



# On the relationship between total ozone and atmospheric dynamics and chemistry at mid-latitudes – Part 1: Statistical models and spatial fingerprints of atmospheric dynamics and chemistry

L. Frossard<sup>1</sup>, H. E. Rieder<sup>2,\*</sup>, M. Ribatet<sup>1,\*\*</sup>, J. Staehelin<sup>2</sup>, J. A. Maeder<sup>2</sup>, S. Di Rocco<sup>2,3</sup>, A. C. Davison<sup>1</sup>, and T. Peter<sup>2</sup>

<sup>1</sup>Mathematics Institute for Analysis and Applications, EPF Lausanne, Lausanne, Switzerland

<sup>2</sup>Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

<sup>3</sup>Department for Geography, University of Zurich, Zurich, Switzerland

\* now at: Lamont-Doherty Earth Observatory and Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

\*\* now at: Institute of Mathematics and Mathematical Modeling, University Montpellier II, Montpellier, France

*Correspondence to:* L. Frossard (linda.frossard@epfl.ch) and H. E. Rieder (hr2302@columbia.edu)

Received: 30 January 2012 – Published in Atmos. Chem. Phys. Discuss.: 25 May 2012

Revised: 5 November 2012 – Accepted: 28 November 2012 – Published: 8 January 2013

**Abstract.** We use statistical models for mean and extreme values of total column ozone to analyze “fingerprints” of atmospheric dynamics and chemistry on long-term ozone changes at northern and southern mid-latitudes on grid cell basis. At each grid cell, the  $r$ -largest order statistics method is used for the analysis of extreme events in low and high total ozone (termed ELOs and EHOs, respectively), and an autoregressive moving average (ARMA) model is used for the corresponding mean value analysis. In order to describe the dynamical and chemical state of the atmosphere, the statistical models include important atmospheric covariates: the solar cycle, the Quasi-Biennial Oscillation (QBO), ozone depleting substances (ODS) in terms of equivalent effective stratospheric chlorine (EESC), the North Atlantic Oscillation (NAO), the Antarctic Oscillation (AAO), the El Niño/Southern Oscillation (ENSO), and aerosol load after the volcanic eruptions of El Chichón and Mt. Pinatubo. The influence of the individual covariates on mean and extreme levels in total column ozone is derived on a grid cell basis. The results show that “fingerprints”, i.e., significant influence, of dynamical and chemical features are captured in both the “bulk” and the tails of the statistical distribution of ozone, respectively described by mean values and EHOs/ELOs. While results for the solar cycle, QBO, and EESC are in good agreement with findings of earlier studies, unprecedented spatial fingerprints are retrieved for

the dynamical covariates. Column ozone is enhanced over Labrador/Greenland, the North Atlantic sector and over the Norwegian Sea, but is reduced over Europe, Russia and the Eastern United States during the positive NAO phase, and vice-versa during the negative phase. The NAO’s southern counterpart, the AAO, strongly influences column ozone at lower southern mid-latitudes, including the southern parts of South America and the Antarctic Peninsula, and the central southern mid-latitudes. Results for both NAO and AAO confirm the importance of atmospheric dynamics for ozone variability and changes from local/regional to global scales.

## 1 Introduction

Interest in changes in total ozone is linked to its direct influence on biologically active UV radiation (e.g., Calbó et al., 2005), and since the detection of the Antarctic ozone hole (Farman et al., 1985) the development of the Earth’s ozone layer has been a key focus in atmospheric research. The global decrease in column ozone between the 1980s and the 1990s raised major concerns in the scientific community and general public. These paved the way for the “Montreal Protocol for the Protection of the Ozone Layer” (e.g., WMO, 1995, 2003, 2007, 2011), whose successful implementation (e.g., WMO, 2007; Mäder et al., 2010) led to a discussion on

future ozone recovery and possible super-recovery expected in about 50 yr (e.g., WMO, 2007, 2011; Eyring et al., 2007, 2010; SPARC CCMVal, 2010; Hegglin and Shepherd, 2009; Shepherd, 2008).

In previous studies ozone changes on spatial scales have been addressed in statistical terms by large-scale averages over latitude bands (e.g., WMO, 2003, 2007, 2011; Fioletov and Shepherd, 2005; Fioletov et al., 2002), or the application of multiple regression models (e.g., Steinbrecht et al., 2006). The former approach has difficulties, because the large-scale spatial averages usually used for analysis, such as zonal mean values, do not take spatial variability fully into account, while ozone changes vary greatly in time and space, and therefore cannot be addressed adequately by a large-scale average. Applications of multiple regression models on large spatial scales and analyses on a grid cell basis, as for satellite data sets, are rare for column ozone, and the results mostly describe the influence of atmospheric dynamics and chemistry on mean column ozone, leaving the extremes unaccounted for. Steinbrecht et al. (2006) compared multiple regression model output for the winter season on a global scale for observational data from TOMS and SBUV instruments with results from Chemistry-Climate Models (CCMs). Further, ozone changes have been addressed using global CCMs (e.g., Eyring et al., 2007, 2010; SPARC CCMVal, 2010; Waugh et al., 2009; Austin and Wilson, 2006). Such CCM analyses take spatial variability into account, as ozone changes are computed on a grid cell basis, according to changes in atmospheric dynamics and chemistry. Each CCM has its own parameterizations for specific processes and therefore its own strengths and deficiencies in its representation of atmospheric dynamics and chemistry. However, a lot of progress has recently taken place in this field (e.g., Eyring et al., 2010; SPARC CCMVal, 2010).

Statistical modeling<sup>1</sup> of the spatial and temporal variability of column ozone requires one to account for the stochastic nature of physicochemical processes and their spatial and temporal variability (e.g., Chilès and Delfiner, 1999; Diggle and Ribeiro, 2007). In particular, a better description of extreme events and an assessment of whether their distribution has changed over time are needed to make progress in answering important questions about ozone-climate interactions. Although geostatistics is well developed for the treatment of average values, despite recent progress (e.g., Cooley et al., 2006; Naveau et al., 2009; Padoan et al., 2010; Davison et al., 2012) a geostatistics of extremes that could be applied to massive data sets is still lacking. The modeling of extremes is becoming increasingly standard as extreme value theory (EVT) develops (e.g., Coles, 2001; de Haan and Ferreira, 2006), but only a few authors have so far analyzed spatial extremes (e.g., Buishand et al., 2008; Padoan et al., 2010)

<sup>1</sup>In this paper about a statistical analysis of past ozone data, the only “models” considered are statistical ones, even if this is not always explicitly stated.

and/or have dealt with huge gridded data sets (e.g., Cooley et al., 2007; Schliep et al., 2010). Although several frameworks exist for dealing with such data sets using Gaussian processes, such as approximation by Gaussian Markov fields (e.g., Rue and Held, 2005) or through composite likelihoods (e.g., Varin, 2008), there is much less guidance for extremes.

But why is it important to distinguish between models for extreme and non-extreme observations (e.g., mean values)? Since extreme events are situated in the tails of the probability distribution and the focus of interest in this context is often an event beyond any observed value, an accurate statistical description of the tails of the distribution is required. A set of observations, however, contains only few extreme observations by nature, so that a model fitted to an entire data set most likely approximates the bulk of the observations well, but has little capacity to also precisely describe its tails. For the latter purpose, EVT provides specific asymptotically-justified models which in practice are fitted just to extremal data.

Recently, methods from EVT have been applied in total ozone research on local/regional scales (Rieder et al., 2010a,b, 2011). The present study extends these earlier statistical analyses to large spatial data sets and compares the results for ozone extremes with those for mean values. The use of separate, more appropriate, models for these two entities provides better insight into the behavior of total ozone. Mean values very often have a Gaussian distribution and hence can be addressed by standard time series tools like autoregressive moving average (ARMA) models. In contrast, the asymptotic theory behind the EVT models establishes that the limiting distribution of sample maxima is not Gaussian. Since these EVT-based methods are less widely known, they will be presented in Sect. 3.1 of this paper.

Our analysis focuses on the influence of several atmospheric covariates on total ozone across space. Because the data cover the whole globe, a spatial model for extremes would be natural. One possibility is max-stable processes (see Davison et al., 2012, for an application-oriented review), but as current fitting methods are computationally infeasible for massive (gridded) data sets, we fit a model for univariate extremes individually to each grid cell, as if the neighboring cells did not exist. This pointwise approach naturally accounts for non-stationarity in space and avoids averaging effects, caused, for example, by averaging over zonal bands. It also allows us to assess how spatial variation of the model parameters affects the grid cell extremes (which is of major interest for the study), but it does not allow us to model the joint behavior of extremes. This is a topic of continuing research.

In order to ensure that the covariate effects are comparable across the different models and space, the same covariate information in form of a multiple linear regression term is used for both means and extremes and for each grid cell. For each of the three targets of interest (EHOs, ELOs, and mean values), we then estimate the coefficients of the covariates

**Table 1.** Overview of data sets used in this study. Data sets indicated with \* are not available for the entire study period.

Data	Time range (resolution)	Source
Total Ozone	1979–2007 (daily)	Bodeker Scientific, NIWA 2.7 data set <a href="http://www.bodekerscientific.com/data/total-column-ozone">http://www.bodekerscientific.com/data/total-column-ozone</a>
Solar Cycle	1979–2007 (monthly)	NOAA National Geophysical Data Center <a href="ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/Penticton_Adjusted/monthly/MONTHPLT.ADJ">ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/Penticton_Adjusted/monthly/MONTHPLT.ADJ</a>
QBO at 30 and 50 hPa	1979–2007 (monthly)	NOAA National Weather Service Climate Prediction Center <a href="http://www.cpc.noaa.gov/data/indices/">http://www.cpc.noaa.gov/data/indices/</a>
Nino 3.4 Index	1979–2007 (monthly)	NCAR/UCAR Climate and Global Dynamics <a href="http://www.cgd.ucar.edu/cas/catalog/climind/TNI_N34/index.html">http://www.cgd.ucar.edu/cas/catalog/climind/TNI_N34/index.html</a>
NAO Index	1979–2007 (monthly)	NCAR/UCAR Climate and Global Dynamics <a href="http://climatedataguide.ucar.edu/guidance/hurrell-north-atlantic-oscillation-nao-index-pc-based">http://climatedataguide.ucar.edu/guidance/hurrell-north-atlantic-oscillation-nao-index-pc-based</a>
AAO Index	1979–2007 (monthly)	NOAA National Weather Service Climate Prediction Center <a href="http://www.cpc.noaa.gov/products/precip/CWlink/daily_ao_index/ao/monthly.ao.index.b79.current.ascii">http://www.cpc.noaa.gov/products/precip/CWlink/daily_ao_index/ao/monthly.ao.index.b79.current.ascii</a>
Sato Index	1979–2000 (monthly)*	NASA Goddard Institute for Space Studies <a href="http://data.giss.nasa.gov/modelforce/strataer/">http://data.giss.nasa.gov/modelforce/strataer/</a>
ODS (in terms of EESC)	1979–2007 (monthly)	NASA Goddard Institute for Space Studies <a href="http://acdb-ext.gsfc.nasa.gov/Data_services/automailer/">http://acdb-ext.gsfc.nasa.gov/Data_services/automailer/</a>

for each grid cell and produce maps of these estimates for interpretation and comparison.

## 2 Data

### 2.1 Spatial ozone data

In this study, version 2.7 of the NIWA (National Institute of Water and Atmospheric Research, New Zealand) total ozone data set is analyzed for the time period 1979–2007 for the northern (30–60° N) and southern (30–60° S) mid-latitudes. NIWA 2.7 contains daily data at a spatial resolution of 1.25° longitude by 1.0° latitude, adding up to 8640 grid cells for the mid-latitudes of each hemisphere. The data set is based on assimilated and homogenized data from the Total Ozone Mapping Spectrometer (TOMS), the Global Ozone Monitoring Experiment (GOME), Solar Backscatter Ultra-Violet (SBUV) retrievals and Ozone Monitoring Instrument (OMI) retrievals. Drifts between measurements of different satellite instruments have been corrected through inter-satellite instrument comparison and comparison with data from Dobson and Brewer ground-based instruments, which contribute to the Global Atmosphere Watch Program (GAW) of the World Meteorological Organization (WMO). For further details on

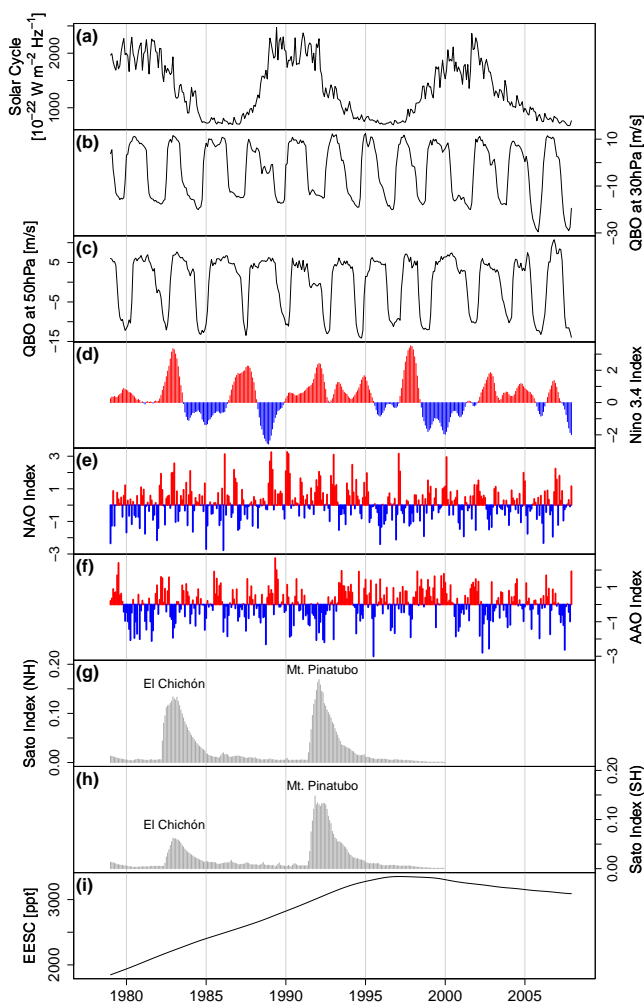
the NIWA assimilated total ozone data set see Bodeker et al. (2005), Müller et al. (2008), and Struthers et al. (2009).

### 2.2 Covariates describing the state of the atmosphere, atmospheric dynamics and chemistry

Various indices describing the dynamical and chemical state of the atmosphere are used as covariates in this study, namely the 11-yr solar cycle, the Quasi-Biennial Oscillation (QBO), the El Niño/Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), the Antarctic Oscillation (AAO), ozone depleting substances (ODS) in terms of equivalent effective stratospheric chlorine (EESC) as calculated by Newman et al. (2007), and finally the stratospheric aerosol load after the major volcanic eruptions of El Chichón and Mt. Pinatubo, as given by Sato et al. (1993). Table 1 contains an overview of the data sets used and Fig. 1 shows the temporal evolution of the indices.

## 3 Methods

In this section we describe the two different models we are using: (i) a model based on extreme value theory (EVT) to analyze extremely high (termed EHOs) and extremely low



**Fig. 1.** Temporal evolution of the covariates used in this study: (a) solar cycle as light flux at 10.7 cm, (b) Quasi-Biennial Oscillation (QBO) Index at 30hPa, (c) as (b) but at 50hPa, (d) Nino 3.4 Index describing the state of the El Niño/Southern Oscillation (ENSO), (e) North Atlantic Oscillation (NAO) Index (principal components of the leading empirically-determined orthogonal function of sea level pressure anomalies over the Atlantic sector (20–80° N, 90° W–40° E), from Hurrell, 2009), (f) Antarctic Oscillation (AAO) Index, (g) volcanic aerosol loading in terms of mean optical thickness, Sato Index (Sato et al., 1993) for the Northern Hemisphere (major volcanic eruptions of El Chichón (1982) and Mt. Pinatubo (1991) are marked), (h) as (g) but for the Southern Hemisphere, and (i) atmospheric loading of ODS in terms of equivalent effective stratospheric chlorine (EESC). All covariates are given on a monthly basis. For information on the sources of the data sets for the individual covariates see Table 1.

(termed ELOs) events in total ozone (see Sect. 3.1); and (ii) an autoregressive moving average model (ARMA) for the analysis of mean values (see Sect. 3.2). In order to assess the effect of the covariates described in Sect. 2.2 on the different regions of the distribution function of total ozone, the same covariate-dependent expression is used in the mod-

els (i) and (ii), while the remaining structure of each model accounts for the nature of the underlying data (EHOs/ELOs or mean values).

### 3.1 Model for total ozone extremes

We first give a short overview on extreme value theory and some associated concepts. An accessible reference on this topic is Coles (2001).

#### 3.1.1 Extreme value theory

A fundamental result of extreme value statistics implies that the behavior of the sample maximum of a large number of identically distributed random variables should, under mild conditions, be well-approximated by the Generalized Extreme Value (GEV) distribution. The GEV has three parameters, the location parameter  $\mu \in \mathbb{R}$ , the scale parameter  $\sigma > 0$ , and the shape parameter  $\xi \in \mathbb{R}$ . Using the notation  $x_+ = \max(x, 0)$ , its probability distribution function is

$$F(y) = \exp\left(-\left(1 + \xi \frac{y - \mu}{\sigma}\right)_+^{-1/\xi}\right), \quad (1)$$

which is defined on the set  $\{y \in \mathbb{R} : 1 + \xi(y - \mu)/\sigma > 0\}$ . The case  $\xi = 0$  is interpreted as the limit when  $\xi$  approaches zero.

Before presenting statistical models involving the GEV distribution, observe that any model for very high extreme events (maxima) can also be used for very low events (minima) due to the equality

$$\min_{i=1, \dots, n} (Y_i) = - \max_{i=1, \dots, n} (-Y_i).$$

Therefore it suffices to describe only models for maxima hereafter.

There are several methods for modeling extremes of univariate stationary time series. In the block maxima model, the data are grouped into blocks of length a given time period (e.g., a year or a month) and the sequence of block maxima is modeled as an independent sample from the GEV distribution. The drawback of this method is that only one observation per block is retained, so some relevant extreme observations may be jettisoned. A remedy to this waste of data is to fit a model to all observations exceeding a high threshold (Davison and Smith, 1990); one possible model being the Generalized Pareto Distribution, which was fitted to total ozone data in Rieder et al. (2010a, 2011). A major difficulty with these threshold-based approaches is the selection of the threshold, which amounts to a trade-off between bias and variance and hence is crucial for the quality of the model. Unfortunately, there are no reliable automatic selection methods, so that peaks over threshold models are unsuitable when numerous thresholds have to be found.

For non-stationary data, threshold selection is even more exasperating, as there is no theoretical guidance on which

part(s) of the model for exceedances should incorporate non-stationarity. Northrop and Jonathan (2011) suggest a rather complex approach that involves models for non-stationarity in both extremes and the threshold, whereas Eastoe and Tawn (2009) propose to preprocess the full data before modeling the extremes.

In light of these difficulties, we use an alternative approach that sidesteps threshold selection by using a fixed number  $r$  of observations per block. This is known as the  $r$ -largest order statistics model (Coles, 2001, Sect. 3.5) and can be seen as a compromise between the block maxima and peaks over threshold methods. The data are split into blocks and the  $r \geq 1$  largest observations of each block are used in the model. This amounts to using the exceedances of a threshold that varies across blocks, while the blocks are assumed to be independent. The analogue to threshold selection in this model is the choice of  $r$ , which therefore also is a trade-off between bias and variance. But once  $r$  has been fixed, the threshold values are determined automatically, so the approach is well-adapted to use with many parallel time series.

We fit the  $r$ -largest order statistics model by maximum likelihood estimation (brief explanations of useful statistical concepts are given in Appendix A; see Sect. A1 for maximum likelihood estimation). This can be done more or less automatically and with acceptable computational effort even for many grid cells, as there exists a closed-form likelihood. Let  $y_t^{(1)} \geq \dots \geq y_t^{(r)}$  be the  $r$  largest observations in block, or equivalently time period,  $t \in \{1, 2, \dots, T\}$ , define the parameters  $\mu$ ,  $\sigma$ , and  $\xi$  as in the GEV distribution (Eq. 1), and again let  $x_+ = \max(x, 0)$ . Then the likelihood  $L_t$  for block  $t$  based on the  $r$ -largest order statistics model is (Coles, 2001, p. 69)

$$L_t(\mu, \sigma, \xi) = \exp \left( - \left( 1 + \xi \frac{y_t^{(r)} - \mu}{\sigma} \right)_+^{-1/\xi} \right) \times \prod_{j=1}^r \frac{1}{\sigma} \left( 1 + \xi \frac{y_t^{(j)} - \mu}{\sigma} \right)_+^{-1/\xi - 1}. \quad (2)$$

This expression is related to the GEV distribution (Eq. 1), whose probability density function is obtained on setting  $r = 1$ , corresponding to using only the largest observation  $y_t^{(1)}$  of the  $t$ -th block for inference. The parameters  $\mu$ ,  $\sigma$ , and  $\xi$  have the same interpretation as in Eq. (1). The assumed independence of the blocks in the  $r$ -largest order statistics model implies that the likelihood for the whole data is the product of the block contributions  $L_t(\mu, \sigma, \xi)$  given in Eq. (2). For a large sample size, i.e., a large number of blocks  $T$ , the asymptotic Gaussian distribution of the maximum likelihood estimator is a good approximation and hence the significance of the parameter estimates can be assessed by  $z$ -tests (see Sect. A2).

The models described so far apply for stationary data, but environmental time series generally show trend and seasonality. One way to deal with this is to let the parameters of

the GEV distribution vary over time. The time-dependent expression is usually based on simple mathematical functions or covariates related to the analyzed phenomenon, and if possible summarized as the product of a design matrix  $\mathbf{Z}$  and a parameter vector  $\boldsymbol{\beta}$ .

All the models described above rely on the implicit assumption that the underlying observations are independent, which considerably simplifies theoretical development and estimation procedures. For temporally correlated data, however, extreme events tend to occur in clusters. The observations retained in the extreme value model are thus dependent in reality, unlike in the associated likelihood function. This is a weakness of the  $r$ -largest order statistics model, because the occurrence times of the extremes are not taken into account, implying that clusters of extremes are unrecognized either in the middle of a block or at the transition between two blocks. It can be shown that ignoring these clusters, and applying standard likelihood theory based on Eq. (2) anyway, leads to estimators that are unbiased with standard errors that are somewhat too small. While declustering procedures exist, estimators based on declustered data tend to be biased (Fawcett and Walshaw, 2012), so we have preferred to use the original, non-declustered data, and to bear in mind that the quoted standard errors are probably 10–20 % too small. Detailed treatment of clusters would be essential when modelling the dynamic behaviour of ELOs or EHOs.

### 3.1.2 Implementation for the ozone data

In this section, we give details on how the  $r$ -largest order statistics model is fitted to every grid cell of the NIWA data set in a region of interest. As mentioned above, we have chosen to use the  $r$ -largest order statistics model instead of the peaks over threshold model mainly because the latter would have required a huge number of manual threshold selections due to the number of grid cells and non-stationarity in time.

Our reasoning for the selection of  $r$ , the number of extremal observations used in each block of the  $r$ -largest order statistics model, is as follows: considering that a month has about 30 days,  $r$  should not exceed 3 for the observations to be “extreme”, because  $r = 3$  corresponds roughly to a threshold at the 90th percentile. In view of the general uncertainty that might govern the estimation due to the high number of covariates, we chose  $r = 3$  to increase the precision of the estimates. Since using  $r$  too large may increase the bias, we also fitted the same model with  $r = 1$  and  $r = 2$ . The parameter estimates in these cases showed the same spatial patterns as for  $r = 3$ , but as anticipated, the standard errors of the estimates were larger than with  $r = 3$ , and hence the  $z$ -statistics were less significant.

For each grid cell, non-stationarity in time is handled in two ways: firstly by including time-dependent covariates in the location parameter (see the next two paragraphs for details) and secondly by choosing the blocks to be months, which yields an individual threshold for each

year-month-combination and also allows for non-stationarity effects other than seasonality.

The  $r$ -largest order statistics model implemented has location parameter  $\mu(x, t)$ , scale parameter  $\sigma(x)$ , and shape parameter  $\xi(x)$ . All parameters depend on grid cell  $x$  because an individual model is fitted to every grid cell, and the additional time-dependence in the location parameter accounts for temporal non-stationarity. As mentioned before, non-stationarity is expressed by a multiple linear regression term, i.e.,

$$\mu(x, t) = \mathbf{Z}(t)\boldsymbol{\beta}(x), \quad (3)$$

where  $\mathbf{Z}(t)$  is a design matrix of purely time-dependent covariates, the same for every grid cell  $x$ , each of which has its own coefficient vector  $\boldsymbol{\beta}(x)$  due to pointwise modeling. The design matrix  $\mathbf{Z}(t)$  is composed of seasonality terms and covariates describing chemical and dynamical processes in the atmosphere (see Eq. (4) below and Fig. 1). All covariates are taken at a monthly resolution, so the same covariate value is used for the  $r$  observations of each block, i.e., of each year-month-combination.

The use of the same design matrix for all grid cells, even though in principle several covariates vary across space, has the advantage that maps of the parameter estimates  $\{\hat{\beta}_i(x) : x \in \mathcal{X}\}$  for each covariate  $i$  are interpretable and comparable. Consequently, an optimal set of covariates must be selected for the whole region of interest in each hemisphere. Since fitting the model to the complete grid requires much computation time, it is infeasible to fit many different models for each grid cell. Instead we identified the most promising set of covariates on a subset of 72 grid cells, using standard model selection techniques. The form of Eq. (3) finally used is

$$\begin{aligned} \mathbf{Z}(t)\boldsymbol{\beta}(x) = & \beta_0(x) + \sum_{i=1}^3 \left( \beta_{2i-1}(x) \cos\left(\frac{2\pi t}{\varphi_i}\right) \right. \\ & \left. + \beta_{2i}(x) \sin\left(\frac{2\pi t}{\varphi_i}\right) \right) + \beta_7(x) \text{EESC}(t) \\ & + \beta_8(x) \text{SOLAR}(t) + \beta_9(x) \text{QBO}_{30}(t) \\ & + \beta_{10}(x) \text{QBO}_{50}(t) + \beta_{11}(x) \text{ENSO}(t) \\ & + \beta_{12}(x) \begin{Bmatrix} \text{AAO}(t) \\ \text{NAO}(t) \end{Bmatrix} + \beta_{13}(x) \text{CHICHON}(t) \\ & + \beta_{14}(x) \text{PINATUBO}(t), \end{aligned} \quad (4)$$

where the  $\varphi_i$  allow for 1-yr, 6-month, and 4-month seasonalities, i.e.,  $(\varphi_1, \varphi_2, \varphi_3) = (12, 6, 4)$ . The other covariates in Eq. (4) are listed in Sect. 2.2. The AAO and NAO are only used for the Southern and the Northern Hemisphere, respectively; and the covariates for the volcanic eruptions of El Chichón and Mt. Pinatubo are extracted from the Sato-Index for the corresponding hemisphere (see Fig. 1). A detailed description of these covariates and their interaction with the ozone levels is given in Sect. 4.

The model with location parameter  $\mu$  given in Eq. (4) will be referred to as the *annual model* because each of the re-

gression parameters  $\beta_0, \dots, \beta_{14}$  expresses the effect of its respective covariate over the whole year. Since the influence of the covariates is likely to vary across the seasons, we fitted a second *seasonal model* in which some covariates are split into four pieces, one for each season defined by December-January-February, March-April-May, June-July-August, and September-October-November. For instance the term  $\beta_7(x)\text{EESC}(t)$  in Eq. (4) can be replaced by

$$\begin{aligned} & \beta_{7, \text{spring}}(x) \text{EESC}_{\text{spring}}(t) + \beta_{7, \text{summer}}(x) \text{EESC}_{\text{summer}}(t) \\ & + \beta_{7, \text{fall}}(x) \text{EESC}_{\text{fall}}(t) + \beta_{7, \text{winter}}(x) \text{EESC}_{\text{winter}}(t), \end{aligned}$$

where  $\text{EESC}_{\diamond}$  equals the EESC values for the months of season  $\diamond$  and equals zero for the other months. In this study EESC, SOLAR, QBO<sub>30</sub> and QBO<sub>50</sub>, ENSO, and AAO/NAO were split into seasonal components. The annual model can be nested into the seasonal one by setting  $\beta_{\cdot, \text{spring}} = \beta_{\cdot, \text{summer}} = \beta_{\cdot, \text{fall}} = \beta_{\cdot, \text{winter}}$ , so likelihood ratio tests allow one to assess whether these covariates have varying impacts over the year (see Sect. 4.1).

### 3.2 ARMA model for total ozone mean values

In analogy to our model for extremes, we fit a univariate model to the monthly mean values of total ozone at each grid cell, even though spatial models for large Gaussian sets have been much more widely investigated than their analogues for extremes and hence would be more easily available.

The multiple linear regression model, a standard tool for analyzing the contributions of individual covariates to changes in a variable of interest (e.g., total ozone), is widely used in atmospheric science. In a first step we therefore fitted a multiple linear regression model in its simplest form, i.e., with independent errors, to the ozone mean values at both northern and southern mid-latitudes. However, the residuals of this model were significantly correlated, showing that the assumption of independent errors does not hold for our data. Autocorrelation in model residuals of total ozone has been reported in previous studies. Vyushin et al. (2007), for example, use a model with autocorrelated noise for total ozone, and von Clarmann et al. (2010) also suggest accounting for correlations when analyzing atmospheric data.

In this study, we address temporal correlation using a classic model for stationary time series, the autoregressive moving average (ARMA) process (e.g., Brockwell and Davis, 2002). The strengths of the ARMA model are its flexibility and parsimony, as a wide range of temporal correlation patterns can be expressed with only few parameters. A stationary sequence of random variables  $\{W_t\}_{t \in \mathbb{T}}$  is an ARMA process of order  $(p, q)$  if it can be written as

$$W_t = \sum_{i=1}^p \phi_i W_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j},$$

where  $\{\varepsilon_t\}_{t \in \mathbb{T}}$  is white noise.

To express non-stationarity and assess the effect of the covariates on the mean values of total ozone at grid cell  $x \in \mathcal{X}$ , we retain the linear term  $\mathbf{Z}(t)\boldsymbol{\beta}(x)$  from Eq. (3). More precisely, the design matrix  $\mathbf{Z}(t)$  is identical with Eq. (4), and the coefficient vector  $\boldsymbol{\beta}(x)$  has the same structure as in the extremes model but takes unrelated values. If  $Y(x, t)$  denotes the mean value at cell  $x \in \mathcal{X}$  for year-month-combination  $t$ , the stationarized process

$$\eta(x, t) = Y(x, t) - \mathbf{Z}(t)\boldsymbol{\beta}(x), \quad (5)$$

is an ARMA( $p, q$ ) process with orders  $p$  and  $q$  to be determined (see next paragraph).

Since the fitting of an ARMA model is much faster than for the  $r$ -largest order statistics model, we fitted several ARMA models with different orders ( $p, q$ ) and compared their AIC (Akaike Information Criterion, see Sect. A4). There is no uniformly best model over all grid cells but in all cases (north, south, annual, seasonal) the ARMA(1,1) model performs adequately in a majority of grid cells, so that  $\eta$  in Eq. (5) becomes

$$\eta(x, t) = \phi(x)\eta(x, t-1) + \varepsilon(x, t) + \theta(x)\varepsilon(x, t-1),$$

where  $\varepsilon(x, t)$  is taken to be Gaussian white noise, which is more appropriate for mean values than for extremes.

Exactly as in the extremes model (see Sect. 3.1) we fitted two ARMA models with different regression terms, an *annual* model, where only one coefficient is used for each covariate, and a *seasonal* model, where some of the covariates are split into four seasons (see the last paragraph of Sect. 3.1.2).

## 4 Results

Inference about the effect of a single covariate based on pointwise modeling of the grid cells may be performed by testing, for each cell, the hypothesis that the covariate has no effect, but interpreting the results is not straightforward, because of the large number of tests that must be performed, based on highly correlated data. If the same false positive rate  $\alpha$  were used for every cell, and if all the null hypotheses were true, then we would expect a proportion  $\alpha$  of them to be falsely rejected. Thus with 8640 cells and with  $\alpha = 0.05$ , we would expect to incorrectly find a significant effect of the covariate at 432 cells, and these cells will tend to be grouped owing to spatial dependence. Ventura et al. (2004) discuss approaches to multiple testing that control the false discovery rate (FDR), i.e., the expected proportion of falsely rejected null hypotheses among the total number of rejected null hypotheses. Using the FDR is attractive because in practice the number of rejected null hypotheses is known but the number of true null hypotheses is not. The basic FDR approach assumes independence of the  $p$ -values, but Ventura et al. (2004) suggest that it can also be used for data

with light spatial correlation. Unfortunately, and as would be anticipated, the NIWA data show very strong correlations, and this undermines the case for using the basic FDR approach. Another method often used in climate science, the field significance method (Livezey and Chen, 1983), seems inappropriate in the setting of extremes. The topic of multiple testing is undergoing rapid development and a definitive treatment in the present context cannot yet be provided, so in order to assess the strength of the conclusions below, we applied four approaches to the  $z$ -statistics for the covariate effects on extremes: (i) making no correction for multiple testing; (ii) false discovery rate (FDR); (iii) a conservative version of FDR that allows for general correlation in the  $z$ -statistics; and (iv) the ultra-conservative Bonferroni correction (see Sect. A5 for a short explanation). We also used various scenarios with inflated standard errors to account for clustering in the extremes. We comment briefly on the results of this sensitivity analysis when discussing the specific effects below, and in the companion paper to this article.

### 4.1 Evaluation of the statistical models

This section briefly discusses two aspects of the evaluation of the statistical models described in Sect. 3, namely the comparison between the annual and seasonal models followed by diagnostics for the models of extremes and mean values.

#### 4.1.1 Comparison between annual and seasonal model

Since the annual model is nested within the seasonal one (see the end of Sect. 3.1.2), these two models can be compared with a likelihood ratio test (see Sect. A3 for an explanation). In the present context, the null hypothesis  $H_0$  is that the simpler annual model fits the data well, i.e., the seasonal model brings no improvement. Performed for every grid cell and every model separately, these tests show that the use of the more complicated seasonal model is justified for both the EVT and the ARMA model, since  $H_0$  is rejected for a majority of grid cells (see Table 2 for a complete summary). For the  $r$ -largest order statistics model the proportions of grid cells where  $H_0$  is rejected at the 5%-level lie far above 80% for both hemispheres and types of extremes, whereas they are lower but still above 50% for the ARMA model fitted to the means. This reduction in the proportion of rejections is probably due to the better incorporation of correlation between the observations in the ARMA model compared to the  $r$ -largest order statistics model, where the blocks are assumed to be independent. Hence the proportions obtained for the EVT models are likely to be too high.

#### 4.1.2 Model validation

Checking the goodness of fit of a statistical model usually involves plots of the residuals or the fitted values. In the present context, both depend not only on space but also on time, making model validation on the complete grid intractable.



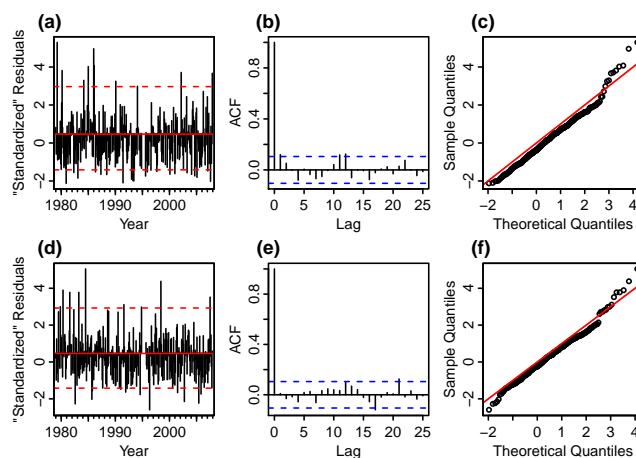
**Table 2.** Summary of the likelihood ratio tests between the annual (null hypothesis  $H_0$ ) and seasonal models for EHOs, ELOs, and mean values (MV) at northern (NM) and southern (SM) mid-latitudes. The column CN indicates the number of grid cells out of 8640 for which  $H_0$  is rejected at the 5 %-level and the column  $R$  gives the proportion of these grid cells (i.e.,  $R = \text{CN}/8640$ ).

Test level		
$\alpha = 5\%$	CN	$R$
EHOs NM	8277	95.8 %
ELOs NM	8632	99.9 %
MV NM	5673	65.7 %
EHOs SM	7581	87.7 %
ELOs SM	8345	96.6 %
MV SM	4713	54.5 %

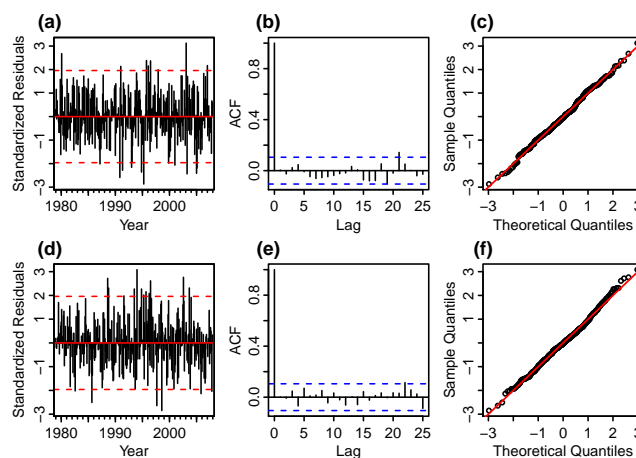
Therefore we provide diagnostic plots for two representative grid cells (one in each hemisphere), using the seasonal models for EHOs in Fig. 2 and those for mean values in Fig. 3. Overall the fit seems reasonable, despite a small bias in the residuals for the EVT model (Fig. 2c, f); this, however, only affects the intercept of our model, and not the regression coefficients of the covariates. A comparison with the same model for  $r = 1$  showed that estimates of the covariate coefficients from the two models are consistent, and diagnostic plots of the model with  $r = 1$  show a good fit and no bias, indicating no fundamental trouble with the model when  $r > 1$ . Simulations suggest that this bias stems from temporal correlation in the underlying time series. In the residuals of both models, however, autocorrelation has practically gone (panels (b) and (e) in Figs. 2 and 3), even though the correlations at some lags are slightly outside the approximate confidence bounds. For the mean values this is expected, as modeling temporal correlation is the main purpose of the fitted ARMA model. The extreme events may arise from more distant observations and hence tend to be less correlated than the underlying daily values.

## 4.2 Spatial patterns

Our modeling approach allows separate analyses of the influence of the individual covariates (see Fig. 1 and Sect. 2.1) on EHOs, ELOs, and mean values. For analytical and illustrative reasons it is convenient to plot maps of the regression coefficient estimates for the covariates, their standard errors, and the  $p$ -values of the likelihood-based  $z$ -test for significance on a grid cell basis; although our analysis provides additional information about the distribution of ozone extremes at each grid cell, such information is beyond the scope of the present study. For lack of space, we only show the maps for estimates and  $p$ -values in this paper, but provide augmented versions of the figures including standard errors in a supplement. The  $p$ -values are useful to assess the significance of a



**Fig. 2.** Diagnostic plots of the extremes model with seasonal covariates for EHOs at two sample grid cells (a)–(c)  $45.5^\circ \text{N}$ ,  $93^\circ \text{W}$  and (d)–(f)  $45.5^\circ \text{S}$ ,  $138^\circ \text{E}$ . Panels (a) and (d) show “standardized” residuals  $(y(t) - \mu(t))/\sigma$  with 95 %-confidence bounds; panels (b) and (e) their correlograms and panels (c) and (f) the GEV(0,1, $\xi$ ) Q-Q plots with the identity line in red.

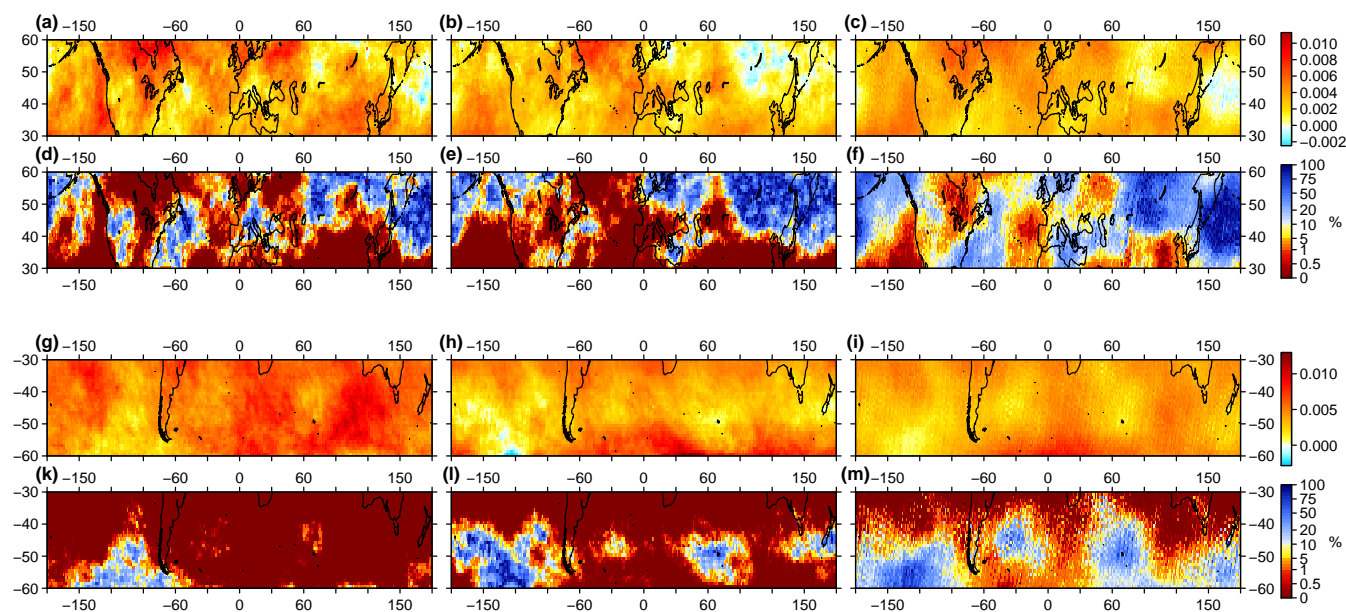


**Fig. 3.** Diagnostic plots of the ARMA(1,1) model with seasonal covariates for monthly mean values of column ozone at two sample grid cells (a)–(c)  $45.5^\circ \text{N}$ ,  $93^\circ \text{W}$  and (d)–(f)  $45.5^\circ \text{S}$ ,  $138^\circ \text{E}$ . Panels (a) and (d) show the standardized residuals with 95 %-confidence bounds; panels (b) and (e) their correlograms and panels (c) and (f) the normal Q-Q plots with the identity line in red.

covariate at a grid cell. We consider a covariate to be significant whenever the corresponding  $p$ -value is less than 1 %, bearing in mind that the calculated  $p$ -values are probably too small, owing to the temporal correlation in the extremes.

Below we focus on spatial patterns in the three “standard” covariates included in analyses of long-term ozone changes (e.g., in the WMO/UNEP Ozone Assessment Reports: WMO, 2003, 2007, 2011), i.e., the solar cycle, QBO, and EESC, and on additional frequently used dynamical covariates, namely the NAO and AAO indices. A companion





**Fig. 4.** Pointwise regression coefficient estimates (in  $\text{DU } 10^{22} \text{ W}^{-1} \text{ m}^2 \text{ Hz}$ ) for the solar cycle on an annual basis for (a) EHOs, (b) ELOs, and (c) mean values of total ozone at  $30^\circ \text{ N}$  to  $60^\circ \text{ N}$ ; (d)–(f) show the  $p$ -values of pointwise significance tests for the estimates in (a)–(c). (g)–(m) as (a)–(f) but at  $30^\circ \text{ S}$  to  $60^\circ \text{ S}$ . An augmented version of this figure including standard errors for all coefficient estimates is provided in Fig. S1.

paper (Rieder et al., 2013, from here on referred to as Part 2) gives a detailed discussion on the spatial fingerprints of the volcanic eruptions of El Chichón and Mt. Pinatubo and of the fingerprints of the El Niño/Southern Oscillation, on which scientific interest has recently focused. Part 2 also reports on the contribution of the individual covariates to long-term total ozone changes for selected regions of interest at northern and southern mid-latitudes.

#### 4.2.1 Solar cycle

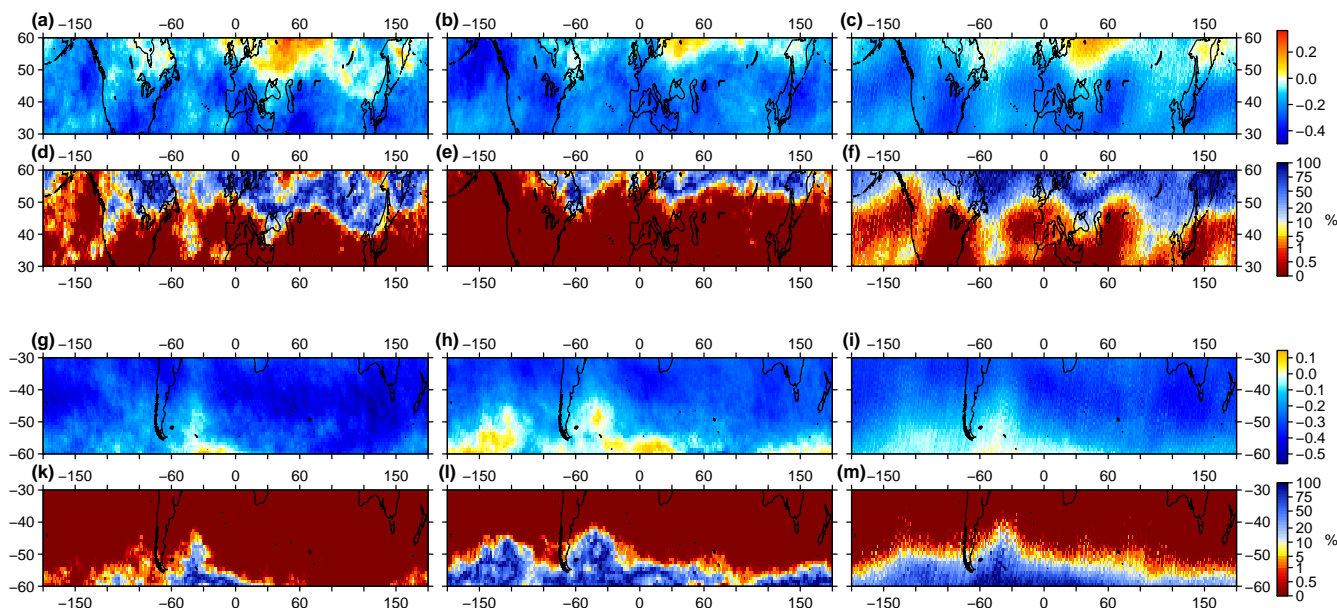
Solar variability described by the 11-yr solar cycle, describing changes in solar irradiance through changes in sunspot number (see Fig. 1a), influences stratospheric ozone, because UV-radiation varies with an amplitude of 6–8 % between solar maxima and minima (e.g., Chandra and McPeters, 1994). Previous studies found that at mid-latitudes about 2 % of total ozone variability can be explained through changes in the solar cycle. However, long-term trends in column ozone over the last decades cannot be explained by solar variability (e.g., Harris et al., 2008; Chandra and McPeters, 1994).

The  $p$ -values show that the solar cycle seems to be significant on an annual basis for much of the northern and southern mid-latitudes (see Fig. 4), in general agreement with earlier studies (e.g., Steinbrecht et al., 2006). However, two things are important to note: the area showing highly significant influence of the solar cycle is much larger for extremes than for mean values; and on a seasonal basis, high variability is found in the spatial extent of the significance area (not shown

here), as during spring and winter significant influence of the solar cycle is restricted towards lower latitudes, in agreement with the strong influence of the solar cycle on ozone production in the tropical region. Sensitivity analysis shows that the effect persists when any of the corrections for multiple testing are applied, though the region of significance becomes smaller.

#### 4.2.2 Quasi-Biennial Oscillation (QBO)

The QBO dominates the variability of the equatorial stratosphere. Seen as a composite of equatorial zonal winds, it shows faster and more regular downward propagation during the westerly phase, and stronger intensity and longer duration during the easterly phase. The mean period of the QBO is about 28 months. Maxima in the variability are larger during the westerly than the easterly phase and are found close to the descending easterly and westerly shear zones (e.g., Baldwin et al., 2001). Although the QBO is a tropical phenomenon, it affects stratospheric flow from pole to pole due to modulation of the effects of extra-tropical waves. Connection between the QBO and the extra-tropical atmosphere (e.g., mid-latitudes) must be seen in the context of the seasonal cycle and variability of the extra-tropical stratosphere. During winter, the high-latitude stratosphere cools and a deep westerly polar vortex is formed. During spring and summer, the vortex diminishes and the westerlies are replaced by easterlies due to increased solar heating. As the Northern Hemisphere has much greater land-mass than the



**Fig. 5.** Pointwise regression coefficient estimates (in  $\text{DU m}^{-1} \text{s}$ ) for the Quasi-Biennial Oscillation at 30 hPa on an annual basis for (a) EHOs, (b) ELOs, and (c) mean values of total ozone at  $30^\circ \text{N}$  to  $60^\circ \text{N}$ ; (d)–(f) show the  $p$ -values of pointwise significance tests for the estimates in (a)–(c). (g)–(m) as (a)–(f) but at  $30^\circ \text{S}$  to  $60^\circ \text{S}$ . An augmented version of this figure including standard errors for all coefficient estimates is provided in Fig. S2.

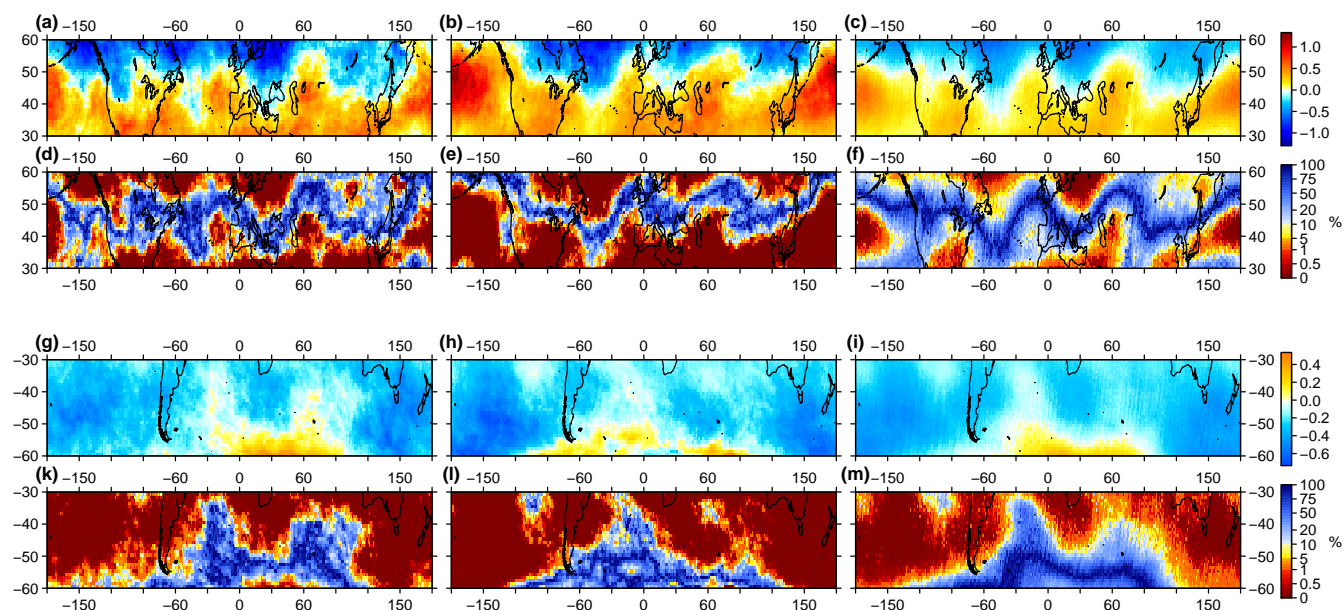
Southern Hemisphere, tropospheric waves have larger amplitudes in the Northern Hemisphere. Therefore, the northern hemispheric winter stratosphere is much more disturbed than its southern counterpart. Consequently the northern polar vortex can already be disrupted by large-scale planetary waves in mid-winter, when the exchange of westerlies with easterlies in high latitudes causes strong sudden stratospheric warming events (e.g., Waugh and Randel, 1999). Further, various studies have also described an influence of the QBO on other constituents of the atmosphere, such as methane and water vapor (e.g., Baldwin et al., 2001).

In this study, the QBO at two different pressure levels (30 and 50 hPa, with no lag) (see Fig. 1b, c) was used as a covariate in the spatial models. The  $p$ -values in Figs. 5 and 6 show that the QBO at both pressure levels seems to be significant over large areas of the northern and southern mid-latitudes. Coefficient estimates for both  $\text{QBO}_{30}$  and  $\text{QBO}_{50}$  are highly significant towards the equatorial regions of the northern and southern mid-latitudes. At northern mid-latitudes the coefficient estimates for  $\text{QBO}_{30}$  show a smooth staggered gradient towards high latitudes, while the coefficient estimates for  $\text{QBO}_{50}$  show a band-like structure where regions towards low and high latitudes with significant effects are split by a band with insignificant effects. While coefficient estimates are positive for lower latitudes, they turn negative after the transition zone towards polar latitudes. This structure is possibly related to a connection between the QBO and the Brewer-Dobson circulation, as Haklander et al. (2006) showed that the mean zonal wind pattern can alter

the wave driving of the Brewer-Dobson circulation. This second significance region towards northern polar regions, identified in our analysis, is in good agreement with earlier work discussing the influence of the QBO on column ozone at high latitudes (e.g., Oltmans and London, 1982; Garcia and Solomon, 1987; Lait et al., 1989; Randel and Cobb, 1994; Baldwin et al., 2001). The effects of  $\text{QBO}_{30}$  are more important than those of  $\text{QBO}_{50}$ , with the latter becoming much less striking under sensitivity analysis.

### 4.2.3 Ozone depleting substances (ODS)

Anthropogenic emissions of ODS (ozone depleting substances such as chlorofluorocarbons) increased from the early 1950s until the late 1980s, when the Montreal Protocol was signed. The scale of EESC (equivalent effective stratospheric chlorine) describes the effect of stratospheric ozone depletion caused by anthropogenic emissions of ODS in an integral way (note that we use the EESC scale for mid-latitudes here). EESC peaked in the second half of the 1990s (see Fig. 1i) because the transport of EESC from the release of ODS near surface into the stratosphere causes a shift between maximal emissions of ODS and EESC. Chemical ozone depletion is particularly large during the winter and spring seasons when ozone destruction occurs inside the polar vortex, caused by heterogeneous chemical reactions taking place on polar stratospheric clouds (e.g., Peter, 1997; Solomon, 1999). Apart from the atmospheric burden in ODS,



**Fig. 6.** Pointwise regression coefficient estimates (in  $\text{DU m}^{-1} \text{s}$ ) for the Quasi-Biennial Oscillation at 50 hPa on an annual basis for (a) EHOs, (b) ELOs, and (c) mean values of total ozone at  $30^\circ \text{N}$  to  $60^\circ \text{N}$ ; (d)–(f) show the  $p$ -values of pointwise significance tests for the estimates in (a)–(c). (g)–(m) as (a)–(f) but at  $30^\circ \text{S}$  to  $60^\circ \text{S}$ . An augmented version of this figure including standard errors for all coefficient estimates is provided in Fig. S3.

stratospheric temperature is the main driver of ozone loss within the polar vortex (e.g., Rex et al., 2004).

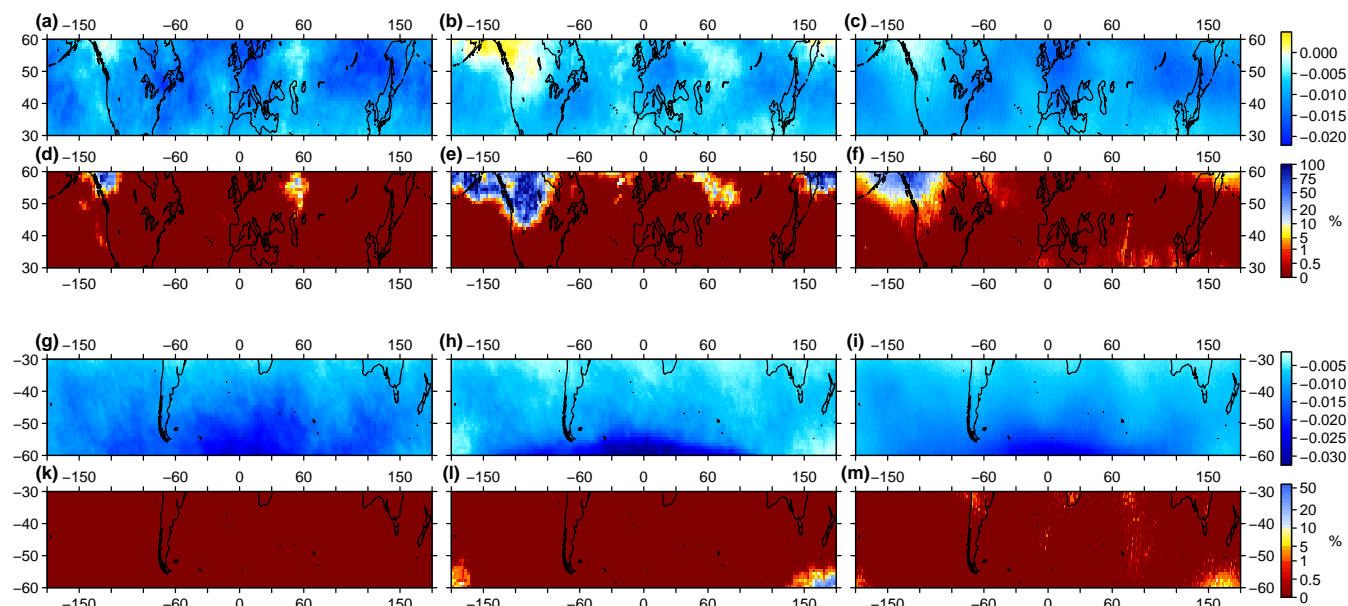
At northern mid-latitudes the relation between EESC and column ozone is negative and significant in almost all areas (see Fig. 7), except in a region north to  $50^\circ \text{N}$  and around  $120^\circ \text{W}$  where EESC is barely significant, probably because dynamic variability disturbs the relation between column ozone and EESC. In our sensitivity analysis, EESC remains the strongest covariate under any multiple testing corrections. Coefficient estimates in Fig. 7 for both extremes and mean values show a gradient increasing towards high latitudes. At southern mid-latitudes, the influence of ODS shows as a “stable staggered” gradient in coefficient estimates which may result from the less disturbed atmospheric flow due to reduced land-mass.

#### 4.2.4 North Atlantic Oscillation (NAO)

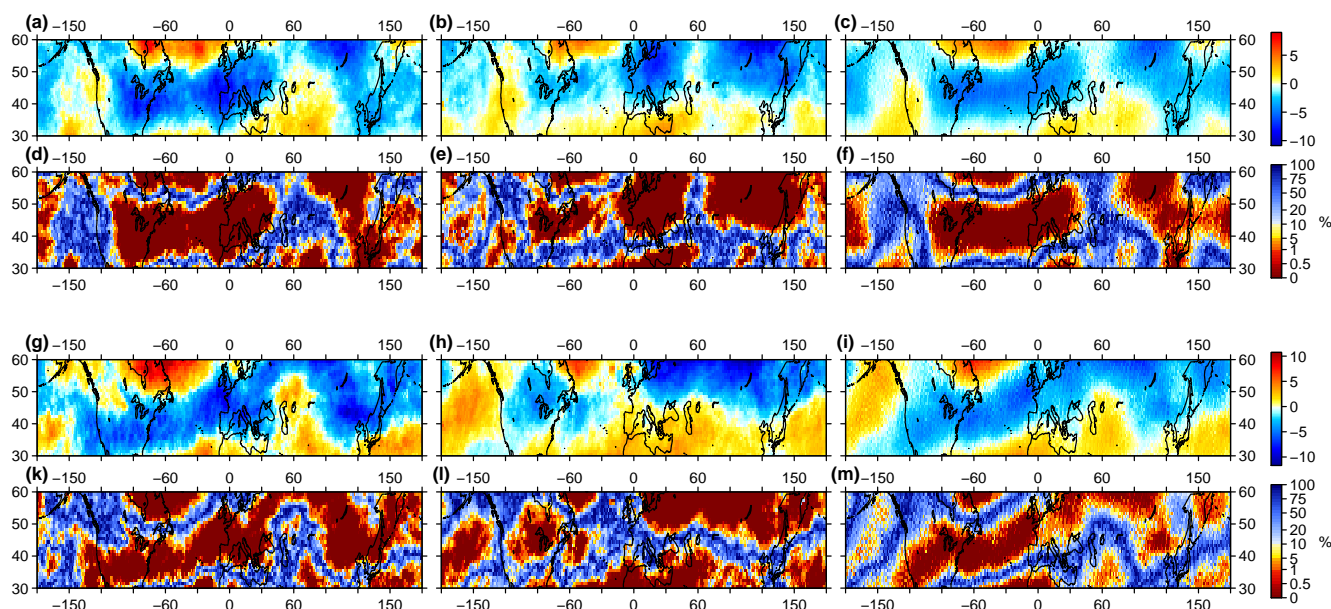
Inter-annual and decadal changes in Northern Hemisphere tropospheric meteorology and stratospheric dynamics are strongly related to the variability in the North Atlantic (NAO) and the Arctic Oscillation (AO). Several studies have shown that the NAO affects changes in the direction and intensity of the dominant westerly tropospheric jet stream (e.g., Orsolini and Limpasuvan, 2001) and thereby influences European winter/spring climate and the strength of the Arctic polar vortex affecting the stratospheric ozone layer (e.g., Appenzeller et al., 2000; Thompson and Wallace, 2000; Orsolini and Limpasuvan, 2001; Hadjinicolaou et al., 2002; Orsolini

and Doblas-Reyes, 2003). Here we use a NAO index following Hurrell (2009) (see Fig. 1e); it is built by the principal components of the leading empirically-determined orthogonal function of seasonal sea level pressure anomalies over the Atlantic sector (defined as:  $20\text{--}80^\circ \text{N}$ ,  $90^\circ \text{W}\text{--}40^\circ \text{E}$ ).

The coefficient estimates for the NAO, in winter and spring (see Fig. 8), are found to be significant for much of the northern mid-latitudes, and this significance varies but persists with any of the multiple testing corrections. Largest positive coefficients are found over Labrador/Greenland, the North Atlantic sector, and over the Norwegian Sea, while largest negative coefficient estimates are found over Europe, Russia, and the Eastern United States. While regions with positive coefficient estimates will show increased column ozone during its positive phases, the converse is true for regions with negative coefficient estimates (see also the contributions of the NAO to long-term ozone changes at different regions presented in Part 2). This relation between column ozone and the mode of the NAO can be explained by pressure gradients, which are increased during positive phases of the NAO due to a deeper than usual Icelandic low and a stronger than usual sub-tropical high pressure system. This increased pressure gradient results in more and stronger winter storms crossing the Atlantic and a shift of storm tracks towards the north. During the negative phase of the NAO the converse occurs, i.e., a weaker sub-tropical high and Icelandic low lead to a reduced pressure gradient, and therefore to fewer and weaker winter storms on a more west-east pathway. Compared to

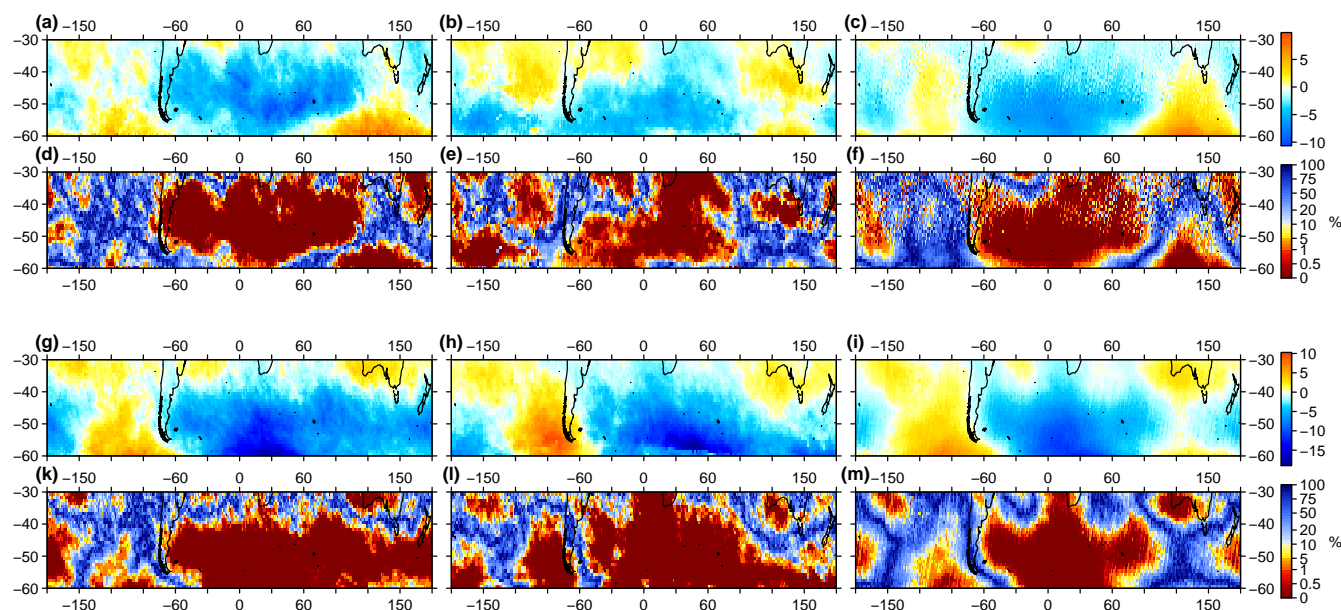


**Fig. 7.** Pointwise regression coefficient estimates (in  $\text{DU ppt}^{-1}$ ) for ozone depleting substances in terms of equivalent effective stratospheric chlorine (EESC) on an annual basis for (a) EHOs, (b) ELOs, and (c) mean values of total ozone at  $30^\circ\text{N}$  to  $60^\circ\text{N}$ ; (d)–(f) show the  $p$ -values of pointwise significance tests for the estimates in (a)–(c). (g)–(m) as (a)–(f) but at  $30^\circ\text{S}$  to  $60^\circ\text{S}$ . An augmented version of this figure including standard errors for all coefficient estimates is provided in Fig. S4.



**Fig. 8.** Pointwise regression coefficient estimates (in  $\text{DU (unit NAO)}^{-1}$ ) for the North Atlantic Oscillation (NAO) for (a) EHOs, (b) ELOs, and (c) mean values of total ozone during winter (DJF) at  $30^\circ\text{N}$  to  $60^\circ\text{N}$ ; (d)–(f) show the  $p$ -values of pointwise significance tests for the estimates in (a)–(c). (g)–(m) as (a)–(f) but during spring (MAM). An augmented version of this figure including standard errors for all coefficient estimates is provided in Fig. S5.





**Fig. 9.** Pointwise regression coefficient estimates (in  $\text{DU}(\text{unit AAO})^{-1}$ ) for the Antarctic Oscillation (AAO) for (a) EHOs, (b) ELOs, and (c) mean values of total ozone during winter (JJA) at  $30^\circ\text{S}$  to  $60^\circ\text{S}$ ; (d)–(f) show the  $p$ -values of pointwise significance tests for the estimates in (a)–(c). (g)–(m) as (a)–(f) but during spring (SON). An augmented version of this figure including standard errors for all coefficient estimates is provided in Fig. S6.

other covariates discussed, the NAO “fingerprint” is of similar spatial extent for both extremes and mean values, but the magnitude of influence on total ozone is larger for EHOs and ELOs than for mean values, confirming the importance of atmospheric dynamics for total ozone variability and changes (see Part 2 and Rieder et al., 2010a,b, 2011).

#### 4.2.5 Antarctic Oscillation (AAO)

The semi-annual oscillation (SAO) at mid and high latitudes in the Southern Hemisphere is related to the depth of the troposphere and to the weakening and expansion of the circumpolar vortex of low pressure surrounding Antarctica from March–June and September–December (e.g., van Loon, 1967, 1972).

During the high phase of the Antarctic Oscillation (AAO), the Lagrangian mean circulation, responsible for the transport of ozone from the tropics to the polar region, is strongly reduced. Wave refraction triggers this process: during the high phase, the polar vortex refracts more wave activity in the tropics and breaking of these waves strengthens the vortex due to the inside transport of momentum (e.g., Thompson and Wallace, 2000). During the low phase of the AAO the opposite occurs: a weaker vortex decelerates more when waves are defracting in.

As for the NAO in the Northern Hemisphere, the contribution of the AAO to extremes and mean values seems to be highly significant over large parts of the southern mid-latitudes (see Fig. 9). Results of the sensitivity analy-

sis for the AAO are similar to those for the NAO, confirming the presence of the effects with any of the multiple testing corrections. Interestingly the central southern mid-latitudes show especially highly significant (negative) coefficient estimates for the AAO. This may be related to enhanced wave activity in the tropics (see also Schnadt Poberaj et al., 2011), leading to enhanced ozone transport from the tropics to the extra-tropics (compare also with results for ENSO shown in Part 2) and a strengthening of the southern ozone “collar”. The AAO was found to have also significant influence on the dynamical masking of the Mt. Pinatubo eruption at southern mid-latitudes, which is discussed in detail in Part 2.

## 5 Discussion and conclusions

In this study statistical models, including important covariates describing the state of the atmosphere (solar cycle, QBO, ENSO, NAO, AAO, EESC, and volcanic eruptions), have been used to analyze changes in extreme values (EHOs and ELOs using the  $r$ -largest order statistics model) and mean values (using an ARMA model) of total ozone at northern and southern mid-latitudes. The results show that “fingerprints” of dynamical and chemical features are captured in both the “bulk” and the tails of the distribution of total ozone time series. However, “fingerprints” of atmospheric dynamics (NAO, AAO) are better represented in the extremes and can be partly overlooked in analysis of the mean values alone. This confirms results from earlier local/regional studies (Rieder et al., 2010b, 2011).

For the three “standard” covariates – solar cycle, QBO, and EESC – significant influence was found almost throughout the northern and southern mid-latitudes, in good agreement with previous studies (e.g., Steinbrecht et al., 2006; WMO, 2003, 2007, 2011). However, several important features should be noted: (i) regarding the “fingerprints” of the solar cycle, the area highly significantly influenced is much larger for extremes than for mean values and there is large seasonal variability (not shown here); (ii) for the QBO at both pressure levels analyzed (30 and 50 hPa) large significant areas are found at both northern and southern mid-latitudes, especially towards equatorial regions. An interesting feature is the band-like structure found for QBO<sub>50</sub>: an insignificant band at central mid-latitudes splits two highly significant areas towards equatorial and polar regions. This structure might be related to a connection between the QBO and the Brewer–Dobson circulation at this pressure level, where zonal winds can alter wave driving (e.g., Haklander et al., 2006); (iii) there is strong influence of EESC on ozone throughout mid-latitudes. At northern mid-latitudes a gradient is found in EESC-coefficients, increasing towards high latitudes, which can be interpreted as resulting from the interaction between ozone production, ozone transport to mid and high latitudes (due to the Brewer–Dobson circulation), and enhanced ozone depletion in polar regions (especially during winter and early spring). This pattern is less pronounced (but still preserved) at southern mid-latitudes, probably due to the reduced land-mass leading to less disturbed atmospheric flow.

For the North Atlantic Oscillation, strong influence on column ozone is found over Labrador/Greenland, the Eastern United States, the Euro-Atlantic sector, and Central Europe. For the NAO’s southern counterpart, the AAO, strong influence on column ozone is found at lower southern mid-latitudes, including the southern parts of South America and the Antarctic Peninsula, and central southern mid-latitudes. At central southern mid-latitudes highly significant negative coefficient estimates found for the AAO can probably be related to enhanced wave activity in the tropics leading to enhanced ozone transport from the tropics to the extra-tropics. Results for both NAO and AAO confirm the importance of atmospheric dynamics for ozone variability and changes from local/regional to global scale.

We refer to the companion paper (Rieder et al., 2013) for the spatial analysis of fingerprints of the volcanic eruptions of El Chichón and Mt. Pinatubo and the El Niño/Southern Oscillation. There we discuss the important role of dynamical covariates (AAO and ENSO) on amplifying/weakening the effect of volcanic eruptions at southern mid-latitudes and provide a detailed overview of the contribution of the individual covariates to long-term total ozone changes (1979–2007), in the mean and extreme values, for various regions of interest.

## Appendix A

### Statistical terms

This section provides an overview of statistical topics relevant to this paper like maximum likelihood estimation, associated hypothesis tests on models and parameters, and the Bonferroni correction for multiple testing. Being far from complete, it merely provides a brief overview. Apart from the Bonferroni correction, these topics all appear in Davison (2003), and a more accessible reference is Coles (2001, Ch. 2). Wilks (2011) covers a broad range of statistical topics in atmospheric sciences, thereunder the Bonferroni correction in Sect. 11.5 and multiple testing in Sect. 5.4.

#### A1 Likelihood estimation

Let  $Y$  be a random vector whose distribution is supposed to be known up to a parameter vector  $\theta$  belonging to a  $p$ -dimensional parameter space with  $p \geq 1$ . The joint density of  $Y$ , denoted  $f_Y(y; \theta)$ , depends on the observation  $y$  and on the parameter vector  $\theta$ , which is unknown. The *likelihood* for  $\theta$  based on the observation  $y$  is defined as  $L(\theta) = f_Y(y; \theta)$ , i.e., the joint density of the observation as a function of the unknown parameter. If the components of  $Y$  are independent and identically distributed random variables, the joint density or the likelihood simplifies to the product of the individual densities.

The *maximum likelihood estimate* of  $\theta$  is the value  $\hat{\theta}$  that maximizes the likelihood  $L(\theta)$ , i.e., the value of  $\theta$  under which the observed outcome  $y$  is most probable. Equivalently,  $\hat{\theta}$  also maximises the logarithm of the likelihood, the so-called log-likelihood. Because sums are easier to handle than products, one generally prefers to work with the log-likelihood. In practice, either  $L$  or  $\log(L)$  is maximized numerically.

To investigate its mathematical properties,  $\hat{\theta}$  has to be regarded as a function of the underlying random vector  $Y$  rather than of the observed data  $y$ . In this case, it is itself a random vector with a probability distribution and therefore called an *estimator*, while the estimate is an “observation” of the estimator. Under some regularity conditions on the likelihood function and the parameter space (Davison, 2003, p. 118), the distribution of the maximum likelihood estimator (MLE)  $\hat{\theta}$  for increasing sample size  $n$  is known. Using this asymptotic result as an approximation for finite samples, the MLE  $\hat{\theta}$  has approximately a Gaussian distribution with mean the true value of  $\theta$  and variance matrix  $\mathbf{J}(\hat{\theta})^{-1}$ , where  $\mathbf{J}(\hat{\theta})$  is the  $p \times p$  matrix of the negative second derivatives of  $\log(L)$ . Since  $\mathbf{J}(\hat{\theta})$  may be computed numerically during the optimization procedure, standard errors  $SE(\hat{\theta}_i)$  for the components  $\hat{\theta}_i$  of  $\hat{\theta}$  are easily produced.

In this work, we estimate the parameters of our statistical models for total ozone by maximum likelihood estimation

and compute standard errors for these estimates as described here.

## A2 $z$ -test and $p$ -value

The distributional result for the MLE given in the previous section implies that the expression  $z_i = (\hat{\theta}_i - \theta_i)/SE(\hat{\theta}_i)$  has a standard Gaussian distribution. This can be used to test the null hypothesis that the unknown parameter  $\theta_i$ , the  $i$ -th component of  $\theta$ , takes a given value  $\theta_i^0$ , i.e.,  $H_0: \theta_i = \theta_i^0$ . Next, we illustrate the principle of hypothesis testing for this so-called  $z$ -test.

If  $H_0$  is true,  $z_{i,0} = (\hat{\theta}_i - \theta_i^0)/SE(\hat{\theta}_i)$  has a standard Gaussian distribution. If  $H_0$  is clearly not true, meaning that the unknown true value  $\theta_i$  is far from  $\theta_i^0$ , the absolute value  $|z_{i,0}|$  is likely to be large because the estimated  $\hat{\theta}_i$  is close to  $\theta_i$ . The size of  $|z_{i,0}|$  is thus an indicator for the plausibility of  $H_0$ . To test whether  $H_0$  is true, one computes the probability of having a  $z_{i,0}$  that is more exceptional under  $H_0$  than the one observed, which is the probability that a standard Gaussian random variable  $Z$  takes a value above  $|z_{i,0}|$  or below  $-|z_{i,0}|$ . If this probability, known as  $p$ -value of the test, is small (typically below  $\alpha = 0.05$  or  $\alpha = 0.01$ ), the observed  $z_{i,0}$ , and hence also  $\hat{\theta}_i$ , are rather unlikely to occur under  $H_0$ , so the latter hypothesis is probably false and therefore should be rejected. If the  $p$ -value is large, however, the observations do not contradict  $H_0$ , but it is not possible to say more (in particular, this does *not* mean that  $H_0$  is true!). The rejection limit  $\alpha$  is called the *level* of the hypothesis test.

In our work, the  $z$ -test is carried out for the maximum likelihood estimates of regression coefficients with  $\theta_i^0 = 0$  for all  $i$  to assess whether the corresponding covariates have an effect on total ozone.

## A3 Likelihood ratio test

Another test related to maximum likelihood theory is the *likelihood ratio test*, which allows a comparison of the fit of two nested parametric models  $A$  and  $B$ . Assume that model  $A$  has  $p$  parameters, that model  $B$  is a restriction of  $A$  with  $q < p$  parameters, and that  $\hat{\ell}_A$  and  $\hat{\ell}_B$  are the maximized log-likelihoods of these models. Provided that model  $B$  is true, which is the null hypothesis in this test, the statistic  $2(\hat{\ell}_A - \hat{\ell}_B)$  has an asymptotic  $\chi^2_{p-q}$ -distribution. Using this result as an approximation for finite samples, one can test whether the additional parameters in model  $A$  bring an improvement on model  $B$ . If not, the simpler model  $B$  would be preferred.

We use the likelihood ratio test to compare the models with annual and seasonal covariates.

## A4 Akaike information criterion

A common criterion to compare non-nested parametric models is the Akaike information criterion (AIC), which involves the maximized log-likelihood  $\hat{\ell}$  and a penalization term for

the number of parameters in the model. More precisely,  $AIC = -2\hat{\ell} + 2p$ , where  $p$  is the number of parameters in the model. The best model is the one for which the AIC is minimal, because a good model should have a large likelihood (i.e., plausible that it generated the observed data) and a small number of parameters (i.e., as simple as possible).

In this work we select the order of the ARMA model for mean values based on the AIC.

## A5 Bonferroni correction

The level  $\alpha$  of a hypothesis test is the probability of rejecting the null hypothesis when it is actually true, or, equivalently, the probability of a false rejection. Often one takes  $\alpha = 0.05$ . When performing  $m$  hypothesis tests each with level  $\alpha$ , the probability  $\delta$  of at least one false rejection among the  $m$  tests is bounded above by  $\delta \leq m\alpha$ , which can be much larger than  $\alpha$ . The *Bonferroni correction* consists in adjusting the level of the individual tests to ensure that the global error probability  $\delta$  is bounded by a given probability. If this probability is chosen to be 0.05, the level of the individual tests must be  $\alpha = 0.05m^{-1}$ . The upper bound  $m\alpha$  is fairly general, so there may be better adjustments for particular cases. Moreover the Bonferroni correction is also very conservative when  $m$  is large.

We use the Bonferroni correction as one method in the assessment of a possible multiple testing effect on the significance tests for the regression coefficients.

**Supplementary material related to this article is available online at:** <http://www.atmos-chem-phys.net/13/147/2013/acp-13-147-2013-supplement.pdf>.

*Acknowledgements.* H. E. R., L. F., M. R., J. S., and A. C. D. acknowledge funding by the Competence Centre for the Environment and Sustainability (CCES) within the ETH-domain in Switzerland within the project EXTREMES: “Spatial extremes and environmental sustainability: statistical methods and applications in geophysics and the environment”. H. E. R. acknowledges also funding of the Swiss National Science Foundation through the Fellowship Grant PBEZP2-134426.

The authors are grateful to the New Zealand National Institute of Water and Atmospheric Research (NIWA) for providing the data of the NIWA assimilated total ozone data set.

The authors are thankful to two anonymous referees for their helpful comments during the discussion phase in ACPD, which have led to an improved version of this paper.

The Software R (R Development Core Team, <http://www.R-project.org>) was used for the statistical analysis.

Edited by: N. Harris



## References

- Appenzeller, C., Weiss, A. K., and Staehelin, J.: North Atlantic oscillation modulates total ozone winter trends, *Geophys. Res. Lett.*, 27, 1131–1134, 2000.
- Austin, J. and Wilson, R. J.: Ensemble simulations of the decline and recovery of stratospheric ozone, *J. Geophys. Res.-Atmos.*, 111, D16314, doi:10.1029/2005jd006907, 2006.
- Baldwin, M. P., 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., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, *Rev. Geophys.*, 39, 179–229, 2001.
- Bodeker, G. E., Shiona, H., and Eskes, H.: Indicators of Antarctic ozone depletion, *Atmos. Chem. Phys.*, 5, 2603–2615, doi:10.5194/acp-5-2603-2005, 2005.
- Brockwell, P. J. and Davis, R. A.: *Introduction to Time Series and Forecasting*, Springer, New York, 2nd Edn., 2002.
- Buishand, T. A., de Haan, L., and Zhou, C.: On spatial extremes: With application to a rainfall problem, *Ann. Appl. Stat.*, 2, 624–642, 2008.
- Calbó, J., Pagès, D., and González, J. A.: Empirical studies of cloud effects on UV radiation: A review, *Rev. Geophys.*, 43, RG2002, doi:10.1029/2004rg000155, 2005.
- Chandra, S. and McPeters, R. D.: The solar-cycle variation of ozone in the stratosphere inferred from NIMBUS-7 and NOAA-11 satellites, *J. Geophys. Res.-Atmos.*, 99, 20665–20671, 1994.
- Chilès, J. P. and Delfiner, P.: *Geostatistics: Modelling Spatial Uncertainty*, Wiley, New York, 1999.
- Coles, S. G.: *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, 2001.
- Cooley, D., Naveau, P., and Poncet, P.: Variogram for spatial max-stable random fields, in: *Dependence in Probability and Statistics*, edited by: Bertail, P., Doukhan, P., and Soulier, P., *Lecture Notes in Statistics*, 373–390, Springer, New York, 2006.
- Cooley, D., Nychka, D., and Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Stat. Assoc.*, 102, 824–840, doi:10.1198/016214506000000780, 2007.
- Davison, A. C.: *Statistical Models*, Cambridge University Press, New York, 2003.
- Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds (with discussion), *J. Roy. Stat. Soc. B*, 52, 393–442, 1990.
- Davison, A. C., Padoan, S. A., and Ribatet, M.: Statistical modeling of spatial extremes (with discussion), *Stat. Sci.*, 27, 161–186, doi:10.1214/11-STS376, 2012.
- de Haan, L. and Ferreira, A.: *Extreme Value Theory: An Introduction*, Springer, New York, 2006.
- Diggle, P. J. and Ribeiro, P. J.: *Model-Based Geostatistics*, Springer, New York, 2007.
- Eastoe, E. F. and Tawn, J. A.: Modelling non-stationary extremes with application to surface level ozone, *Appl. Stat.*, 58, 25–45, 2009.
- Eyring, V., Waugh, D. W., Bodeker, G. E., Cordero, E., Akiyoshi, H., Austin, J., Beagley, S. R., Boville, B. A., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Dameris, M., Deckert, R., Deushi, M., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Mancini, E., Manzini, E., Marsh, D. R., Marches, S., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Scinocca, J. F., Semeniuk, K., Shepherd, T. G., Shibata, K., Steil, B., Stolarski, R. S., Tian, W., and Yoshiki, M.: Multimodel projections of stratospheric ozone in the 21st century, *J. Geophys. Res.*, 112, D16303, doi:10.1029/2006JD008332, 2007.
- Eyring, V., Cionni, I., Bodeker, G. E., Charlton-Perez, A. J., Kinnison, D. E., Scinocca, J. F., Waugh, D. W., Akiyoshi, H., Bekki, S., Chipperfield, M. P., Dameris, M., Dhomse, S., Frith, S. M., Garny, H., Gettelman, A., Kubin, A., Langematz, U., Mancini, E., Marchand, M., Nakamura, T., Oman, L. D., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Shepherd, T. G., Shibata, K., Tian, W., Braesicke, P., Hardiman, S. C., Lamarque, J. F., Morgenstern, O., Pyle, J. A., Smale, D., and Yamashita, Y.: Multimodel assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models, *Atmos. Chem. Phys.*, 10, 9451–9472, doi:10.5194/acp-10-9451-2010, 2010.
- Farman, J. C., Gardiner, B. G., and Shanklin, J. D.: Large losses of total ozone in Antarctica reveal seasonal ClO<sub>x</sub>/NO<sub>x</sub> interaction, *Nature*, 315, 207–210, 1985.
- Fawcett, L. and Walshaw, D.: Estimating return levels from serially dependent extremes, *Environmetrics*, 23, 272–283, 2012.
- Fioletov, V. E. and Shepherd, T. G.: Summertime total ozone variations over middle and polar latitudes, *Geophys. Res. Lett.*, 32, L04807, doi:10.1029/2004gl022080, 2005.
- Fioletov, V. E., Bodeker, G. E., Miller, A. J., McPeters, R. D., and Stolarski, R.: Global and zonal total ozone variations estimated from ground-based and satellite measurements: 1964–2000, *J. Geophys. Res.-Atmos.*, 107, 4647, doi:10.1029/2001jd001350, 2002.
- Garcia, R. R. and Solomon, S.: A possible relationship between interannual variability in Antarctic ozone and the Quasi-Biennial Oscillation, *Geophys. Res. Lett.*, 14, 848–851, 1987.
- Hadjinicolaou, P., Jrrar, A., Pyle, J. A., and Bishop, L.: The dynamically driven long-term trend in stratospheric ozone over northern middle latitudes, *Q. J. Roy. Meteorol. Soc.*, 128, 1393–1412, 2002.
- Haklander, A. J., Siegmund, P. C., and Kelder, H. M.: Analysis of the frequency-dependent response to wave forcing in the extratropics, *Atmos. Chem. Phys.*, 6, 4477–4481, doi:10.5194/acp-6-4477-2006, 2006.
- Harris, N. R. P., Kyrö, E., Staehelin, J., Brunner, D., Andersen, S.-B., Godin-Beekmann, S., Dhomse, S., Hadjinicolaou, P., Hansen, G., Isaksen, I., Jrrar, A., Karpetchko, A., Kivi, R., Knudsen, B., Krizan, P., Lastovicka, J., Maeder, J., Orsolini, Y., Pyle, J. A., Rex, M., Vanicek, K., Weber, M., Wohltmann, I., Zanis, P., and Zerefos, C.: Ozone trends at northern mid- and high latitudes – a European perspective, *Ann. Geophys.*, 26, 1207–1220, doi:10.5194/angeo-26-1207-2008, 2008.
- Heggin, M. I. and Shepherd, T. G.: Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux, *Nat. Geosci.*, 2, 687–691, doi:10.1038/ngeo604, 2009.
- Hurrell, J.: Hurrell PC-Based North Atlantic Oscillation Index (Monthly), available online: <http://www.cgd.ucar.edu/cas/jhurrell/indices.html>, 2009.
- Lait, L. R., Schoeberl, M. R., and Newman, P. A.: Quasi-Biennial modulation of the Antarctic ozone depletion, *J. Geophys. Res.-Atmos.*, 94, 11559–11571, 1989.
- Livezey, R. E. and Chen, W. Y.: Statistical field significance and its determination by Monte-Carlo techniques, *Mon. Weather Rev.*, 111, 46–59, doi:10.1175/1520-

- 0493(1983)111<0046:sfsaid>2.0.CO;2, 1983.
- Mäder, J. A., Staehelin, J., Peter, T., Brunner, D., Rieder, H. E., and Stahel, W. A.: Evidence for the effectiveness of the Montreal Protocol to protect the ozone layer, *Atmos. Chem. Phys.*, 10, 12161–12171, doi:10.5194/acp-10-12161-2010, 2010.
- Müller, R., Groß, J.-U., Lemmen, C., Heinze, D., Dameris, M., and Bodeker, G. E.: Simple measures of ozone depletion in the polar stratosphere, *Atmos. Chem. Phys.*, 8, 251–264, doi:10.5194/acp-8-251-2008, 2008.
- Naveau, P., Guillou, A., and Cooley, D.: Modelling pairwise dependence of maxima in space, *Biometrika*, 96, 1–17, 2009.
- Newman, P. A., Daniel, J. S., Waugh, D. W., and Nash, E. R.: A new formulation of equivalent effective stratospheric chlorine (EESC), *Atmos. Chem. Phys.*, 7, 4537–4552, doi:10.5194/acp-7-4537-2007, 2007.
- Northrop, P. J. and Jonathan, P.: Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights (with discussion), *Environmetrics*, 22, 799–816, 2011.
- Oltmans, S. J. and London, J.: The Quasi-Biennial Oscillation in atmospheric ozone, *J. Geophys. Res.-Oc. Atm.*, 87, 8981–8989, 1982.
- Orsolini, Y. J. and Doblas-Reyes, F. J.: Ozone signatures of climate patterns over the Euro-Atlantic sector in the spring, *Q. J. Roy. Meteorol. Soc.*, 129, 3251–3263, doi:10.1256/qj.02.165, 2003.
- Orsolini, Y. J. and Limpasuvan, V.: The North Atlantic Oscillation and the occurrences of ozone miniholes, *Geophys. Res. Lett.*, 28, 4099–4102, 2001.
- Padoan, S., Ribatet, M., and Sisson, S.: Likelihood-based inference for max-stable processes, *J. Am. Stat. Assoc.*, 105, 263–277, 2010.
- Peter, T.: Microphysics and heterogeneous chemistry of polar stratospheric clouds, *Annu. Rev. Phys. Chem.*, 48, 785–822, 1997.
- Randel, W. J. and Cobb, J. B.: Coherent variations of monthly mean total ozone and lower stratospheric temperature, *J. Geophys. Res.-Atmos.*, 99, 5433–5447, 1994.
- Rex, M., Salawitch, R. J., von der Gathen, P., Harris, N. R. P., Chipperfield, M. P., and Naujokat, B.: Arctic ozone loss and climate change, *Geophys. Res. Lett.*, 31, L04116, doi:10.1029/2003gl018844, 2004.
- Rieder, H. E., Staehelin, J., Maeder, J. A., Peter, T., Ribatet, M., Davison, A. C., Stübi, R., Weihs, P., and Holawe, F.: Extreme events in total ozone over Arosa – Part 1: Application of extreme value theory, *Atmos. Chem. Phys.*, 10, 10021–10031, doi:10.5194/acp-10-10021-2010, 2010a.
- Rieder, H. E., Staehelin, J., Maeder, J. A., Peter, T., Ribatet, M., Davison, A. C., Stübi, R., Weihs, P., and Holawe, F.: Extreme events in total ozone over Arosa – Part 2: Fingerprints of atmospheric dynamics and chemistry and effects on mean values and long-term changes, *Atmos. Chem. Phys.*, 10, 10033–10045, doi:10.5194/acp-10-10033-2010, 2010b.
- Rieder, H. E., Jancso, L. M., Di Rocco, S., Staehelin, J., Maeder, J. A., Peter, T., Ribatet, M., Davison, A. C., De Backer, H., Koehler, U., Krzyscin, J., and Vaníček, K.: Extreme events in total ozone over the northern mid-latitudes: an analysis based on long-term data sets from five European ground-based stations, *Tellus B*, 63, 860–874, doi:10.1111/j.1600-0889.2011.00575.x, 2011.
- Rieder, H. E., Frossard, L., Ribatet, M., Staehelin, J., Maeder, J. A., Di Rocco, S., Davison, A. C., Peter, T., Weihs, P., and Holawe, F.: On the relationship between total ozone and atmospheric dynamics and chemistry at mid-latitudes – Part 2: The effects of the El Niño/Southern Oscillation, volcanic eruptions and contributions of atmospheric dynamics and chemistry to long-term total ozone changes, *Atmos. Chem. Phys.*, 13, 165–179, doi:10.5194/acp-13-165-2013, 2013.
- Rue, H. and Held, L.: *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC, London, 2005.
- Sato, M., Hansen, J. E., McCormick, M. P., and Pollack, J. B.: Stratospheric aerosol optical depths, 1850–1990, *J. Geophys. Res.-Atmos.*, 98, 22987–22994, 1993.
- Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A.: A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling, *Extremes*, 13, 219–239, doi:10.1007/s10687-009-0098-2, 2010.
- Schnadt Poberaj, C., Staehelin, J., and Brunner, D.: Missing stratospheric ozone decrease at Southern Hemisphere middle latitudes after Mt. Pinatubo: a dynamical perspective, *J. Atmos. Sci.*, 68, 1922–1945, 2011.
- Shepherd, T. G.: Dynamics, stratospheric ozone, and climate change, *Atmos.-Ocean*, 46, 117–138, doi:10.3137/ao.460106, 2008.
- Solomon, S.: Stratospheric ozone depletion: a review of concepts and history, *Rev. Geophys.*, 37, 275–316, 1999.
- SPARC CCMVal: SPARC Report on the Evaluation of Chemistry-Climate Models, edited by: Eyring, V., Shepherd, T. G., and Waugh, D. W., SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, available at: <http://www.sparc-climate.org/publications/sparc-reports/>, 2010.
- Steinbrecht, W., Haßler, B., Brühl, C., Dameris, M., Giorgetta, M. A., Grewe, V., Manzini, E., Matthes, S., Schnadt, C., Steil, B., and Winkler, P.: Interannual variation patterns of total ozone and lower stratospheric temperature in observations and model simulations, *Atmos. Chem. Phys.*, 6, 349–374, doi:10.5194/acp-6-349-2006, 2006.
- Struthers, H., Bodeker, G. E., Austin, J., Bekki, S., Cionni, I., Dameris, M., Giorgetta, M. A., Grewe, V., Lefèvre, F., Lott, F., Manzini, E., Peter, T., Rozanov, E., and Schraner, M.: The simulation of the Antarctic ozone hole by chemistry-climate models, *Atmos. Chem. Phys.*, 9, 6363–6376, doi:10.5194/acp-9-6363-2009, 2009.
- Thompson, D. W. J. and Wallace, J. M.: Annular modes in the extratropical circulation. Part I: month-to-month variability, *J. Climate*, 13, 1000–1016, 2000.
- van Loon, H.: Half-yearly oscillations in middle and high southern latitudes and coreless winter, *J. Atmos. Sci.*, 24, 472–486, 1967.
- van Loon, H.: Half-yearly oscillations in Drake Passage, *Deep-Sea Res.*, 19, 525–527, 1972.
- Varin, C.: On composite marginal likelihoods, *Adv. Stat. Anal.*, 92, 1–28, 2008.
- Ventura, V., Paciorek, C. J., and Risbey, J. S.: Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data, *J. Climate*, 17, 4343–4356, doi:10.1175/3199.1, 2004.
- von Clarmann, T., Stiller, G., Grabowski, U., Eckert, E., and Orphal, J.: Technical Note: Trend estimation from irregularly sampled, correlated data, *Atmos. Chem. Phys.*, 10, 6737–6747,

- doi:10.5194/acp-10-6737-2010, 2010.
- Vyushin, D., Fioletov, V. E., and Shepherd, T. G.: Impact of long-range correlations on trend detection in total ozone, *J. Geophys. Res.*, 112, D14307, doi:10.1029/2006JD008168, 2007.
- Waugh, D. W. and Randel, W. J.: Climatology of arctic and antarctic polar vortices using elliptical diagnostics, *J. Atmos. Sci.*, 56, 1594–1613, 1999.
- Waugh, D. W., Oman, L., Kawa, S. R., Stolarski, R. S., Pawson, S., Douglass, A. R., Newman, P. A., and Nielsen, J. E.: Impacts of climate change on stratospheric ozone recovery, *Geophys. Res. Lett.*, 36, L03805, doi:10.1029/2008gl036223, 2009.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, Academic Press, Amsterdam, 3rd Edn., 2011.
- WMO: Scientific Assessment of Ozone Depletion: 1994, Global Ozone Research and Monitoring Project – Report No. 37, World Meteorological Organization, Geneva, 1995.
- WMO: Scientific Assessment of Ozone Depletion: 2002, Global Ozone Research and Monitoring Project – Report No. 47, World Meteorological Organization, Geneva, 2003.
- WMO: Scientific Assessment of Ozone Depletion: 2006, Global Ozone Research and Monitoring Project – Report No. 50, World Meteorological Organization, Geneva, 2007.
- WMO: Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project – Report No. 52, World Meteorological Organization, Geneva, 2011.