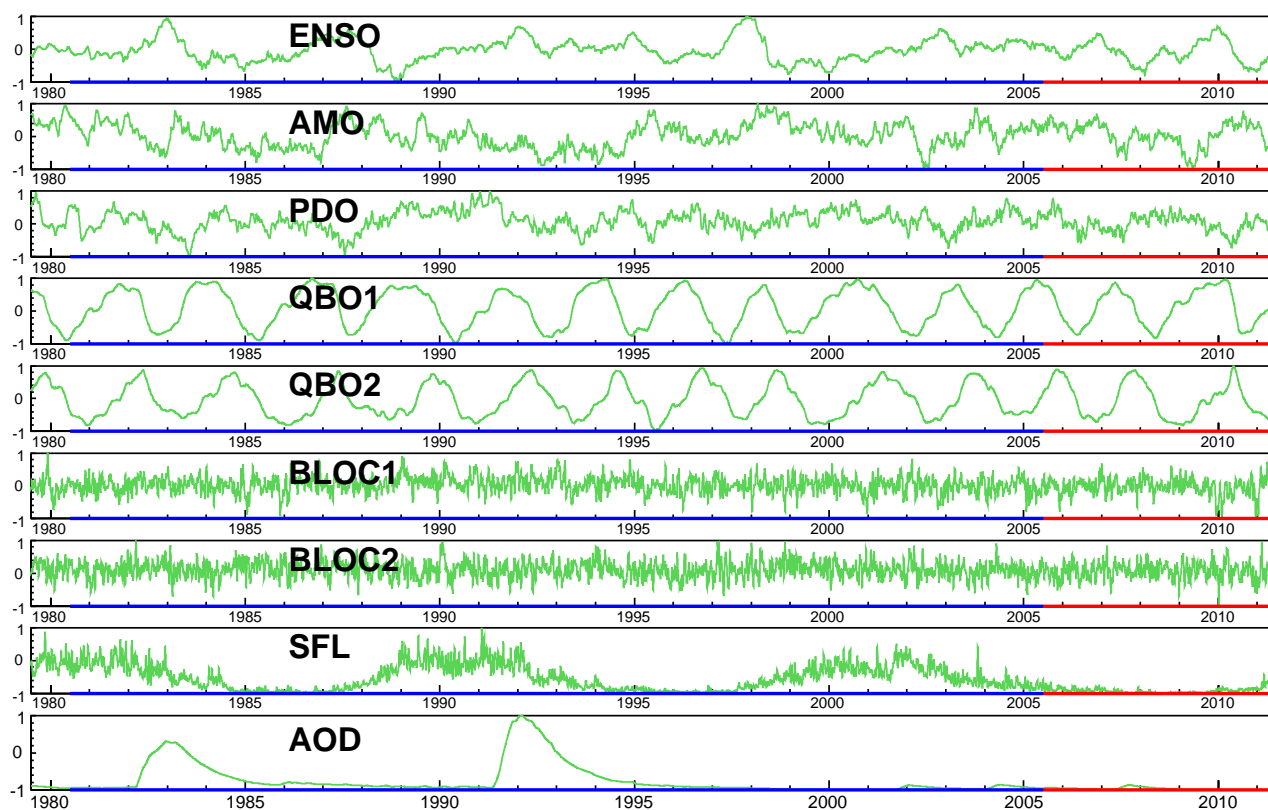


Fig. 1. The different external factors in green as used in this study, omitting sine, cosine, and trend terms for simplicity. Blue denotes the training and red the hindcast period.

adjusts weights and biases. A weight is given at each synapse (connection between two neurons) and a bias at each neuron. Each neuron calculates a weighted linear combination of its inputs from neighboring neurons. This linear combination is then post-processed with the sigmoid function. In this work, MLP has two tuning parameters: the number of neurons in the first hidden layer $L_1 > 0$ and the number of neurons in the second hidden layer $L_2 \geq 0$.

4. The support vector regression (SVR) (Basak et al., 2007), which is also stationary but nonlinear. SVR is a support vector machine (Vapnik, 1995) adjusted to solve regression problems. Approximation and generalization performance of a support vector machine stem from the nonlinear kernel that is used to transform the feature space into a higher dimensional space. In this work, the popular radial basis kernel is used. This kernel was shown to be the most efficient while being very simple and including the linear case (Keerthi and Lin, 2003). SVR has two tuning parameters: the scaling parameter γ in the radial basis kernel and the parameter δ controlling the trade-off between approximating the training data and generalizing to unseen data.

It is worth noting that LDA and FEM-VARX can in principle also be used in a nonlinear fashion by transforming the external factors with a nonlinear function (e.g. higher order polynomials, cross-products of different factors, etc.) in a first step. However, this has not been done here because the corresponding functions are unknown a priori. The nonlinearity in MLP and SVR stems from the nonlinear transfer function and the nonlinear kernel in MLP and SVR, respectively, so that the explicit knowledge of a transfer function is not necessary.

It is interesting to see in which areas of atmospheric research more advanced statistical methods are already in use. For instance, Walter and Schönwiese (2002) addressed the detection and attribution of observed global climate change in global temperature anomalies using a neural network. Lu et al. (2009) used a nonlinear neural network to receive an alternative representation of the QBO. Coughlin and Gray (2009) use the K-Means cluster algorithm to determine two distinct states in the polar stratosphere, one representing normal, the other weak vortex conditions. Nao et al. (2006) presents a technique based on support vector machines to estimate the surface area of polar stratospheric clouds. Franzke et al. (2009) used the nonstationary FEM-VARX clustering approach to identify large-scale dynamical circulation patterns in GCM simulations.

As previously indicated, all models but LDA depend on a set of free tuning parameters that needs to be determined, which is called the model architecture. The optimal model architecture (combination of tuning parameters) aims at meeting the principle of Occam's Razor (Ariew, 1976), stating that the simplest model is the preferred if it contains just as much information as any of the more complicated models. There are two major branches found in the literature of information theory (Burnham and Anderson, 2002) aiming at selecting the optimal model: information criteria and cross-validation. The approach to be used depends on the statistical method and the specific application.

For FEM-VARX, the optimal architecture was determined with the use of the Akaike information criterion (Akaike, 1974; Horenko, 2011) where the parameter setting leading to the smallest criterion is preferred. For MLP and SVR, a 5-fold cross-validation (Kohavi, 1995) was conducted in which the training data were partitioned into 5 equally-sized contiguous subsets (folds). The model architecture with the largest correlation calculated from these tested subsets was selected. In the following, the optimal values are given in parentheses for geopotential and temperature, respectively. For FEM-VARX, K (5, 5) denotes the number of clusters and C (146, 112) the persistency threshold. For MLP, L_1 (8, 3) and L_2 (5, 0) denote the number of neurons in the first and second hidden layer, respectively. For SVR, γ (1, 0.2) denotes the radial scaling parameter and δ (0.3, 0.3) the trade-off parameter.

LDA and SVR only lead to global solutions, whereas FEM-VARX and MLP might run into local minima during training. In order to reduce this effect, a total of 30 models were trained for FEM-VARX and MLP. For FEM-VARX, K is fixed and values of C were chosen slightly different from the optimal value. For MLP, those pairs of L_1 and L_2 were chosen that were ranked highest according to the cross-validation. The regression and forecast results are the average across these realizations.

4 Results for training and hindcast period

The statistical models are trained with data from the training period (1980–2005) while being set up with optimal model architectures as described above. After being trained, the models are used to hindcast the period from 2005 through 2011, meaning that the models are evaluated with the available external factors from this period. The result of this procedure is presented for geopotential in Fig. 2, where the truth is shown in gray. The correlation coefficient between each model and truth is given in parentheses. The training period is modeled well ($R \approx 0.9$) by all models except LDA ($R \approx 0.3$). FEM-VARX possesses the highest explanatory power over the training period. Please note that all external factors are used in a resolution of a few days.

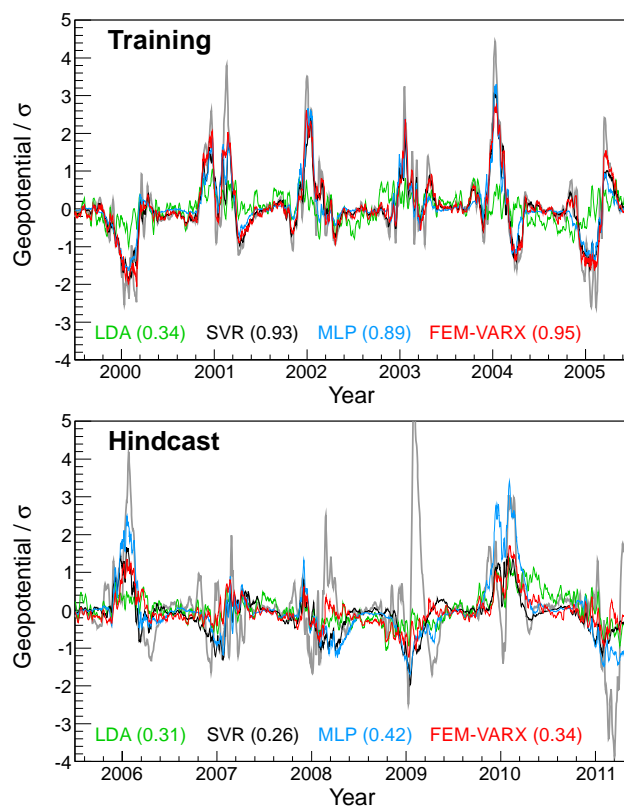


Fig. 2. Geopotential results for the training period (top) and the hindcast period (bottom) for each of the statistical models together with the correlation coefficient R (in parentheses) calculated between the particular model and truth (gray). The hindcast is shown for the full hindcast period whereas the training results are only shown for a representative period (last six years). Labeled year is the first of January of the particular year. The 95 % confidence interval of the correlation factors are ± 0.02 for the training and ± 0.04 for the hindcast period.

For the hindcast period a large drop in correlation is observed for all methods. Please note that when the performance on the hindcast period is much worse than the performance on the training period, it is generally referred to as overfitting. As observed, the effect of overfitting is largest for SVR and also still large for MLP and FEM-VARX. There is no significant overfitting found for the LDA response. Overfitting is not necessarily a problem but it shows that approximating training data does not imply a proper generalization to unseen data.

As mentioned earlier, the truths are anomalies with respect to a climatology. This means that any statistically significant correlation larger than zero between truth and model beats the climatology. As shown in Fig. 2, this is true for all models on the training and the hindcast period. SVR ($R = 0.26$) performs significantly worse than the other models on the hindcast period. FEM-VARX ($R = 0.34$) leads to a higher correlation than LDA ($R = 0.31$) which are, however,

not significantly different from each other. MLP shows the best overall hindcast performance. The MLP correlation of 0.42 is significantly larger than that of the remaining model responses.

Despite only moderate correlations between truth and model response, FEM-VARX and MLP are able to approximate most anomalies and hindcast the general behavior in 5 out of 6 winters. Looking more closely, significant differences between truth and hindcast become evident. The most obvious is the sudden stratospheric warming in January 2009. This is an extraordinary strong warming (Labitzke and Kunze, 2009b) during solar minimum and QBO west, which was not expected according to the Solar-QBO relationships by Labitzke and Kunze (2009a) and Camp and Tung (2007a). This is an example of variability that cannot be explained using the current statistical models. This may be due to errors and deficiencies in the statistical models or due to the internal chaotic nature of the system. Also, the present set of external factors might not be optimal, needing further investigation.

The temperature results (not shown) are similar to the geopotential results but lead to smaller correlations. For the training period, correlations of 0.25 (LDA), 0.53 (SVR), 0.86 (MLP), and 0.92 (FEM-VARX) are obtained. For the hindcast period, correlations of 0.22 (LDA), 0.16 (SVR), 0.25 (MLP), and 0.24 (FEM-VARX) are computed. These correlations, except for SVR, are statistically indistinguishable. The hindcast performance for temperature is not satisfactory and aims at improving it should be made in future statistical analysis. However, we still obtain the same correlation ranking of the different models as for geopotential, pointing to the MLP as the model with the best generalization performance.

It can be seen that it is possible to statistically model and hindcast polar stratospheric variability to a certain extent. The MERRA reanalysis is utilized to validate the regression results by evaluating each statistical model with MERRA data on both the training and the hindcast period. This leads to very similar results compared to ERA (not shown). We conclude that the ERA results can be considered robust and trustworthy.

5 Impact of external factors

The statistical importance (or *impact*) of each of the external factors on the statistical models is calculated. The impact I_k is the standard deviation of the difference between model responses so that $I_k = \sigma(Y - Y^{(k)})$, where Y is the original model response and $Y^{(k)}$ is the model response for external factor k held constant at its median. I_k represents the averaged response deviation from the equilibrium response given by Y . The relative impact is then simply I_k divided by the sum of all impacts for one statistical model. This is shown in Fig. 3 for geopotential and temperature along with a weighted average over all four models. The weights were determined from the correlation coefficients of the training

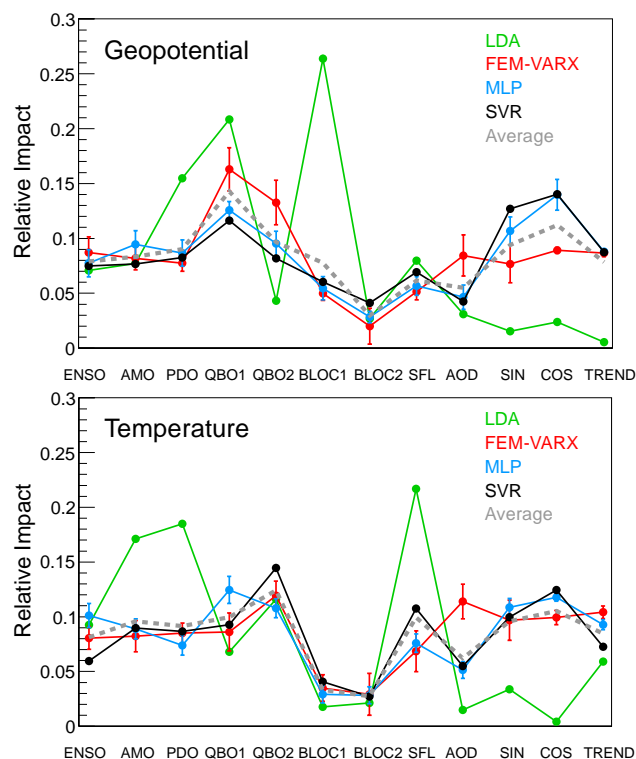


Fig. 3. Relative impact of the external factors on each of the statistical models for geopotential (top) and temperature (bottom). The average impact (gray) is calculated as a weighted mean over the different models where the weights are calculated from the correlation coefficients (see left panel of Fig. 2) on the training period. The error bars of the FEM-VARX and MLP impacts denote the 95 % confidence interval calculated from the 30 model realizations.

period, meaning that FEM-VARX is given the largest weight and LDA the smallest. It is observed that the impacts of FEM-VARX, MLP, and SVR are very similar, whereas LDA misinterprets the importance of factors, such as the impact of high-latitude blockings (BLOC1) on the geopotential or the solar cycle (SFL) on the temperature. LDA assumes linear and stationary relationships, which is not valid for the polar stratosphere (e.g. Calvo et al., 2009; Richter et al., 2011).

Apart from LDA, the impacts in geopotential and temperature are very similar across the different models. A large impact of the QBO terms and a medium impact of SFL are observed, in agreement with e.g. Holton and Tan (1982), Labitzke and Kunze (2009a) and Camp and Tung (2007a). QBO1 is more important than QBO2 for geopotential and vice versa for temperature. The ENSO impact on vortex variability is moderate, as also found by Camp and Tung (2007b) and Mitchell et al. (2011). The AMO and PDO impacts are of similar magnitude. There are only two sufficiently powerful volcanic eruptions (El Chichón in 1982 and Mt. Pinatubo in 1991) (Robock, 2000). Therefore, the impact of the aerosol optical depth (AOD) index is very small for this period. It is worth noting that the AOD impacts vary

significantly across the four models, reflecting a large uncertainty for this forcing, possibly caused by the small number of eruptions important for the stratosphere. Surprisingly, the two factors representing tropospheric high-latitude blockings (BLOC1/2) show a relatively small importance. Especially for the modeling of temperature, they can be omitted. However, the BLOC1 impact on geopotential is of the same order as the SFL impact and needs to be accounted for. However, as stressed by Woollings and Hoskins (2008), BLOC1 represents high-latitude blockings in both the Atlantic and Pacific sectors simultaneously and cannot be used as a proxy for all blocking situations. Another challenge with blockings is that they are shown to precede SSWs but they also appear without an SSW following (Martius et al., 2009), making it difficult to use them for statistical modeling and therefore resulting in small statistical impacts. The sine and cosine terms largely influence the model response, which reflects the strong seasonal dependence of the dynamics in the polar stratosphere. The linear trend term was also found to be relatively strong ($\approx 10\%$).

6 The winter 2011/12 forecast

MLP performs best over the hindcast period (see Fig. 2) and is therefore used to predict the winter 2011/12. As this winter lies in the future, we need to make assumptions about the external factors while taking into account the optimal lags from Sect. 2. For SST variability along with SFL, we used predictions from the NOAA Climate Prediction Center (<http://www.cpc.ncep.noaa.gov>). We obtain -0.5 for ENSO, 0 for AMO, -0.3 for PDO, and -0.5 for SFL. Please note that the external factors are normalized between -1 and $+1$ for the period from 1980 to 2011. BLOC2 has a very small impact (see Fig. 3), which is why it is set to zero. AOD is held at -1 , as future volcanic eruptions that might affect the stratosphere are unknown. The trend term is held at one (its value in 2011), as an approximate value for the extension of only one winter. We selected 0.8 for QBO1 and 0 for QBO2 by extending the corresponding oscillations with a period of 28 months.

Figure 4 shows the resulting MLP forecast for the winter 2011/12 by only varying the sine and cosine terms for geopotential and temperature and for three different conditions of BLOC1. A value of -1 represents extremely pronounced high-latitude blocking situations (Woollings and Hoskins, 2008), whereas $+1$ represents no high-latitude blockings at all. For moderate values of BLOC1, the synoptic situation remains unclear and regional blocking situations may still occur. It is shown in Fig. 4 that the geopotential forecast changes significantly with varying BLOC1. However, for minimum and average BLOC1 conditions, the geopotential forecast is well above one standard deviation. This also holds for the temperature forecast, which is almost unaffected by BLOC1 changes, indicating the small statistical importance

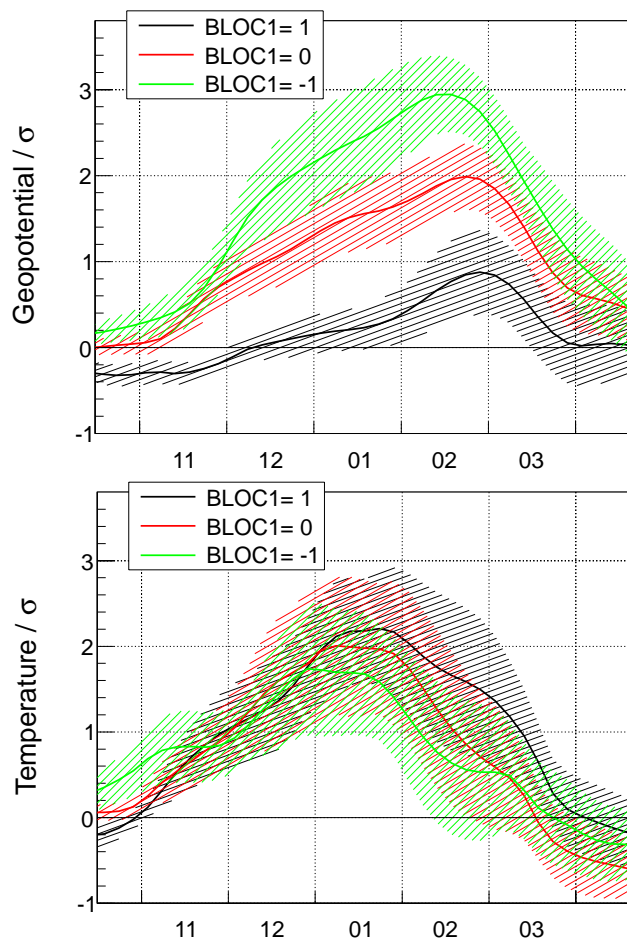


Fig. 4. MLP forecast for the winter 2011/12, holding the external factors constant and varying only the sine and cosine terms. The assumptions about the external factors are partly received from predictions made by the NOAA and partly from scientific reasoning (see text). The forecast is shown for geopotential (top) and temperature (bottom) for three different conditions of BLOC1. The hatched area denotes the 95% confidence interval calculated from the 30 model realizations. Please note the additional uncertainty imposed by the only moderate hindcasting performance (see right panel of Fig. 2).

of BLOC1 on the temperature response (see Fig. 3). To summarize, both forecasts tend to be positive and well above one sigma, indicating extreme variability and a warm stratospheric winter with a weak stratospheric vortex. Since the anomalies in Fig. 4 are quite large, denoting extreme conditions, a sudden stratospheric warming is likely to take place in late January, early February 2012. The temperature anomaly leads and is preceded by the geopotential anomaly.

The winter 2011/12 will most probably coincide with a westerly QBO in 50 hPa and weak solar activity (NOAA). Hence, our finding contrasts the Solar-QBO relationship found by Labitzke and Kunze (2009a), which predicts a cold and undisturbed polar stratosphere under these conditions. Correspondingly, Camp and Tung (2007a) found that solar

minimum conditions and a westerly QBO point to the least disturbed vortex state. Moreover, work performed by, e.g. Camp and Tung (2007b) and Mitchell et al. (2011), indicates that a warm and disturbed polar stratosphere is more likely to take place during warm ENSO phases (El Niño) than during cold ENSO phases (La Niña). This is also in contrast to our forecast, since the ENSO index is most likely to be moderately negative for the winter 2011/12 according to the NOAA predictions. However, since the impacts of the individual external factors do not add up linearly, a nonlinear statistical method is certainly more appropriate. Our analysis, in addition to being nonlinear, incorporates all the significant external factors simultaneously.

7 The winter 2011/12 – observations

During the review process of this paper, the winter 2011/12 has passed, giving us now the opportunity to directly compare our prediction of the polar vortex conditions with actual observations. The NCEP/NCAR reanalysis (Kalnay et al., 1996) is used as a reference as it provides a gridded, freely available data set until almost present day. This is shown in Fig. 5 for geopotential and temperature, which represent polar cap anomalies and were computed as described in Sect. 2. Hence, they can be directly compared to the statistical forecast presented in Fig. 4. In order to measure if the vortex broke down, the zonal mean zonal wind at 60° N and 10 hPa is shown on the bottom panel of Fig. 5. During a vortex breakdown, i.e. a sudden stratospheric warming, the zonal wind is smaller than zero (easterlies).

It is observed in Fig. 5 that the polar vortex in early winter was stronger and colder than usual, indicated by the negative anomalies in November and December 2011. Also, the zonal wind is slightly stronger than the climatology during early winter with a maximum of 48 m s^{-1} , indicating a strong polar vortex. Then, suddenly, the temperature rises by almost 3σ at the end of December within a few days. The geopotential follows approx. two weeks after and both reach maximum values of 2.5σ (geopotential) and 3.5σ (temperature) in mid January to then decrease to climatological values within approx. four weeks (geopotential) and two weeks (temperature). The zonal wind follows somewhat the progression of the geopotential and reaches a first minimum of 8 m s^{-1} in mid January. The zonal wind, however, does not drop below values of 6 m s^{-1} for the rest of the actual winter, so that no vortex breakdown, i.e. no major stratospheric warming took place. At the end of March, the transition toward the stratospheric summer circulation is observed.

There was no vortex breakdown, but the vortex conditions of the winter 2011/12 were extreme. The vortex was anomalously weak from mid January to mid February and anomalously warm from end of December to end of January. This is a classical example of a minor warming during mid-winter (Labitzke and Naujokat, 2000). Therefore, a winter 2011/12

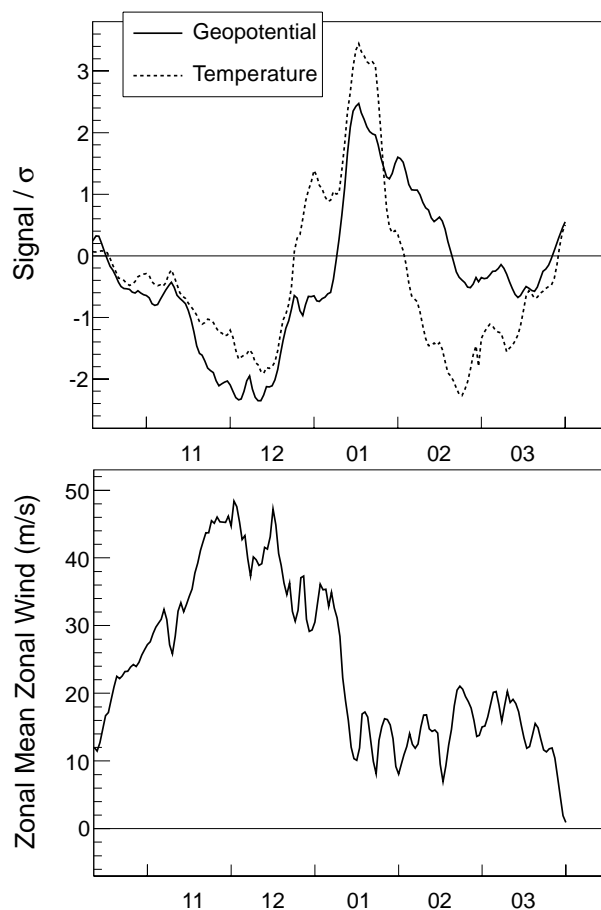


Fig. 5. Observed polar stratospheric variability of the winter 2011/12 until March 31st for geopotential and temperature (top) and the zonal mean zonal wind at 60° N and 10 hPa (bottom).

with warm and weak stratospheric vortex conditions was predicted correctly in Sect. 6. Even the temperature maximum for mid January was predicted correctly, as shown in Fig. 4. The forecast for the geopotential is already large in mid January (for BLOC1 = $-1, 0$) but does not peak until February. It is notable that the magnitudes of the forecast (except for BLOC1 = 1) are similar to those found in observations. In addition to correctly forecasting weak and warm vortex conditions in Sect. 6, a vortex breakdown for late January, early February was also predicted. This did not take place, as mentioned earlier, and was not forecasted correctly. However, the vortex remained the weakest from mid January to mid February, as observed in Fig. 5 for the geopotential and the zonal wind. The overall forecast was correct to a large extent concerning the positive anomalies representing the minor warming in mid-winter. The strong negative anomalies in early and late winter, representing strong and cold vortex conditions, could not be forecasted correctly. It appears to be a particular challenge to statistically forecast strong vortex events, as also observed in the hindcast results shown in Fig. 2.

8 Conclusions

We have presented a novel statistical treatment of variabilities in the polar middle stratosphere, making use of four independent and different statistical models. For the first time, partly nonstationary and nonlinear statistical methods were trained with polar stratospheric geopotential and temperature anomalies incorporating all significant external factors simultaneously (Fig. 1). It was shown that, with the help of external factors, SVR, FEM-VARX, and MLP are able to efficiently approximate geopotential (Fig. 2, left) and temperature anomalies on the training period (1980–2005), whereas LDA performs significantly worse. On the hindcast period (2005–2011) for geopotential (Fig. 2, right), MLP performs significantly better than the other models. MLP and FEM-VARX are able to hindcast most anomalies and the general winter behavior in 5 out of 6 winters. It should be noted, however, that the FEM-VARX hindcast is not significantly better than that of LDA in terms of the correlation coefficient. The temperature hindcast correlations are generally smaller but still lead to the same ranking, pointing to MLP as the model with the best generalization performance. However, the current methodology needs significant future improvements in order to use statistical models to reliably hindcast polar stratospheric temperature anomalies. But the MLP hindcast performance for geopotential in particular is promising. However, a degree of variability remains, seen in the sudden stratospheric warming of January 2009, that cannot be forecasted using the current framework. This may be due to model errors or due to the chaotic nature of the system. Also, the present set of external factors might not be optimal.

The statistical impact of each of the external factors on the statistical models was computed (Fig. 3). It was shown that the QBO factors, the seasonal terms, and the trend term show the greatest impact. The solar cycle and the SST variabilities have a medium impact along with high-latitude large-scale blockings (BLOC1). Volcanic eruptions (AOD) only point to a small but more uncertain statistical importance. It was observed that relative impacts of external factors are very similar for FEM-VARX, MLP, and SVR, whereas those of LDA differ significantly from the model-averaged impact. Therefore, LDA should not be used for a study like this as it does not weight the external factors correctly.

Since the multi-layer perceptron (MLP) showed the best generalization performance, it was used to predict the winter 2011/12 under reasonable assumptions about the external factors (Fig. 4). It predicts a disturbed and warm polar stratosphere, with a sudden stratospheric warming likely to take place during late January, early February 2012. This is in contrast to previous studies which expect a cold and less disturbed polar stratosphere given the same external factors (weak solar, QBO west, La Niña). However, standard analysis is based on linear models and does not consider more than a few external factors at the same time. Our prediction

is based on a nonlinear statistical method incorporating all significant external factors simultaneously.

During the review process of this paper, the winter 2011/12 has passed so that the statistical forecast could be compared to actual observations (Fig. 5). It was found that the prediction of warm and weak vortex conditions was correct. The prediction of a vortex breakdown was not. However, a strong minor warming was observed during mid-winter and the zonal flow slowed down and was the smallest from mid January to mid February. Hence, the forecast of positive anomalies during the winter 2011/12 was correct to a large extent, which shows the great potential in using nonlinear statistical models for the modeling and forecasting of polar stratospheric variability. However, it was not possible to forecast the large negative anomalies in early and late winter, representing strong and cold vortex conditions.

There are several improvements that could be made to this analysis. There may exist other, currently not included external factors that may improve the statistical forecasting of polar stratospheric variability. For instance, a different and possibly more regional blocking index should be tested within the current framework. In the current study, a set of factors was held constant and the optimal model architecture was computed for each statistical method. However, it would be favorable to optimize on the set of external factors plus the internal model setting. For each tested set of external factors, the optimal model setting would have to be estimated using information criteria or cross-validation. This is usually computationally expensive but may be feasible for a reasonable number of training events and external factors.

Instead of a linear lag correlation analysis, the lags should be computed separately for each statistical model, with a grid search technique using cross-validation. Unfortunately, these lag calculations would be computationally extremely expensive. It would also be interesting to decrease the temporal resolution of the considered time series to see if the modeling improves. Additionally, the nonlinear interrelationships between external factors should be further investigated using the introduced methods.

Acknowledgements. This work has been carried out within the Helmholtz-University Young Investigators Group NATHAN funded by the Helmholtz-Association through the President's Initiative and Networking Fund, the GFZ Potsdam, and by FU Berlin. We thank the members of the Stratosphere Research Group at the Institute for Meteorology of FU Berlin for helpful discussions. We are grateful to Edwin Gerber and two anonymous reviewers for their constructive comments.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: M. Palm

References

- Akaike, H.: A new look at the statistical model identification, *IEEE T. Automat. Contr.*, 19, 716–723, 1974.
- Ariew, R.: Ockham's Razor: A Historical and Philosophical Analysis of Ockham's Principle of Parsimony, Ph.D. thesis, Champaign-Urbana, University of Illinois, 1976.
- Avriel, M.: *Nonlinear Programming: Analysis and Methods*, Dover Publishing, 2003.
- Baldwin, M. and Dunkerton, T.: Stratospheric harbingers of anomalous Weather Regimes, *Science*, 294, 581–584, 2001.
- Baldwin, M. and Thompson, D. W. J.: A critical comparison of stratosphere-troposphere coupling indices, *Q. J. Roy. Meteor. Soc.*, 135, 1661–1672, 2009.
- Baldwin, M., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, *Rev. Geophys.*, 39, 179–229, 2001.
- Baldwin, M. P. and Holton, J. R.: Climatology of the stratospheric polar vortex and planetary wave breaking, *J. Atmos. Sci.*, 45, 1123–1142, 1988.
- Basak, D., Pal, S., and Patranabis, D. C.: Support Vector Regression, *Neural Info. Proc.*, 11, 203–224, 2007.
- Bishop, C.: *Neural networks for pattern recognition*, Oxford: University Press, 1995.
- Blume, C., Matthes, K., and Horenko, I.: Supervised learning approaches to classify sudden stratospheric warming events, *J. Atmos. Sci.*, 69, 1824–1840, 2012.
- Burnham, K. P. and Anderson, D. R.: *Model selection and multimodel inference: a practical information-theoretic approach*, Springer, 2002.
- Calvo, N., Giorgetta, M. A., Garcia-Herrera, R., and Manzini, E.: Nonlinearity of the combined warm ENSO and QBO effects on the Northern Hemisphere polar vortex in MAECHAM5 simulations, *J. Geophys. Res.*, 114, D13109, doi:10.1029/2008JD011445, 2009.
- Camp, C. D. and Tung, K.-K.: The Influence of the Solar Cycle and QBO on the Late-Winter Stratospheric Polar Vortex, *J. Atmos. Sci.*, 64, 1267–1283, 2007a.
- Camp, C. D. and Tung, K.-K.: Stratospheric polar warming by ENSO in winter: A statistical study, *Geophys. Res. Lett.*, 34, L04809, doi:10.1029/2006GL028521, 2007b.
- Coughlin, K. and Gray, L. J.: A continuum of sudden stratospheric warmings, *J. Atmos. Sci.*, 66, 531–540, 2009.
- Crooks, S. and Gray, L.: Characterization of the 11-year solar signal using a multiple regression analysis of the ERA-40 dataset, *J. Climate*, 18, 996–1015, 2005.
- Delworth, T. L. and Mann, M. E.: Observed and simulated multidecadal variability in the Northern Hemisphere, *Climate Dynam.*, 16, 661–676, 2000.
- Deser, C., Alexander, M. A., Xie, S.-P., and Phillips, A. S.: Sea Surface Temperature Variability: Patterns and Mechanisms, *Ann. Rev. Mar. Sci.*, 2, 115–143, 2010.
- Franzke, C., Horenko, I., Majda, A. J., and Klein, R.: Systematic metastable atmospheric regime identification in an AGCM, *J. Atmos. Sci.*, 66, 1997–2012, 2009.
- Garfinkel, C. I. and Hartmann, D. L.: Effects of the El-Niño Southern Oscillation and the Quasi-Biennial Oscillation on polar temperatures in the stratosphere, *J. Geophys. Res.*, 112, D19112, doi:10.1029/2007JD008481, 2007.
- Gerber, E. P., Orbe, C., and Polvani, L. M.: Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts, *Geophys. Res. Lett.*, 36, L24801, doi:10.1029/2009GL040913, 2009.
- Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., Luterbacher, J., Meehl, G. A., Shindell, D., van Geel, B., and White, W.: Solar Influences on Climate, *Rev. Geophys.*, 48, RG4001, doi:10.1029/2009RG000282, 2010.
- Holton, J. R. and Tan, H. C.: The influence of the equatorial Quasi-Biennial Oscillation on the global atmospheric circulation at 50mb, *J. Atmos. Sci.*, 37, 2200–2208, 1980.
- Holton, J. R. and Tan, H. C.: The quasi-biennial oscillation in the Northern Hemisphere lower stratosphere, *J. Meteorol. Soc. Jpn.*, 60, 140–148, 1982.
- Horenko, I.: On the Identification of Nonstationary Factor Models and Their Application to Atmospheric Data Analysis, *J. Atmos. Sci.*, 67, 1559–1574, 2010.
- Horenko, I.: Nonstationarity in multifactor models of discrete jump processes, memory and application to cloud modeling, *J. Atmos. Sci.*, 68, 1493–1506, 2011.
- Jolliffe, I. T.: *Principal Component Analysis*, Springer Series in Statistics, 2002.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteor. Soc.*, 77, 437–472, 1996.
- Keerthi, S. S. and Lin, C.-J.: Asymptotic behaviors of support vector machines with Gaussian kernel, *Neural Comput.*, 15, 1667–1689, 2003.
- Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, vol. 2, 1137–1143, Morgan Kaufmann, 1995.
- Kurkova, V.: Kolmogorov's theorem and multilayer neural networks, *Neural Networks*, 5, 501–506, 1992.
- Kuroda, Y.: High initial-time sensitivity of medium-range forecasting observed for a stratospheric sudden warming, *Geophys. Res. Lett.*, 37, L16804, doi:10.1029/2010GL044119, 2010.
- Labitzke, K.: Sunspots, the QBO, and the stratospheric temperatures in the north polar region, *Geophys. Res. Lett.*, 14, 535–537, 1987.
- Labitzke, K. and Kunze, M.: Variability in the stratosphere: The Sun and the QBO, 257–278, *Climate and Weather of the Sun-Earth System (CAWSES): Selected Papers from the 2007 Kyoto Symposium*, Terrapub, Tokyo, 2009a.
- Labitzke, K. and Kunze, M.: On the Remarkable Arctic Winter in 2008/09, *J. Geophys. Res.*, 114, D00I02, doi:10.1029/2009JD012273, 2009b.
- Labitzke, K. and Naujokat, B.: The lower arctic stratosphere in winter since 1952, *SPARC Newsletter No. 15*, 11–14, 2000.
- Labitzke, K. and van Loon, H.: Associations between the 11-year solar cycle, the QBO and the atmosphere. Part I: the troposphere and stratosphere in the northern hemisphere winter, *J. Atmos. Terr. Phys.*, 50, 197–206, 1988.
- Lu, B., Pandolfo, L., and Hamilton, K.: Nonlinear representation of the quasi-biennial oscillation, *J. Atmos. Sci.*, 66, 1886–1904,

- 2009.
- MacDonald, G. M. and Case, R. A.: Variations in the Pacific Decadal Oscillation over the past millennium, *Geophys. Res. Lett.*, 320, L08703, doi:10.1029/2005GL022478, 2005.
- Manzini, E., Giorgetta, M., Esch, M., Kornblueh, L., and Roeckner, E.: The influence of sea surface temperatures on the Northern winter stratosphere: Ensemble simulations with the MAECHAM5 model, *J. Climate*, 19, 3863–3881, 2006.
- Martius, O., Polvani, L. M., and Davies, H. C.: Blocking precursors to stratospheric sudden warming events, *Geophys. Res. Lett.*, 36, L14806, doi:10.1029/2009GL038776, 2009.
- Mitchell, D. M., Gray, L. J., and Charlton-Perez, A. J.: The structure and evolution of the stratospheric vortex in response to natural forcings, *Geophys. Res. Lett.*, 116, D15110, doi:10.1029/2011JD015788, 2011.
- Montgomery, D. C., Peck, E. A., and Vining, G. G.: *Introduction to linear regression analysis*, Wiley-Interscience, 2001.
- Nao, I., Naomi, A., and Sachiko, H.: Estimation of the surface area density of polar stratospheric clouds by using the support vector machines, *IPSI SIG Tech. Reports*, Z0031, 2006B.
- Randel, W. J., Garcia, R. R., Calvo, N., and Marsh, D.: ENSO influence on zonal mean temperature and ozone in the tropical lower stratosphere, *Geophys. Res. Lett.*, 36, L15822, doi:10.1029/2009GL039343, 2009.
- Richter, J. H., Matthes, K., Calvo, N., and Gray, L. J.: Influence of the Quasi-Biennial Oscillation and El Niño-Southern Oscillation on the frequency of sudden stratospheric warmings, *J. Geophys. Res.*, 116, D20111, doi:10.1029/2011JD015757, 2011.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *J. Climate*, 24, 3624–3648, 2011.
- Robock, A.: Volcanic eruptions and climate, *Rev. Geophys.*, 38, 191–219, 2000.
- Simmons, A., Uppala, S. M., Dee, D., and Kobayashi, S.: ERA-Interim: New ECMWF reanalysis products from 1989 onwards, *ECMWF Newsletter*, 110, 26–35, 2006.
- SPARC CCMVal: SPARC Report on the Evaluation of Chemistry-Climate Models, SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, 2010.
- Thompson, D. W. J.: Stratospheric connection to northern hemisphere wintertime weather: Implications for prediction, *J. Climate*, 15, 1421–1428, 2003.
- Trenberth, K. E.: The Definition of El Niño, *B. Am. Meteor. Soc.*, 78, 2771–2777, 1997.
- Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- Wallace, J. M., Panetta, R. L., and Estberg, J.: A phase-space representation of the equatorial stratospheric quasi-biennial oscillation, *J. Atmos. Sci.*, 50, 1751–1762, 1993.
- Walter, A. and Schönwiese, C.-D.: Attribution and detection of anthropogenic climate change using a backpropagation neural network, *Meteorol. Z.*, 11, 335–343, 2002.
- Woollings, T. and Hoskins, B.: Simultaneous Atlantic-Pacific blocking and the Northern Annular Mode, *Q. J. Roy. Meteor. Soc.*, 134, 1635–1646, 2008.
- Zhang, G., Patuwo, B. E., and Hu, M. Y.: Forecasting with artificial neural networks: The state of the art, *Int. J. Forecasting*, 14, 35–62, 1998.