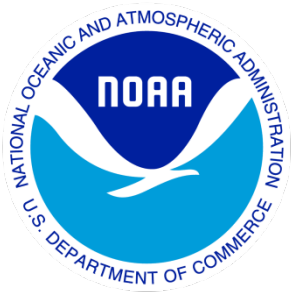
Climate Data Record (CDR) Program

Climate Algorithm Theoretical Basis Document (C-ATBD)

Ozone Binary Database of Profiles (BDBP)



CDR Program Document Number: CDRP-ATBD-0555
Configuration Item Number: 01B-28 (Ozone - ESRL)
Revision 1 / September 17, 2014

REVISION HISTORY

Rev.	Author	DSR No.	Description	Date
1	Birgit Hassler	DSR-705	Initial Submission to CDR Program	09/15/2014

TABLE of CONTENTS

LIST of FIGURES

LIST of TABLES

1. Introduction

1.1 Purpose

This document describes the algorithm used to create the Climate Data Record (CDR) for the Zonal Mean Ozone Binary Database of Profiles (BDBP) (Hassler et al., 2008). The algorithm and CDR product has been submitted to the National Climatic Data Center (NCDC) by Dr. K. Rosenlof of NOAA Earth System Research Laboratory (ESRL) and Dr. Birgit Hassler/NOAA ESRL. The actual algorithm is provided in the computer programs (code) that accompany this document. Thus the intent here is to provide a guide to understanding that algorithm from a scientific perspective and to assist a software engineers or end-users performing an evaluation of the code.

1.2 Definitions

Following is a summary of the symbols used to define the algorithm.

Pinatubo basis function parameters:

y = description of the stratospheric aerosol loading (1)

Onset = temporal expression of the onset of aerosol in the stratosphere (July, 1991) (2)

DecimalYear = time index (3)

Regresssion parameters (Section 3.4.1):

Ozone = modeled ozone values (4)

t = time index in regression model (5)

φ = latitude index in regression model (6)

Γ = age of air index in regression model (7)

EESC = EESC basis function (8)

QBO = QBO basis function (9)

QBOorthog = Orthogonalized QBO basis function (10)

ENSO = ENSO basis function (11)

Solar = 11-year solar cycle basis function (12)

Pinatubo = Strat. aerosol loading from Pinatubo basis function (13)

FG = first guess of the regression model fit coefficient (14)

PP = perturbation pattern *(15)*

1.3 Referencing this Document

This document should be referenced as follows:

Ozone Binary Database of Profiles (BDBP) - Climate Algorithm Theoretical Basis Document, NOAA Climate Data Record Program CDRP-ATBD-0555 Rev. 1 (2014). Available at http://www.ncdc.noaa.gov/cdr/operationalcdrs.html

1.4 Document Maintenance

This document describes the initial submission, version 1.0 revision 0 (v01r00) of the processing algorithm and resulting data. This document is under configuration management. The version and revision numbers will be incremented to include subsequent enhancements to the processing or algorithm when proposed changes have been coordinated with and submitted to the CDR Program Office.

2. Observing Systems Overview

2.1 Products Generated

The objective of this algorithm is to create a global, gap-free, vertically resolved, zonal mean ozone data set that spans the period 1979-2007. Since this data set was created using multiple linear regressions, it is possible to provide several versions of this product according to various basis function contributions. To understand the different contributions of basis functions, the data product is provided in five different "Tiers" (Bodeker et al., 2013).

- Tier 0: raw monthly mean data that was used in the regression model
- Tier 1.1: Anthropogenic influences (as determined by the regression model)
- Tier 1.2: Natural influences (as determined by the regression model)
- Tier 1.3: Natural and volcanic influences (as determined by the regression model)
- Tier 1.4: All influences (as determined by the regression model)

2.2 Instrument Characteristics

Ozone profile measurements from several different satellite systems and ozonesondes are combined to create the monthly means that are provided for the regression model. The main criteria for the selection of the satellite systems that have been used to generate the Ozone BDBP product were to have: (1) high vertical resolution of the profiles in their native grids, (2) measurements that include an uncertainty value, (3) preferably span a long time period, (4) cover a data sparse region of the globe (Hassler et al., 2008). Using these criteria, data from the following measurement systems have been used to calculate the monthly mean, zonal mean ozone values (see also Table 1):

- Stratospheric Aerosol and Gas Experiment I: SAGE I (McCormick et al., 1989)
- Stratospheric Aerosol and Gas Experiment II: SAGE II (McCormick et al., 1989; Wang et al., 2006)
- Halogen Occultation Experiment: HALOE (Russell et al., 1993)
- Polar Ozone and Aerosol Measurement II: POAM II (Glaccum et al., 1996; Lumpe et al., 1997)
- Polar Ozone and Aerosol Measurement III: POAM III (Lucke et al., 1999; Lumpe et al., 2002)
- Improved Limb Array Spectrometer: ILAS (Sasano et al., 1999)
- Improved Limb Array Spectrometer II: ILAS II (Nakajima, 2006)

- Limb Infrared Monitor of the Stratosphere: LIMS (Remsberg et al., 1984)
- Ozonesondes (Smit, 2002)

Both SAGE instruments, HALOE, both POAM instruments, and both ILAS instruments use the solar occultation technique for their measurements. For the solar occultation technique transmission of sunlight through the Earth's atmosphere is measured and then put into a ratio to exoatmospheric measurements (solar measurements recorded with no atmospheric attenuation). This is carried out at a series of tangent heights, increasing in altitude during a sunrise or decreasing in altitude during a sunset, therefore obtaining information about different atmospheric layers. The solar occultation technique measurements have a very high signal-to-noise ratio allowing the detection of species with low concentrations without the need for averaging measurements. All mentioned measurement systems and their characteristics concerning ozone profile measurements are described in more detail in Hassler et al. (2014).

Table 1a: Measurement systems that are used in the creation of the BDBP.

Instrument name	Measurement technique	Time period	Vertical resolution	Latitudinal coverage
SAGE I	Solar occultation	02/1979 – 11/1981	~ 1 km	79°S – 79°N
SAGE II	Solar occultation	10/1984 – 08/2005	~ 1 km	80°S – 80°N
HALOE	Solar occultation	10/1991 – 11/2005	~ 1.6 km	80°S – 80°N
POAM II	Solar occultation	10/1993 – 11/1996	~ 1 km	88°S – 63°S, 55°N – 71°N
POAM III	Solar occultation	04/1998 – 11/2005	~ 1 km	88°S – 62°S, 54°N – 71°N
ILAS	Solar occultation	09/1996 – 06/1997	~ 1.9 km	88°S – 64°S, 57°N – 71°N
ILAS II	Solar occultation	03/2003 – 10/2003	~ 1.6 km	88°S – 64°S, 57°N – 71°N
LIMS	Limb emission	10/1978 – 05/1979	~ 3.7 km	64°S – 84°N
Ozonesondes	Chemical reaction of ozone with potassium iodine in an aqueous solution	01/1979 – 12/2007	100 – 150 m	Global, but point measurements (see Figure 1)

Table 2b: Measurement systems that are used in the creation of the BDBP (continued).

Instrument name	Number of measurements	Covered vertical region	Covered wavelength region
SAGE I	Up to 15 sunrise and sunset events per day	Tropopause/10km – 55km	0.385 - 1.00 µm
SAGE II	Up to 15 sunrise and sunset events per day	Tropopause/10km – 55km	0.385 - 1.02 µm
HALOE	Up to 15 sunrise and sunset events per day	Cloud top/tropopause – 0.005hPa	9.2 to 10.4 µm
POAM II	14 sunrise and sunset events per day	Cloud top/5km – 60km	0.35 to 1.06 µm
POAM III	14 sunrise and sunset events per day	Cloud top/5km – 60km	0.35 to 1.02 µm
ILAS	14 sunrise and sunset events per day	Cloud top/10km – top of atmosphere	6 to 12 µm
ILAS II	14 sunrise and sunset events per day	Cloud top/10km – top of atmosphere	6 to 12 µm
LIMS	~3000 profiles per day	Cloud top/~300hPa – 0.01hPa	9 to 10 µm
Ozonesondes	1-2 soundings per week	Surface – 30 to 35km	n/a

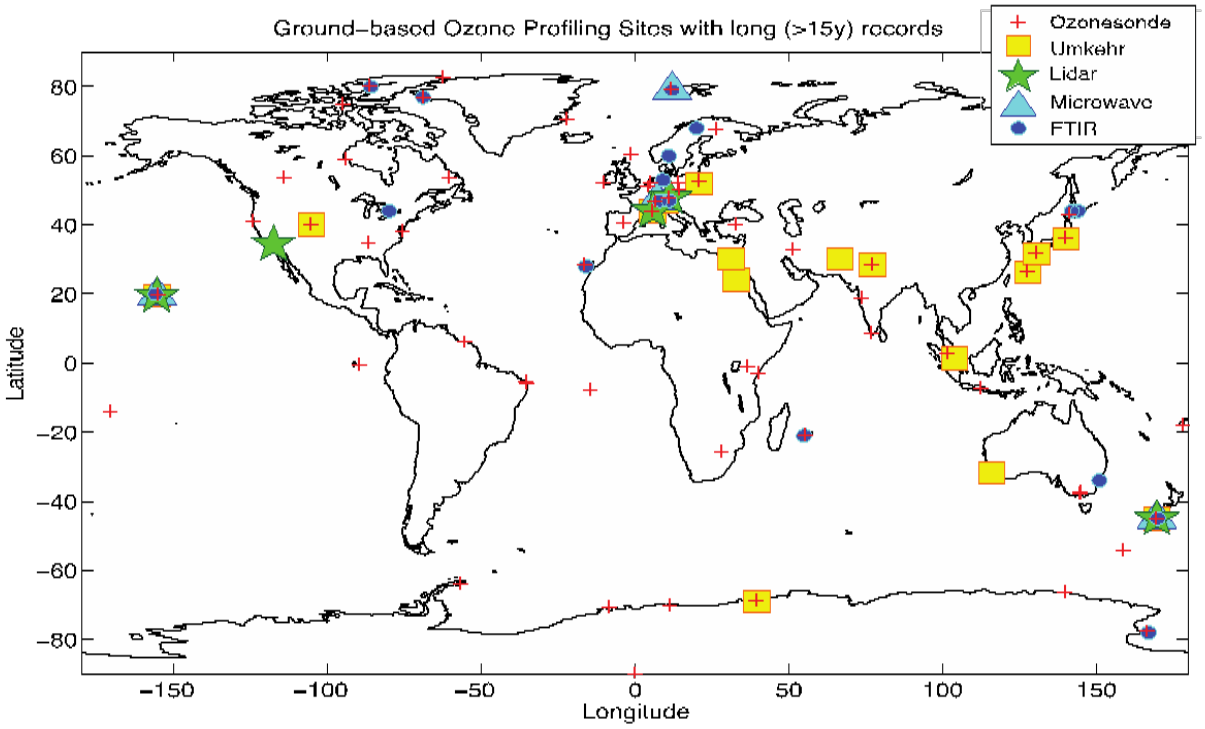


Figure 1: World map with locations of long-term (> 15 years) ground-based measurement sites: ozone sounding sites (red plus signs), lidar sites (green stars), microwave sites (light blue triangles), Dobson/Brewer sites with Umkehr measurements (yellow squares), and FTIR sites (dark blue circles). More stations with shorter records (< 15 years) can be found in the WOUDC and NDACC data bases (Figure taken from Hassler et al., 2014).

3. Algorithm Description

3.1 Algorithm Overview

Creating the global, zonal mean monthly mean data sets (Tier 1.1 – Tier 1.4) requires three steps of processing: (1) combining (pre-processing) the individual data points of the different measurement systems to calculate zonal mean monthly mean ozone values for specific altitude/pressure region, latitudinal region, and months where available, (2) run the multiple linear regression model using basis function time series (time series of natural and anthropogenic factors that can influence ozone concentrations; see Section 3.3.2) to obtain a statistical representation of the input data and fill in data gaps, (3) combine the contributions of the different basis functions in several variations to obtain four Tiers for the zonal mean, monthly mean data set (post-processing).

3.2 Processing Outline

The three processing steps included in the algorithm overview are outlined in Figure 3.1, where input data are given by blue boxes, yellow boxes illustrate the performed processing, and the green boxes highlight the resulting product. A more detailed description of the processing is described below:

Pre-Processing: Individual ozone measurements from the different satellite systems and ozone soundings are extracted from the BDBP database (Hassler et al., 2008) and accumulated into monthly means in 5° latitude bands. The individual data points were quality screened* before they were added to the BDBP database. Weights are then calculated for each ozone value according to its uncertainty. The weights are applied to the different data points and monthly means are calculated as average of the available ozone values for each given latitude band for various altitudes/pressure levels and months to produce the Tier 0 BDBP dataset, which is then used as an input to the statistical model. The program used to calculate the monthly means is called “CalculateMonthlyMeans.dpr”. However, since the Tier 0 dataset is provided, it should not be necessary to recalculate the monthly means to be able to rerun the statistical regression model.

* Quality screening (for adding data to the BDBP database): Measurements from the different measurement systems were quality screened before they were used to calculate zonal mean monthly means. The quality screening is instrument dependent, but can be summarized as follows:

- Following suggestions (quality flags) from the individual profiles in the original measurement files about data exclusion or filtering.
- Application of an altitude correction of the SAGE I measurements according to Wang et al. (1996).
- Check for negative ozone values in ozonesonde profiles.
- Check for unrealistic outliers in ozonesonde profiles.

For more details about the quality screening see Hassler et al. (2008).

Statistical Model: The type of statistical model applied in this processing is a multiple linear least squares regression (see Bodeker et al., 2013). The raw monthly mean ozone values (Tier 0) of each pressure/altitude level provide the input for this model. A linear combination of several basis functions (pre-selected according to their ability to explain ozone variability; see Section 3.3.2) is ozone variability and determine the coefficients that define the contribution of used to describe the observed each basis function. Tiers 1.1-1.4 of the BDBP dataset are produced from the statistical model. The program used to calculate the regression is called "MakeTier1Databases.dpr".

Post-Processing: Different combinations of basis function contributions (basis function time series * basis function coefficient determined by the statistical model) are then used to create the four different Tiers of data (i.e., Tier 1.1 – Tier 1.4: see Section 2.1). The obtained data arrays are then stored in netCDF format.

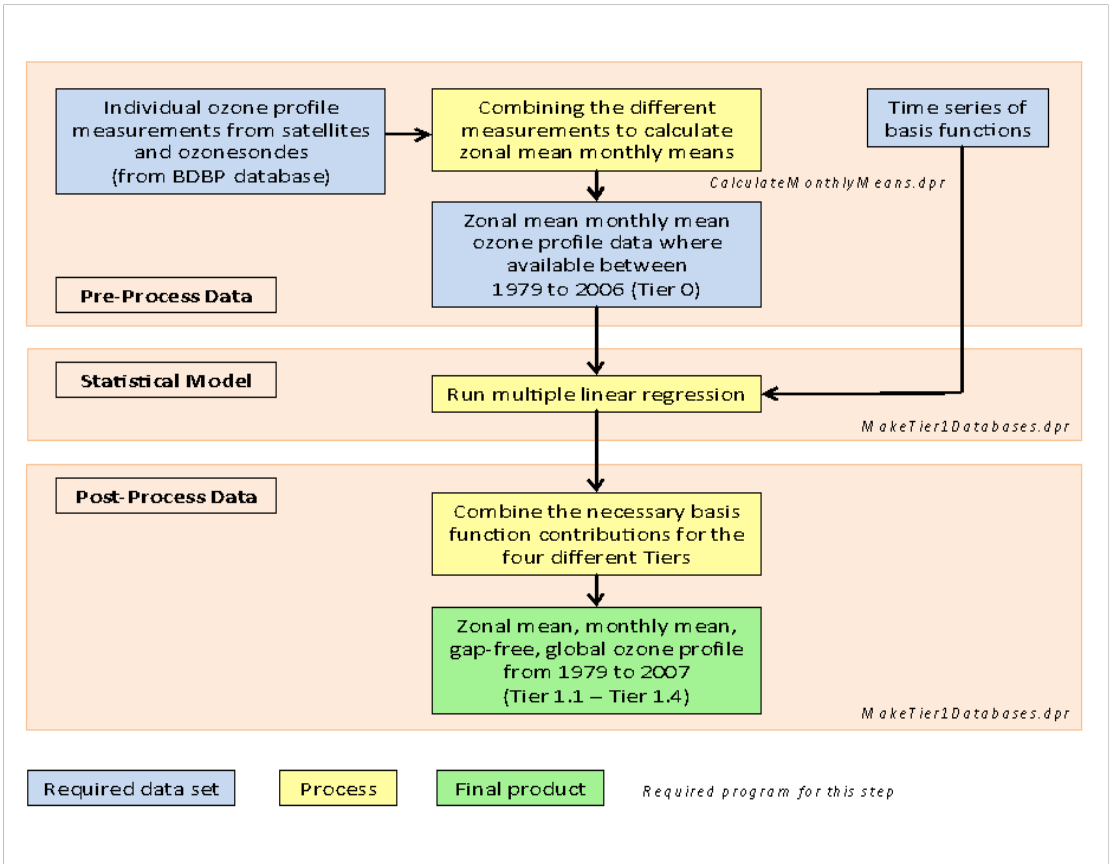


Figure 2: Flow chart of processing steps to create the zonal mean ozone climate data record.

3.3 Algorithm Input

3.3.1 Primary Sensor Data

Ozone profile measurements from several different satellite systems (SAGE I, SAGE II, HALOE, ILAS, ILAS II, POAM II, POAM III and LIMS; see Section 2.2) and ozonesondes are used as input data for the calculation of monthly mean ozone values. These data are stored in the BDBP database (see Hassler et al., 2008 for more detail and all the information of from the different data versions). They have been quality screened (see Section 3.2) and interpolated onto a common altitude and pressure grid. BDBP ozone data are provided in two different units. These include 1) mixing ratio and 2) number density (molecules/m3). Data in the BDBP is stored in a binary format.

3.3.2 Ancillary Data

There are several different ancillary data sets needed as input for the statistical model: basis function time series and global total column ozone fields. The basis function time series include the Quasi-biennial Oscillation (QBO), Solar Flux Data, El Niño Southern Oscillation (ENSO), stratospheric aerosol loading (here: from the Mt. Pinatubo eruption), Equivalent Effective Stratospheric Chlorine (EESC), and the first guess of regression model fit coefficients. The first four basis functions were chosen because they have been shown to explain a high percentage of observed total column ozone and ozone profile variations (Harris et al., 1998; Bodeker et al, 1998; Daniel et al., 1995; Newman et al., 2007), whereas the last basis function was added for the first time to a regression model for the creation of the BDBP. More details to these ancillary data sets are listed below:

- Quasi-biennial Oscillation (QBO): It has been shown that the QBO in zonal wind and temperature in the tropical stratosphere influences the ozone profile in the tropics and the subtropics (e.g. Zawodny and McCormick, 1991). The QBO basis function was specified as the monthly mean 50 hPa Singapore zonal wind. The time series is provided by the Institute of Meteorological from the Free University of Berlin and can be downloaded from an anonymous ftp-server (http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/index.html). The time series is updated on a monthly basis.

- Solar flux data (10.7cm wavelength): It has been shown that UV irradiance changes related to the 11-year solar cycle influence the ozone production and therefore cause ozone variations (e.g. Zerefos et al., 1997). The time series of solar flux data is used as a proxy for the 11-year solar cycle. It is provided by NOAA's National Geophysical Data Center (NGDC) in different formats (daily, monthly, yearly values), and can be downloaded from the NGDC website (http://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-radio/noontime-flux/penticton/penticton_observed/tables/).

● El Niño Southern Oscillation (ENSO): ENSO is large-scale ocean-atmosphere climate phenomenon linked to a periodic warming in sea-surface temperatures across the central and east-central equatorial Pacific. The strength of this phenomenon is often expressed with the Southern Oscillation Index (SOI), that is defined as the difference in sea level pressure between Tahiti and Darwin, Australia. It influences the sea surface temperatures and the state of the troposphere, and therefore results in ozone variability (Kiladis et al., 2001). The time series of monthly mean SOI values, used as proxy for the strength of the ENSO, can be downloaded from NOAA's Climate Prediction Center (CPC): http://www.cpc.ncep.noaa.gov/data/indices/soi

● Stratospheric aerosol loading: The eruption of Mt. Pinatubo (Philippines) in June 1991 put a huge amount of aerosol in the atmosphere that significantly affected global ozone and temperatures (Rosenfield et al., 1997). The aerosol concentration is approximated by the following equation:

$$y = exp(Onset - DecimalYear) \times \{1\text{-}exp[6 \times (Onset - DecimalYear)]\} \qquad (3.1)$$

for all times after the Onset time and 0 for all times before the Onset time (see Bodeker et al., 1998).

● Equivalent effective stratospheric chlorine (EESC): The EESC time series represents the total halogen loading of the stratosphere effective in ozone depletion (Daniel et al., 1995). The EESC differs with age of air (Newman et al., 2007), resulting in a different shape and time of EESC peak depending on which part of the stratosphere and the globe is considered. The EESC basis function is therefore adjusted according to the annually averaged zonal mean of the mean age of air shown in Fig. 7 of Waugh and Hall (2002).

● First guess of regression model fit coefficients: These are obtained by running a regression model on total column ozone data (that has a more complete global coverage) with the same basis functions as planned for the vertically resolved ozone data to enable the regression model to describe finer ozone variability (for a detailed description of this method see Bodeker et al., 2013). The total column ozone and the performed regression on that data are described in Bodeker et al. (2001). The total column ozone data can be obtained from http://www.bodekerscientific.com/data/total-column-ozone

3.3.3 Derived Data

The statistical model outlined in Section 3.2 and Section 3.4 produces not only a zonal mean ozone data set that describes the observed ozone variability, but also several additional data sets that describe only the part of the ozone variability that is caused by specific causes:

● Tier 1.1: All ozone variability that is caused by anthropogenic factors like the effects of climate change and the effects of ozone depleting substances are combined here.

● Tier 1.2: All ozone variability that is caused by natural factors like the month-to-month variability, the effects of the QBO, ENSO and solar cycle are combined here.

● Tier 1.3: All ozone variability that is caused by natural factors (see Tier 1.2) and the effects of major volcanic eruptions on the ozone variability are combined here.

3.3.4 Forward Models

N/A

3.4 Theoretical Description

The objective of the algorithm is to calculate a monthly mean zonal mean, gap-free, vertically resolved ozone data set with the help of a multiple linear regression. To this end, monthly mean ozone values are calculated from several different data sources, for each defined latitude band and vertical level. These monthly means are then used in the regression model to produce a gap-free estimate of the global, vertically resolved ozone distribution. The different produced Tiers of the database can help with comparisons of climate or chemistry-climate model output for an evaluation of model performance.

3.4.1 Physical and Mathematical Description

The description of the regression method used to create the data set is taken from Bodeker et al. (2013). The least squares regression model fit to Tier 0 values at each pressure or altitude level is of the form:

$$\begin{aligned} Ozone(t,\varphi) = {} & A(t,\varphi)+ \\ & B(t,\varphi)\times EESC(t,\Gamma)+ \\ & C(t,\varphi)\times t+ \\ & D(t,\varphi)\times QBO(t)+ \\ & E(t,\varphi)\times QBOorthog(t)+ \\ & F(t,\varphi)\times ENSO(t)+ \\ & G(t,\varphi)\times Solar(t)+ \\ & H(t,\varphi)\times Pinatubo(t)+ \\ & R(t) \end{aligned} \qquad (3.2)$$

where Ozone(t,φ) is the regression modeled ozone number density on some pressure or altitude surface as a function of time (t) and latitude (φ). Note that a single fit is applied to all available Tier 0 data on a given surface; the model is not applied separately in each 5° zone. The advantage of this approach is that the regression model can then be used to interpolate/extrapolate to latitudes where there would be insufficient data to apply a

zonal regression model. Letters, A to I in (3.2) are the regression model coefficients calculated using the standard least squares regression (Press et al., 1989).

The first term in the regression model (A coefficient) represents a constant offset and when expanded in a Fourier expansion (see below), represents the mean annual cycle.

The EESC (equivalent effective stratospheric chlorine; Daniel et al. (1995)) basis function represents the total halogen load within the stratosphere that effectively leads to ozone depletion. The EESC differs with age of air (Newman et al., 2007), as denoted by the Γ parameter in equation (3.1). If this resulted only in a linear scaling of the EESC basis function, this would not be an issue since the B coefficient would adjust accordingly. However, changes in the age of air change the date at which the EESC time series peaks and also the shape of the EESC basis function. To incorporate these effects, the annually averaged zonal mean of the mean air age is shown in Figure 7 of Waugh and Hall (2002). This information was used to appropriately weigh EESC time series generated in 1 year age increments obtained from Paul Newman (personal communication). Since the EESC basis function is a stratospheric quantity, it was excluded from the fitting below 10 km in all regions, below 11 km equatorward of 50°, below 12 km equatorward of 40°, below 13 km equatorward of 35°, below 14 km equatorward of 30°, below 15 km equatorward of 27.5°, and below 16 km equatorward of 25° (approximately following the mean tropopause; see Table 2).

Table 3: Altitudes where the EESC basis function was not included in the regression model (symmetric for both hemispheres).

90° - 50°	50° - 40°	40° - 35°	35° - 30°	30° - 27.5°	27.5° - 25°	25° - 0°
< 10 km	< 11 km	< 12 km	< 13 km	< 14 km	< 15 km	< 16 km

The linear trend term (C coefficient) is used to account for linear changes in ozone that may result from changes in greenhouse gases. The EESC and trend basis functions are far from orthogonal but since this regression model is not used for attribution (quantitative description of which basis function causes how much of the observed ozone variability) in any way, it is inconsequential whether trend-like variance is assigned to the EESC basis function or to the trend basis function. Including both EESC and trend basis functions resulted in a better quality fit to the raw ozone data compared to using each separately.

The QBO basis function was specified as the monthly mean 50 hPa Singapore zonal wind. The phase of the QBO varies with latitude and altitude. In attempt to fit the phase, a second QBO basis function that is orthogonalized to the first, was included in the regression model as was done in Austin et al. (2008).

The El Niño Southern Oscillation (ENSO), solar cycle, and stratospheric aerosol (Mt. Pinatubo) basis functions were the same as those used in Bodeker et al. (2001). Note

that no basis function is included for the El Chichón eruption since there were insufficient data to adequately constrain the fit.

As in Bodeker et al. (2001), a two-term autocorrelation model is used to account for the effects of autocorrelation when calculating the uncertainties on the fit coefficients.

For ozone below 1 ppm or 1×10^18 molec cm−3 the values are transformed using:

$$O3' = \ln(O3) + 1.0 \qquad (3.3)$$

before being passed to the regression model. An inverse transform is applied to the ozone values obtained from the regression model. This logarithmic transformation at low values results in large additions to the regression model cost function (the sum of the squared residuals that is calculated to express the goodness of the fit) when the regression does not track very low ozone measurements. In this way the low ozone values are tracked well by the model. An example of this is found in the Antarctic winter stratosphere.

Because a single instance of the regression model is applied across all latitudes and seasons, the regression model fit coefficients are expanded as follows:

$$X = X0 \times FG(t,\varphi) + X1 + PP(t,\varphi) \qquad (3.4)$$

where X0 and X1 are fit coefficients, FG is a first guess of the regression model fit coefficient and is a function of season and latitude and PP is a perturbation pattern that is also a function of season and latitude. If the first guess describes the complete season and latitude structure of the regression model fit coefficient, then X0 would be 1.0, X1 would zero, and PP would be uniformly zero. If the first guess describes the shape of the season and latitude structure of the regression model fit coefficient but with the wrong amplitude, then X0 ≠ 1.0. If the first guess describes the shape of the season/latitude structure of the regression model fit coefficient but with a systematic bias, then X1 would become non-zero. The shape of the first guess is not correct; this is modified by the perturbation pattern (PP). PP may be visualized as a two-dimensional field where the horizontal structure represents the seasonal dependence and the vertical structure represents the latitudinal dependence. To this end PP is constructed from Fourier series to account for the seasonality as:

$$\sum_{k=1}^{N}\left[X2_{2k-1}\sin\left(2\Pi k\left(M-0.5\right)/12\right)+X2_{2k}\cos\left(2\Pi k\left(M-0.5\right)/12\right)\right] \qquad (3.5)$$

where N is the number of Fourier pairs in which the coefficient is expanded. The X2i coefficients in equation 3.5 are then further expanded in spherical harmonics to account for the latitudinal structure. Since the data being fitted are zonal means, the spherical harmonics reduce to Legendre polynomials of the form Pn(cosθ).

Operationally, the regression is first performed at 20 km/58.19 hPa using the latitude/season structure of the fit coefficients derive by applying a regression model similar to that presented in equation 1 to total column ozone fields as described in Bodeker et al. (2001). For levels above this, the pattern response from the level below is used as the first guess which is then perturbed by the offset and perturbation pattern. Similarly, for

lower levels, the pattern response from the level above is used. In this way the fit 'evolves' upward and downward from 20 km/58.19 hPa in a way that is constrained by a prescribed number of terms in the Fourier and Legendre expansions. The number of terms prescribed for the Fourier and Legendre expansions were selected empirically by slowly increasing the values until the marginal improvement in the quality of the fit was insignificant. The offset coefficient, when expanded in Fourier terms, accounts for the mean annual cycle and by including 4 Fourier pairs in the perturbation pattern very subtle changes in the mean annual cycle from one level to the next can be accurately tracked. The number of Fourier pairs included in the perturbation pattern for EESC is decreased at upper levels since the change in the annual cycle from one level to the next at these upper levels is small. For ENSO, the solar cycle, and the Pinatubo eruption, the effect on ozone is assumed to be independent of season. The number of terms in the Legendre expansions constrains how the latitudinal structure in the perturbation pattern can change from one level to the next. For the offset basis function, which describes the mean annual cycle, 10 Legendre polynomials are used since the annual cycle is a very robust feature in the data and this allows for the amplitude of the annual cycle to adapt to small changes in meridional structure from one level to the next. This is reduced to 5 Legendre terms at the upper levels where changes from one level to the next are smaller. Similarly, for the EESC basis function, where it is necessary to track steep meridional gradients in the response of ozone to EESC, 6 Legendre polynomials are used at altitudes below 50 km which is reduced to 3 at levels above 50 km. For all other basis functions, 3 Legendre terms are used in the perturbation pattern. The structure of the regression model results in up to 188 fit coefficients.

The advantage of including a first guess of the pattern response of ozone to the basis function is the Fourier and Legendre expansions constituting the perturbation pattern can be truncated at a fewer number of terms since steep gradients in season or latitude are likely to be captured by the first guess pattern. This prevents overfitting of the regression model which would likely cause anomalies in regions where there are few data to constrain the fit.

3.4.2 Data Merging Strategy

Monthly mean ozone values are calculated for each latitude band (5°) and each vertical level by combining individual ozone measurements from the latitude/altitude or latitude/pressure grids of the BDBP. Some screening of the source data was performed before the monthly means were calculated, specifically all SAGE data below 18 km, all SAGE II data below 10 km, and all LIMS data below 25 km were excluded since they were found to include occasional anomalous values which biased the monthly means (Bodeker et al., 2013). Data from ozonesonde flights were only added to the monthly mean calculation from flights where normalization factors (normalization factors: integrated ozonesonde profile divided by independent total column ozone measurement) ranged between 0.9 and 1.1. The normalization factors were applied to correct the ozonesonde data before being added to the monthly mean calculation.

Additionally, weights dependent on the measurement uncertainty ($1/\sigma 2$; σ is the uncertainty on the measurement) were calculated for each available ozone value. A

latitudinal weighting was further applied to the ozone values using the cosine of the latitude where the measurement was made. Both weights were applied to each individual ozone value for each latitude band and vertical level in the calculation of a weighted monthly mean.

To avoid spatial biasing of the monthly means (for example, if measurements are only available from a narrow longitudinal region within a latitude band), an additional spatial bias correction was applied: each ozone measurement was scaled by the monthly mean zonal mean total column ozone divided by the daily total column ozone at the latitude and longitude of that measurement (Bodeker et al., 2013). The total column ozone data for this bias correction was obtained from the same source as for the first guess of regression model fit coefficients Section 3.3.2).

To actually calculate a monthly mean then, at least 6 measurements for each latitude band and each vertical level was required. If only ozonesonde data were available, this requirement was omitted. In a two-step process the actual monthly mean was determined: the standard deviation (σ) of each monthly mean was calculated (step 1), and then only data within 3σ of the calculated mean of step 1 were used to calculate a revised monthly mean and standard deviation (step 2). These monthly means constitute Tier 0 data and were used as input to a least squares regression model to generate the Tier 1.1 to Tier 1.4 data sets (Bodeker et al., 2013).

3.4.3 Numerical Strategy

N/A

3.4.4 Calculations

See section 3.4.1

3.4.5 Look-Up Table Description

N/A

3.4.6 Parameterization

N/A

3.4.7 Algorithm Output

Four different tiers of database were constructed as follows:

● Tier 1.1 (Anthropogenic): summing the contributions from the offset (A), EESC (B) and linear trend (C) basis functions.

● Tier 1.2 (Natural): summing the contributions from the offset (A), QBO (D and E), ENSO (F), and solar cycle (G) basis functions.

● Tier 1.3 (Natural and Volcanoes): summing the contri- butions from the offset (A), QBO (D and E), ENSO (F), solar cycle (G), and Mt. Pinatubo volcanic eruption (H) basis functions.

● Tier 1.4 (Anthropogenic and natural): summing the contributions from all basis functions.

A netCDF file with each of these Tiers was created, where Tier 0 was also added. Each file contains ozone monthly means, zonal means, in the ozone units 'number density' and 'mixing ratio', on the vertical grids 'altitude' and 'pressure'. This results in two files altogether, each containing all five Tiers for both given ozone units (10 Tiers altogether) for each vertical grid, with the size for each file of about 35. 1MB.

4. Test Datasets and Outputs

4.1 Test Input Datasets

A further means of validating the database developed here is to compare vertically integrated ozone profiles (ozone columns measured in Dobson Units (DU); 1DU=2.69×1016 molec cm−2) from the Tier 1.4 database with independent monthly mean total column ozone times series (see Fig. 3).

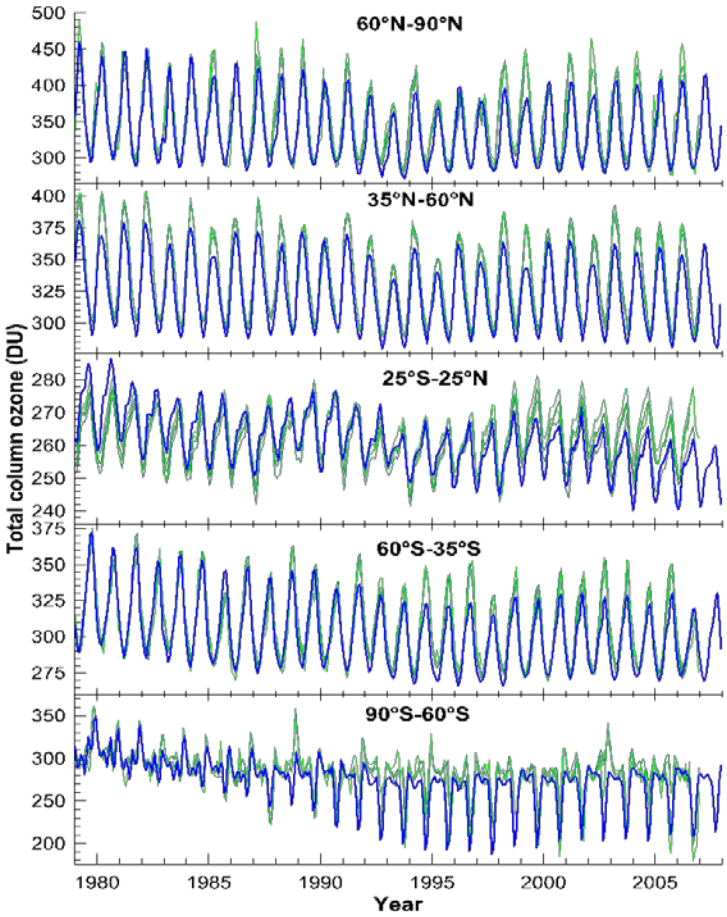


Figure 3: A comparison of zonal mean total column ozone from the four monthly mean observational time series updated from Fioletov et al. (2002) shown in green traces, and zonal mean total column ozone calculated by vertically integrating the Tier 1.4 database (blue trace). It is not necessary to distinguish between the four independent total column ozone time series and so these are not labeled individually (Bodeker et al., 2013).

The four monthly mean total column ozone time series shown in Fig. 3 are updates of those from Fioletov et al. (2002). Note that at no stage in the generation of the

Tier 1.4 ozone database are the values normalized to ensure agreement with an independent total column ozone value.

Additionally, individual zonal mean ozone values of specific months and altitudes from the Tier 0 and Tier 1.4 databases were compared with values extracted from a database compiled by Randel and Wu (2007). This database was most commonly used to constrain Atmosphere-Ocean General Circulation Model (AOGCM) simulations, to calculate ozone radiative forcing, and for analysis of long-term changes in the vertical distribution of ozone. Figure 4 shows an example of such comparisons for selected March mean ozone time series.

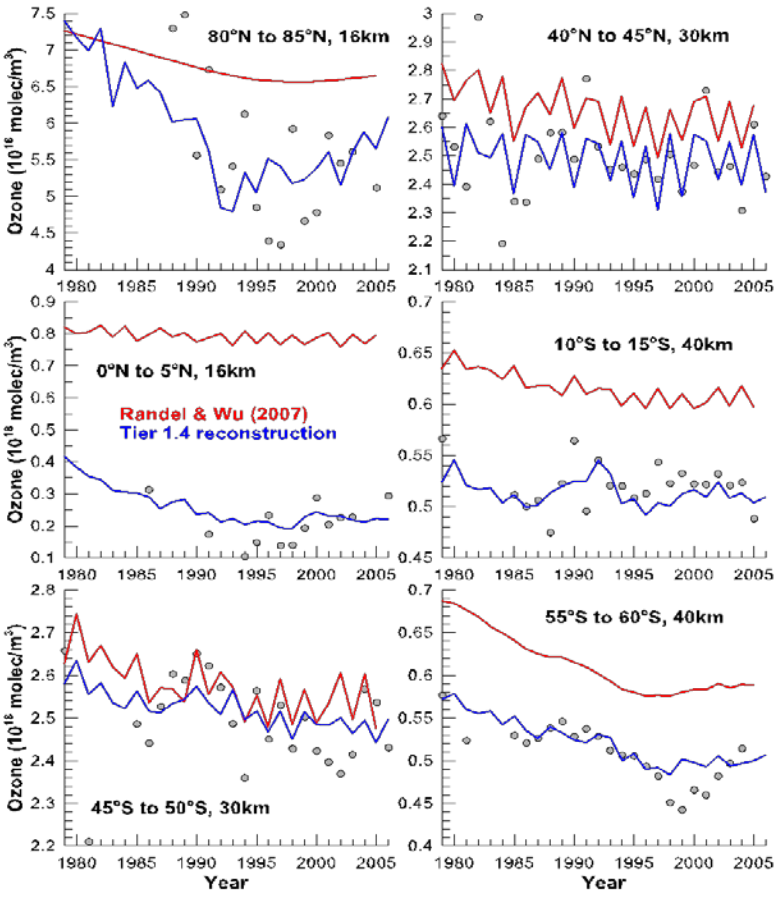


Figure 4: Selected March mean ozone time series from the Tier 0 database (grey dots), from Randel and Wu (2007) (red trace), and from the Tier 1.4 database (blue trace) (Bodeker et al., 2013).

4.2 Test Output Analysis

4.2.1 Reproducibility

The raw ozone data (Tier 0), and complete auxiliary data sets or links to their web destination are provided. Applying the processing algorithm to the input dataset (Tier

0), the user should recover exact results for each of final results. Differences in any of the final results are indicative of an error.

4.2.2 Precision and Accuracy

Each of the individual data sources (ozone profiles from the different satellites) is provided with an uncertainty estimate in the BDBP database (Hassler et al., 2008). When the monthly means are calculated these uncertainties are propagated through that process, so that each calculated monthly mean is provided with an uncertainty as well.

It is possible to estimate the uncertainty of the regression fit for regular multiple linear regression depending on how many data points are available and how well the basis functions can represent the variability that is to be described. Since the regression for calculation the BDBP data set is extended with several features that are not present in regular regressions (e.g. expansion in Legendre polynomials, usage of first guess patterns etc.), the regular way to estimate the uncertainty the regression results cannot be applied. At this stage, attempts to quantify the uncertainties on the BDBP data set have not been successful, so no uncertainties are available for the data set.

Comparisons with independent data sources (e.g. total column ozone data or profile data, see Section 5.3) can only provide a qualitative estimate on the accuracy of the data set since the independent data sources often do not cover all latitudes or all vertical levels.

An attempt to describe the BDBP data set in a more quantitative way is described in Hassler et al. (2013) where the squared differences between monthly means calculated from independent data sources and the BDBP were determined. Figures 5 and 6 show the results for these comparisons for BDBP with SAGE II and ozonesonde data.

4.2.3 Error Budget

The error budget for this CDR product has not been established for the global dataset. As described in the previous section, independent data sources such as ozonesonde data have been used to validate results. However, a comprehensive error budget has not been established for this product.

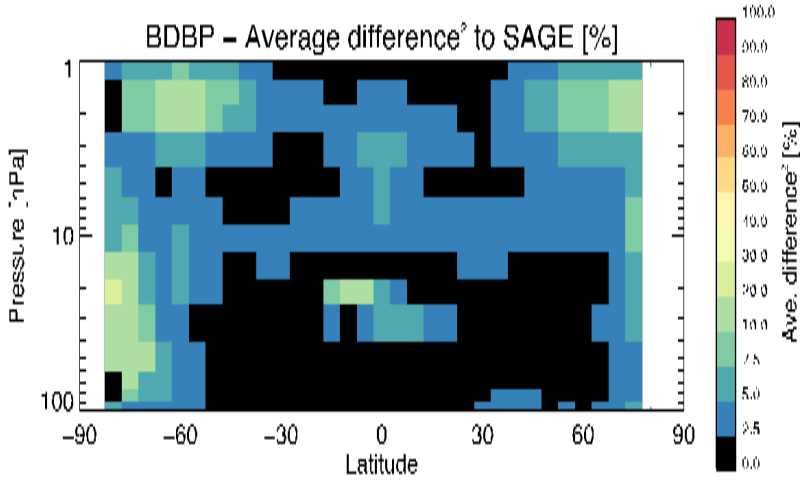


Figure 5: Average squared difference between all available monthly mean SAGE II values for given latitude bands and pressure levels, and the BDBP data set, given in percent. Colors changing from black to red indicate a change in average squared differences from 0% to 100%. (adopted from Hassler et al., 2013).

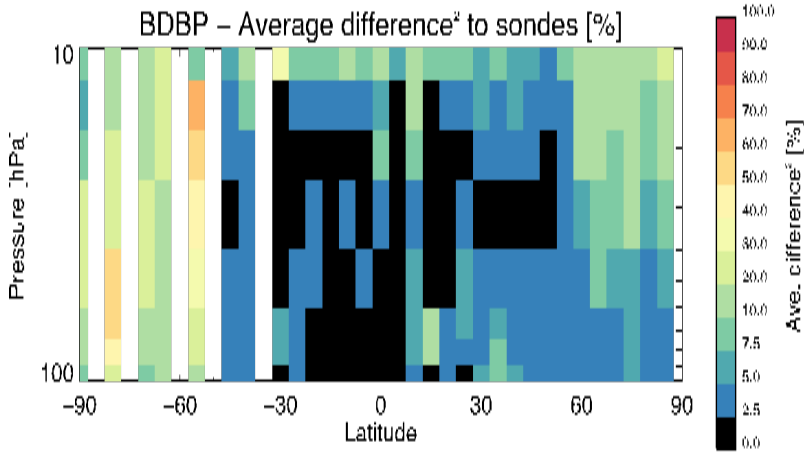


Figure 6: Average squared difference between all available monthly mean ozonesonde values for given latitude bands and pressure levels, and the BDBP data set, given in percent. Colors changing from black to red indicate a change in average squared differences from 0% to 100% (adopted from Hassler et al., 2013).

5. Practical Considerations

5.1 Numerical Computation Considerations

The whole algorithm is written in Delphi, in a very modular way. No steps are parallelized. The program needs less than two hours to run completely.

5.2 Programming and Procedural Considerations

File path in the main program have to be adjusted so that the program will find the necessary auxiliary data and procedures to be able to run. As soon as the program is compiled, Delphi creates an executable file that can be run outside a Delphi environment.

5.3 Quality Assessment and Diagnostics

Not all output products can be assessed easily for their quality. Tiers 1.1 to 1.3 reflect only parts of the observed ozone variability, so a comparison to independent measurements is not possible. Tier 1.4, however, can be compared to measurements from several different measurement systems to assess how well the regression result (Tier 1.4) is able to track observed ozone variability.

One part of the quality control during the creation process was described already in Section 4.1 where the integrated value of each monthly profile was compared to independent total column ozone measurements.

Another way of assessing the quality of Tier 1.4 obtained from the regression is to compare these profiles to actual ozone profile measurements, either used as input data for calculating Tier 0 (raw data) of the BDBP dataset (e.g. measurements from SAGE II, HALOE, POAM II, etc.), or from totally independent measurement systems (e.g. from the Solar Backscatter Ultraviolet Radiometer (SBUV)). The following figures (Figure 7 and 8) show some of these comparisons for different latitude zones and pressure levels, taken from Hassler et al. (2013).

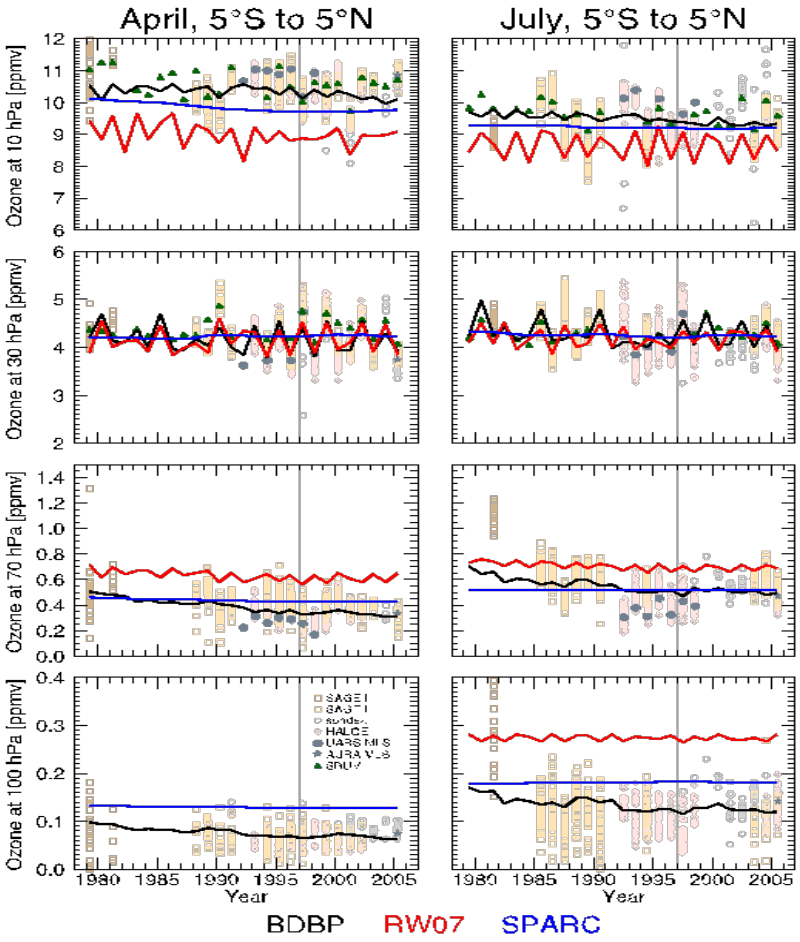


Figure 7: April (left) and July (right) ozone time series between 5°S to 5°N for four different pressure levels. Individual ozonesonde (grey open circles), SAGE I and II (orange open squares), HALOE (red open diamonds), and SBUV/SBUV2 (dark green, filled triangles), UARS MLS monthly means (dark grey, filled circles), AURA MLS monthly means (dark grey, filled stars), and the BDBP (Tier 1.4, black), RW07 (red) and SPARC (blue) time series are shown. The vertical grey line denotes the inflection point selected for piecewise linear trend calculations (taken from Hassler et al., 2013).

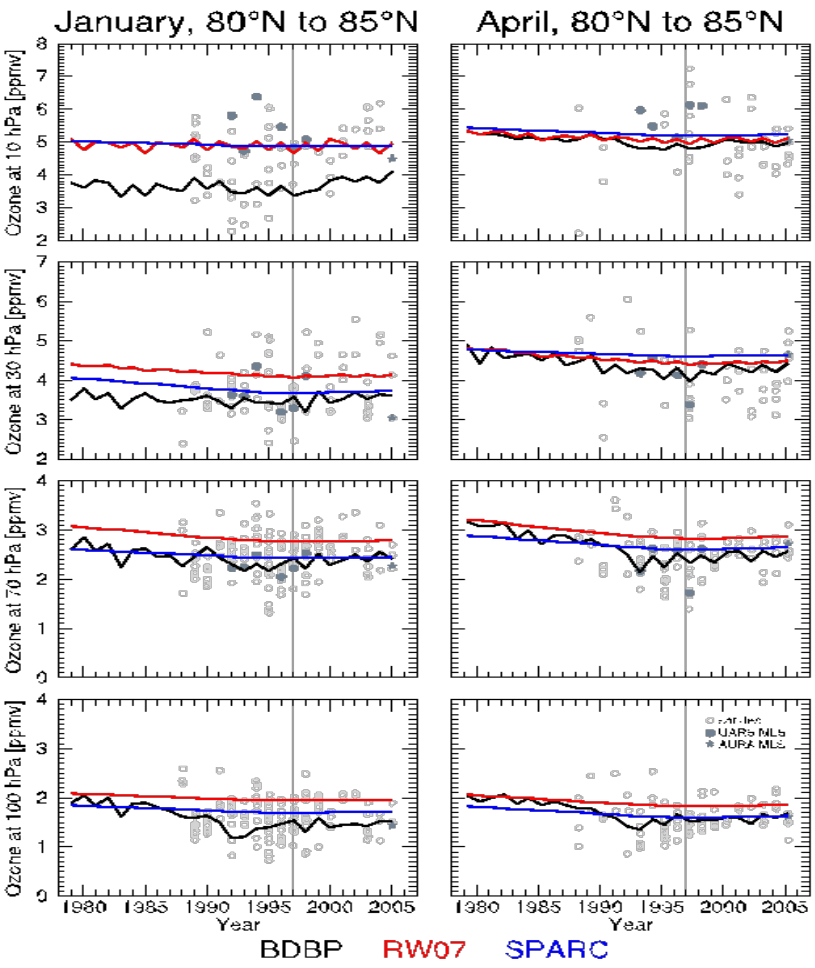


Figure 8: January (left) and April (right) ozone time series 80°N to 85°N for four different pressure levels. Individual ozonesonde measurements available in this zone are shown for these months (grey open circles), together with UARS MLS monthly means (dark grey, filled circles) and AURA MLS monthly means (dark grey, filled stars), and the BDBP (Tier 1.4, black), RW07 (red) and SPARC (blue) time series. The vertical grey line denotes the inflection point selected for piecewise linear trend calculations (taken from Hassler et al., 2013).

5.4 Exception Handling

The program will stop and print out an informative message for all known error conditions.

5.5 Algorithm Validation

A global gapfree ozone product has not been available or routinely produced.

This lack has precipitated the need to develop the Ozone BDBP product to better understand ozone measurements variability and the dynamics that impact these changes. While validation has been completed at select times and locations using ozonesonde data, the algorithms used in the model derived components have also been validated see Bodeker et al., (2013). Per those results, four monthly mean total column ozone time series updated from Fioletov et al. (2002) were evaluated. The integrated total column ozone from the Tier 1.4 database agrees well with the four independent time series over the Arctic. The suppression of total column ozone following the eruption of Mt. Pinatubo, seen in the independent observations, is tracked well, although the onset may be slightly too early. Through the late 1990s and early 2000s, the spring maximum in total column ozone is occasionally underestimated in the Tier 1.4 database, particularly in 1998, 1999, 2001, 2002, 2004 and 2006. This underestimation is likely the result of dynamically forced increases in ozone (Hadjinicolaou et al., 2005) which cannot be tracked in the regression model since it does not include an appropriate basis function.

Over the northern midlatitudes, the integrated Tier 1.4 time series tracks the independent total column ozone time series very well, but does show some underestimation of the springtime peak in many years. The suppression of ozone following the eruption of Mt. Pinatubo is captured well in the Tier 1.4 time series.

5.6 Processing Environment and Resources

Embarcadero Delphi XE 2010 was used to calculate the monthly means and run the regression model. It was run on a Windows 7 laptop, with a Windows XP operation system. The program ran maximum 2 hours to finish creating the data set.

6. Assumptions and Limitations

Assumptions: It is assumed that ozone variability is well enough understood to be able to describe it with a statistical model. It is further assumed that a comparison of the integration of ozone values over the profile with independent total column ozone data can provide insight about the quality and reality of the ozone profiles.

Limitations: Any ozone variability observed that cannot be described by the used basis functions, will not be present in the calculated data set. This means that only *known* sources of variability are included in the data set.

6.1 Algorithm Performance

N/A

6.2 Sensor Performance

N/A

7. Future Enhancements

There are no plans at this stage to enhance the algorithm.

8. References

Austin, J., Tourpali, K., Rozanov, E., Akiyoshi, H., Bekki, S., Bodeker, G.E., Brühl, C., Butchart, N., Chipperfield, M., Deushi, M., Fomichev, V.I., Giorgetta, M.A., Gray, L., Kodera, K., Lott, F., Manzini, E., Marsh, D., Matthes, K., Nagashima, T., Shibata, K., Stolarski, R.S., Struthers, H., and Tian, W. (2008). Coupled chemistry climate model simulations of the solar cycle in ozone and temperature. *J. Geophys. Res*., 113, D11306, doi:10.1029/2007JD009391.

Bodeker, G.E., Boyd, I.S., and Matthews, W.A. (1998). Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986–1996. *J. Geophys. Res*., 103, D22, 28661–28681.

Bodeker, G.E., Hassler, B., Young, P.J., and Portmann, R.W. (2013). A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations. *Earth Syst. Sci. Data*, 5(1), doi:10.5194/essd-5-31-2013, 31-43.

Bodeker, G.E., Scott, J.C., Kreher, K., and McKenzie, R.L. (2001). Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978–1998. *J. Geophys. Res*., 106, 23029–23042.

Daniel, J.S., Solomon, S., and Albritton, D.L. (1995). On the evaluation of halocarbon radiative forcing and global warming potentials. *J. Geophys. Res*., 100, 1271–1285.

Fioletov, V.E., Bodeker, G.E., Miller, A.J., McPeters, R.D., and Stolarski, R. (2002). Global and zonal total ozone variations estimated from ground-based and satellite measurements: 1964–2000. *J. Geophys. Res*., 107, 4647, doi:10.1029/2001JD001350, 2002.

Glaccum, W., Lucke, R., Bevilacqua, R.M., Shettle, E.P., Hornstein, J.S., Chen, D.T., Lumpe, J.D., Krigman, S.S., Debrestian, D.J., Fromm, M.D., Dalaudier, F., Chassefiere, E., Deniel, C., Randall, C.E., Rusch, D.W., Olivero, J.J., Brogniez, C., Lenoble, J., and Kremer, R. (1996). The Polar Ozone and Aerosol Measurement (POAM II) Instrument. *J. Geophys. Res*., 101(D9), 14 479–14 487.

Harris, N., Hudson, R. and Phillips C. (Eds.) (1998). Assessment of trends in the vertical distribution of ozone. SPARC Rep. 1, Ozone Res. and Monit. Proj. Rep. 43, World Meteorol. Organ., Geneva.

Hassler, B., Bodeker, G.E., and Dameris M. (2008). Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements. *Atmos. Chem. Phys*., 8(17), 5403-5421.

Hassler, B., Young, P.J., Portmann, R.W., Bodeker, G.E., Daniel, J.S., Rosenlof, K.H., and Solomon, S. (2013). Comparison of three vertically resolved ozone data sets: climatology, trends and radiative forcings. *Atmos. Chem. Phys*., 13(11), 5533–5550.

Hassler, B. *et al.* (2014). SI2N overview paper: ozone profile measurements: techniques, uncertainties and availability. *Atmos. Meas. Tech. Discuss*., 6, doi:10.5194/amtd-6-9857-2013, 9857–9938.

Hadjinicolaou, P., Pyle, J. A., and Harris, N. R. P. (2005). The recent turnaround in stratospheric

ozone over northern middle latitudes:A dynamical modeling perspective, Geophys. Res. Lett., 32, L12821, doi:10.1029/2005GL022476.

Kiladis, G.B., Straub, K.H., Reid, G.C., and Gage, K.S. (2001). Aspects of interannual and intra seasonal variability of the tropopause and lower stratosphere. *Q. J. Roy. Met. Soc*., 127, 1961–1983.

Lucke, R.L., Korwan, D., Bevilacqua, R.M., Hornstein, J.S., Shettle, E.P., Chen, D.T., Daehler, M., Lumpe, J.D., Fromm, M.D., Debrestian, D., Neff, B., Squire, M., König-Langlo, G., and Davies, J. (1999). The Polar Ozone and Aerosol Measurement (POAM III) Instrument and Early Validation Results. *J. Geophys. Res*., 104(D15), 18 785–18 799.

Lumpe, J.D., Bevilacqua, R.M., Hoppel, K.W., Krigman, S.S., Kriebel, D.L., Randall, C.E., Rusch, D.W., Brogniez, C., Ramananaherosa, R., Shettle, E.P., Olivero, J.J., Lenoble, J., and Pruvost, P. (1997). POAM II Retrieval Algorithm and Error Analysis. *J. Geophys. Res*., 102(D19), 23 593–23 614.

Lumpe, J.D., Bevilacqua, R.M., Hoppel, K.W., and Randall, C.E. (2002). POAM III Retrieval Algorithm and Error Analysis. *J. Geophys. Res*., 107(D21), 4575, doi:10.1029/2002JD002137.

McCormick, M.P., Zawodny, J.M., Veiga, R.E., Larsen, J.C., and Wang, P.-H. (1989). An overview of SAGE I and SAGE II ozone measurements. *Planet. Space Sci*., 37(12), 1567–1586.

Nakajima, H. (2006). Preface to special section on ILAS-II: The Improved Limb Atmospheric Spectrometer-II. *J. Geophys. Res*., 111, D20S90, doi:10.1029/2006JD007412.

Newman, P.A., Daniel, J.S., Waugh, D.W., and Nash, E.R. (2007). A new formulation of equivalent effective stratospheric chlorine (EESC). *Atmos. Chem. Phys*., 7, 4537–4552, doi:10.5194/acp-7-4537-2007.

Press, W.H., Flannery, B.R., Teukolsky, S.A., and Vettering, W.T. (1998). Numerical Recipes in Pascal, Cambridge Univ. Press, New York, 759 pp.

Randel, W.J. and Wu, F. (2007). A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data. *J. Geophys. Res*., 112, D06313, doi:10.1029/2006JD007339.

Remsberg, E.E., Russell III, J.M., Gille, J.C., Gordley, L.L., Bailey, P.L., Planet, W.G., and Harries, J.E. (1984). The validation of Nimbus 7 LIMS measurements of ozone. *J. Geophys. Res*., 89, 5161–5178.

Rosenfield, J.E., Considine, D.B., Meade, P.E., Bacmeister, J.T., Jackman, C.H., and Schoeberl, M.R. (1991). Stratospheric effects of Mount Pinatubo aerosol studied with a coupled two-dimensional model. *J. Geophys.Res*., 102, 3649-3670.

Russell, J.M. III, Gordley, L.L., Park, J.H., Drayson, S.R., Hesketh, D.H., Cicerone, R.J., Tuck, A.F., Frederick, J.E., Harries, J.E., and Crutzen, P. (1993). The Halogen Occultation Experiment. *J. Geophys. Res*., 98(D6), 10 777–10 797.

Sasano, Y., Suzuki, M., Yokota, T., and Kanzawa, H. (1999). Improved limb atmospheric spectrometer (ILAS) for stratospheric ozone layer measurements by solar occultation technique. *Geophys. Res. Lett.*, 26, 197-200.

Smit, H.G.J. (2002). Ozonesondes, in: Encyclopedia of Atmospheric Sciences, edited by: Holton, J., Pyle, J., and Curry, J., Academic Press, London, 1469–1476, 2002.

Wang, H.J., Cunnold, D.M., and Bao, X. (1996). A critical analysis of Stratospheric Aerosol and Gas Experiment ozone trends. *J. Geophys. Res*., 101, D7, 12495–12514.

Wang, P.-H., Cunnold, D.M., Trepte, C.R., Wang, H.J., Jing, P., Fishman, J., Brackett, V.G., Zawodny, J.M., and Bodeker, G.E. (2006). Ozone variability in the midlatitude upper troposphere and lower stratosphere diagnosed from a monthly SAGE II climatology relative to the tropopause. *J. Geophys. Res*., 111, D21304, doi:10.1029/2005JD006108.

Waugh, D.W. and Hall, T.M. (2002). Age of stratospheric air: theory, observations and models. *Rev. Geophys*., 40, 1010, doi:10.1029/2000RG000101.

Zawodny, J.M., and McCormick, M.P. (1991). Stratospheric Aerosol and Gas Experiment II measurements of the quasi-biennial oscillation in ozone and nitrogen dioxide. *J. Geophys. Res*., 96, 9371-9377.

Zerefos, C.S., Tourpali, K., Bojkov, B.R., Balis, D.S., Rognerund, B., and Isaksen, I.S.A. (1997). Solar activity-total column ozone relationships: Observations and model studies with heterogeneous chemistry. *J. Geophys. Res*., 102, 1561–1569.

Appendix A. Acronyms and Abbreviations

Acronym or Abbreviation	Definition
AOGCM	Atmosphere-Ocean General Circulation Model
AURA	AURA satellite
BDBP	Binary DataBase of Profiles
C-ATBD	Climate Algorithm Theoretical Basis Document
CDR	Climate Data Record
CPC	Climate Prediction Center
DU	Dobson Unit
EESC	Equivalent Effective Stratospheric Chlorine
ENSO	El Niño Southern Oscillation
ESRL	Earth System Research Laboratory
FG	First Guess of the regression model fit coefficient
FTIR	Fourier Transform Infrared Spectrometer
HALOE	Halogen Occultation Experiment
ILAS	Improved Limb Array Spectrometer
LIMS	Limb Infrared Monitor of the Stratosphere
MLS	Microwave Limb Sounder
NCDC	National Climatic Data Center
NDACC	Network for the Detection of Atmospheric Composition Change
NGDC	National Geophysical Data Center
NOAA	National Oceanic and Atmospheric Administration
POAM	Polar Ozone and Aerosol Measurement
PP	Pertubation Pattern
QBO	Quasi-Biennial Oscillation
SAGE	Stratospheric Aerosol and Gas Experiment
SBUV	Solar Backscatter Ultraviolet Radiometer
SOI	Southern Oscillation Index
SPARC	Stratosphere-troposphere Processes And their Role in Climate
UARS	Upper Atmosphere Research Satellite
WOUDC	World Ozone and Ultraviolet Radiation Data Centre