White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/104096/

Version: Published Version

Article:
Sofen, E. D., Bowdalo, D., Evans, M. J. et al. (2016) Gridded global surface ozone metrics for atmospheric chemistry model evaluation. Earth System Science Data. pp. 41-59. ISSN: 1866-3516

https://doi.org/10.5194/essd-8-41-2016

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Earth Syst. Sci. Data, 8, 41–59, 2016
www.earth-syst-sci-data.net/8/41/2016/
doi:10.5194/essd-8-41-2016
© Author(s) 2016. CC Attribution 3.0 License.

Open Access Earth System Science Data

Gridded global surface ozone metrics for atmospheric chemistry model evaluation

E. D. Sofen[1], D. Bowdalo[1], M. J. Evans[1,2], F. Apadula[3], P. Bonasoni[4], M. Cupeiro[5], R. Ellul[6], I. E. Galbally[7], R. Girgzdiene[8], S. Luppo[9], M. Mimouni[10], A. C. Nahas[11], M. Saliba[6], and K. Tørseth[12]

[1]Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK
[2]National Centre for Atmospheric Sciences, University of York, Heslington, York, YO10 5DD, UK
[3]Ricerca sul Sistema Energetico – RSE, via R. Rubattino 54, 20134, Milan, Italy
[4]Institute of Atmospheric Sciences and Climate, Italian National Research Council (ISAC-CNR), Bologna, Italy
[5]Servicio Meteorológico Nacional, Ushuaia GAW Station, Tierra del Fuego, Argentina
[6]Department of Geosciences, University of Malta, Msida MSD 2080, Malta
[7]CSIRO Oceans and Atmosphere Flagship, Aspendale Vic. 3195, Australia
[8]State Research Institute Center for Physical Sciences and Technology, Savanorių ave. 231, 02300 Vilnius, Lithuania
[9]Universidad Nacional de Tierra del Fuego, ICPA, Tierra del Fuego, Argentina
[10]Algerian Meteorology Service, P.O. Box 31, Tamanrasset, Algeria
[11]Global Atmosphere Watch (GAW) Station Bukit Kototabang, The Indonesian Agency for Meteorological, Climatological, and Geophysics (BMKG), Jl. Raya Bukittinggi – Medan Km. 17 Palupuh, Agam, West Sumatera, Indonesia
[12]NILU – Norwegian Institute for Air Research, P.O. Box 100, 2027 Kjeller, Norway

Correspondence to: E. D. Sofen (esofen@gmail.com) and M. J. Evans (mat.evans@york.ac.uk)

Received: 1 July 2015 – Published in Earth Syst. Sci. Data Discuss.: 30 July 2015
Revised: 22 December 2015 – Accepted: 22 January 2016 – Published: 11 February 2016

Abstract. The concentration of ozone at the Earth's surface is measured at many locations across the globe for the purposes of air quality monitoring and atmospheric chemistry research. We have brought together all publicly available surface ozone observations from online databases from the modern era to build a consistent data set for the evaluation of chemical transport and chemistry-climate (Earth System) models for projects such as the Chemistry-Climate Model Initiative and Aer-Chem-MIP. From a total data set of approximately 6600 sites and 500 million hourly observations from 1971–2015, approximately 2200 sites and 200 million hourly observations pass screening as high-quality sites in regionally representative locations that are appropriate for use in global model evaluation. There is generally good data volume since the start of air quality monitoring networks in 1990 through 2013. Ozone observations are biased heavily toward North America and Europe with sparse coverage over the rest of the globe. This data set is made available for the purposes of model evaluation as a set of gridded metrics intended to describe the distribution of ozone concentrations on monthly and annual timescales. Metrics include the moments of the distribution, percentiles, maximum daily 8-hour average (MDA8), sum of means over 35 ppb (daily maximum 8-h; SOMO35), accumulated ozone exposure above a threshold of 40 ppbv (AOT40), and metrics related to air quality regulatory thresholds. Gridded data sets are stored as netCDF-4 files and are available to download from the British Atmospheric Data Centre (doi:10.5285/08fbe63d-fa6d-4a7a-b952-5932e3ab0452). We provide recommendations to the ozone measurement community regarding improving metadata reporting to simplify ongoing and future efforts in working with ozone data from disparate networks in a consistent manner.

Published by Copernicus Publications.

1 Introduction

Tropospheric ozone (O_3) is an air pollutant that is detrimental to human health and vegetation, an important atmospheric oxidant, and a short-lived greenhouse gas. Ozone is a secondary species; it is not directly emitted, rather, it is produced from the photochemical oxidation of non-methane volatile organic compounds (VOCs), methane (CH_4), or carbon monoxide (CO) in the presence of nitrogen oxides (NO_x). Ozone is also produced in the stratosphere (from the photolysis of oxygen) and transported into the troposphere. Ozone is destroyed both photochemically and through deposition to the surface (Monks et al., 2015, and references therein). Chemical transport models attempt to describe the physical, chemical, and biological processes important for determining the composition of the troposphere. A strong scientific community has focused heavily on the modeling of ozone because of its importance in atmospheric chemistry, air quality, and climate. As well as providing a tool for the scientific understanding of atmospheric chemistry these models are used to develop air quality and climate mitigation policies. Our confidence in the unbiased nature of these models and hence the value of these models in developing mitigation policies is assessed by comparing model simulations against observations. Comparisons of model simulations to observations for ozone are central to this evaluation.

In many ways, this evaluation has been limited by the availability of data. For example, field campaign based aircraft observations can offer some vertical information but they are limited in space and time (Parrington et al., 2013; Shim et al., 2008). Ozonesondes, on the other hand, provide near global coverage and vertical information but they are released infrequently and are spatially sparse (Logan, 1999; Tilmes et al., 2012; Liu et al., 2013).

Model evaluation based on satellite observations of ozone can provide some degree of global coverage at high spatial resolution (depending on orbital parameters), but their interpretation is complicated by the radiative transfer calculation central to the observation, which can only provide vertical resolution that can, at best, separate the upper troposphere from the lower troposphere (Chandra et al., 2003; Ziemke et al., 2006; Zhang et al., 2010).

Much of the model-measurement comparison has relied upon comparisons with surface observations. Surface observations have the advantage of occurring at relatively high time resolution (typically hourly) and at fixed locations that can, for some sites, go back decades with very stable instruments. Surface observations also have the advantage of observing the portion of the atmosphere that is most relevant to human and plant life. Many of these observations are collected as part of regional or global networks that aim to monitor either the background state of the atmosphere or air quality regulatory compliance.

A range of comparisons between surface ozone and model predictions have been made over the last decades. Some of these have been to investigate the performance of a single model; some as part of a multimodel comparison. All of these models have in some way had to limit their comparison to some form of metric for model evaluation. The annual mean ozone concentration is probably the most basic assessment for the model performance (e.g. Voulgarakis et al., 2009). The comparison of monthly mean ozone mixing ratios has probably been studied most extensively (e.g Wang et al., 1998; Fiore et al., 2009; Voulgarakis et al., 2009), as storage of model output at higher temporal resolutions has historically been considered expensive. Studies focused on air quality, often using regional models, typically devote more attention to metrics related to air quality regulations, extreme ozone events, and human health. These metrics include the maximum daily 8 h average (MDA8) ozone concentration (e.g. Jacob and Winner, 2009; Reidmiller et al., 2009; Pfister et al., 2014; Simon et al., 2015), 95th and 99th percentiles (e.g. Pfister et al., 2014), histogram of hourly concentrations (Solazzo et al., 2012), the W126 sigmoidal weighting function for human ozone exposure (Lefohn and Runeckles, 1987; Lapina et al., 2014), the cumulative exposure to mixing ratios above 35 ppbv (1 ppbv = 1 nmol mol^{-1}) (SOMO35) (Amman et al., 2005; Colette et al., 2012), and the AOT40 plant exposure metric (Ashmore and Wilson, 1994).

A consistent challenge in model evaluation is the appropriate way to relate point observations to model output that represents the average over a grid box that may be hundreds of kilometers on a side. One method is to average a number of observations over a wider region in the expectation that potential biases from a single site will be minimized. For example, Fiore et al. (2009) compared the average seasonal cycles for nine different regions across the US, Europe, and Japan between monthly averaged observations from the EPA CASTNET, EMEP, and EANET data sets and 21 different models. This work represents one of the larger efforts in terms of combining different ozone data sets, but still neglects much of the world, essentially restricting the evaluation to "regional background" sites in industrialized Northern Hemispheric continental regions. Several more detailed regional efforts have been undertaken to evaluate regional air quality models against US and European regulatory networks (Appel et al., 2007; Emery et al., 2012; Kim et al., 2015; Colette et al., 2012; Katragkou et al., 2015). Schnell et al. (2014) used an inverse distance interpolation algorithm to interpolate EPA AQS data for the US and EMEP and AirBase data for Europe to construct continuous fields of surface ozone to evaluate the ability to model extreme air pollution events over the time period 2000–2009. Other model evaluations adopt the approach of interpolating the model to the locations of the measurements, including a vertical correction to account

for the influence that ozone deposition to the surface has on near-surface concentrations (Brown-Steiner et al., 2015).

Most previous model evaluations have emphasized one regional regulatory or background network for their comparisons. Regional models are evaluated over the domain for which they have been developed in support of air quality policy. In the evaluation of global models, the comparison with a single network, while expedient, has tended to ignore the other networks and data sets which obviously limits the veracity of the comparison. In this work, we describe a combined global surface ozone data set that brings together all readily available public modern surface ozone data beginning in 1971 with the WMO GAW network (other networks start between 1974 and 2000) and ending with complete data through 2013. The objective is to provide a consistent, homogenized data set and metrics that are easily accessible to the atmospheric chemistry modeling community, primarily for global model evaluation, as part of efforts such as the Chemistry Climate Model Initiative (CCMI) and Aerosols Chemistry Model Intercomparison Project (AerChemMIP). In Sect. 2, we describe the contributing data sets, and Sect. 3 describes the data processing to construct a unified data set and issues that arise during the processing. Section 4 describes the spatial and temporal extent of the data set and Sect. 5 provides a statistical summary of the ozone data. Section 6 describes the production of the gridded version of the data set that is being distributed to other modeling groups and provides examples of model-data comparison. Finally, in Sect. 7, we provide some recommendations to the measurement, data set manager, and modeling communities.

2 Contributing data sets

As described in Sect. 1 there are a variety of significant, publicly available surface ozone data sets. For simplicity and efficiency we define "significant, publicly available" as being complete data sets that are directly available from the internet, extend over more than 10 sites and over more than 5 years, report concentration in either ppbv or $\mu g\,m^{-3}$ (i.e. not air quality index), which extend beyond solely being for national in-city pollution evaluation. We have implemented a semi-automated system for processing the ozone observations from these networks, dealing with the disparate file formats and discrepancies in the data and metadata, to produce a single data set suitable for model evaluation. The combined data set consists of data from the following data sets:

- World Data Center for Greenhouse Gases (WDCGG; http://ds.data.jma.go.jp/gmd/wdcgg/) from the World Meteorological Organization (WMO) Global Atmospheric Watch (GAW; http://www.wmo.int/pages/prog/arep/gaw/gaw_home_en.html),
- Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP; http://ebas.nilu.no/default.aspx),
- European Environment Agency AirBase (http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8),
- US Environmental Protection Agency Clean Air Status and Trends Network (US EPA CASTNET; ftp://ftp.epa.gov/castnet/data),
- US EPA Air Quality System (AQS; http://www.epa.gov/airquality/airdata/ad_data.html),
- Environment Canada's Air and Precipitation Monitoring Network (CAPMoN; https://www.ec.gc.ca/rs-mn/default.asp?lang=En&n=6C8C66C5-1),
- Canadian National Air Pollution Survey Program (NAPS; http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx), and
- Acid Deposition Monitoring Network in East Asia (EANET; http://www.eanet.asia).

Data were last downloaded on 29 June 2015. All told, the data set represents 7288 station records, with measurements between 1971 through 2014 totaling 578 041 289 hourly surface observations.

The WMO GAW observations are made at sites across the globe with the goal of "maintaining and applying global, long-term observations of the chemical composition and selected physical characteristics of the atmosphere and emphasizing quality assurance and quality control" (Müller et al., 2007). GAW sites are classified into global, regional, and contributing stations. Global stations are defined as sites that "contribute data required to address global environmental issues of global scale and importance". Regional stations are designed "primarily to address regional aspects of global environmental issues and environmental problems of regional scale and importance". Contributing stations "belong to other organizations or international programs" and share data through mutual agreements (Müller et al., 2007). GAW ozone observations begin in 1971 at Hohenpeissenberg and now represent 108 stations. We include all of the global, regional and contributing stations in this analysis with provisions to handle duplicate sites in multiple data sets (see Sect. 3.10).

EMEP was founded "to provide sound scientific support of the Convention on Long-range Transboundary Air Pollutants (LRTAP)" to convention member states (EMEP Steering Body, 2012). EMEP sites are generally intended to provide representative regional observations to monitor long-range transport in Europe and constitute 203 stations, with the earliest records beginning in 1977 (Tørseth et al., 2012).

The European Environment Agency AirBase data set is made up of national air pollution monitoring networks from 40 European countries and some overseas territories (e.g. Azores), comprising a total of 3711 stations (European Environment Agency, 2002; European Topic Centre on Air Pollution and Climate change mitigation, 2015). Many of these stations are in urban areas and are filtered out of the data set in order to produce a final data set that represents regional average consistent with global chemical transport models (see Sect. 3.11). There are some common sites between AirBase and EMEP, which are addressed by removing the duplicate sites with shorter records (see Sect. 3.10).

The US EPA CASTNET is a network of 126 regional background stations constructed to assess trends in air quality (AMEC Environment and Infrastructure, Inc., 2014). Since 2011, it has also contributed to US air quality regulatory monitoring and is included in the EPA AQS database as well.

The EPA AQS ozone measurements are used to monitor compliance with the Clean Air Act. The data set consists of 2326 stations. Like AirBase, AQS also includes many urban sites that are removed in this analysis (see Sect. 3.11).

In parallel with the structure of the US EPA networks, Environment Canada operates the CAPMoN background network "designed to study the regional patterns and trends of atmospheric pollutants such as acid rain, smog, particulate matter and mercury, in both air and precipitation", with 19 sites and the NAPS air quality regulatory network "to monitor and assess the quality of ambient (outdoor) air in the populated regions of Canada" with 369 sites.

EANET is an intergovernmental monitoring network, primarily focused on acid deposition, that includes hourly ozone observations from a total of 16 sites in Japan, Malaysia, and Thailand. EANET has ozone measurements from a number of other sites across Asia, but they are provided at daily or monthly resolution and are not included in our database at this point.

We acknowledge that other surface ozone data sets do exist especially in regions unserved by a major monitoring network. However they are (as far as we can tell) not readily available, unlikely to conform to the quality assurance standards followed by the above networks, or not be in a standard data format. For example, one notable gap in this data set is Australia. Australia has similar air quality legislation to the US, Canada, and Europe, but measurements are collected and stored at the state level. States differ in what data they make accessible and in data format. Archived data from past years is generally not readily available.

In Sect. 7 we suggest that all surface ozone monitoring data sets should be reported to a central repository such as that offered by the WMO GAW, or a data portal and access interface that can provide data in a uniform manner, so that both the wider community and the groups making the measurements can benefit from broad consistency in the handling and formatting of ozone data.

3 Methods: data set processing

Here, we describe the data processing scheme, the related data quality issues revealed during processing, and how the issues are addressed in the effort to homogenize these data sets. Most of the data quality issues have been resolved and are described here, but a number of ambiguities remain. In Sect. 7, we provide some recommendations to the measurement community that we hope will improve data consistency in the future. Appendix A provides specific comments on a site by site basis.

3.1 Initial file parsing

First, each file is parsed into a time series for each site. The ozone networks provide data in formats ranging from NASA Ames (EMEP) and other fixed format text files (WMO GAW) to comma-separated value (CSV) files laid out in a variety of ways (CAPMON, NAPS, AQS, CASTNET, and EANET). WMO GAW, EMEP, CAPMON, and EANET provide one file per site per year. AQS and NAPS provides one file per year that includes all sites. WMO GAW, EMEP, and CAPMON include metadata within each site/year file. All CASTNET data are provided in a single file, with a separate CSV file providing site-specific metadata. EPA AQS data are provided with one file per year including all sites and their respective metadata. EANET is provided with one file per site per year, but metadata text has to be stripped out of a PDF file containing site descriptions and manually cleaned-up prior to processing the EANET data. NAPS data are provided with one CSV file for each year including all sites plus a separate metadata file. AirBase provides data in separate files by year by site. AirBase metadata is provided via CSV files of stations by country, a corresponding list of measurement configurations, and a single XML file that provides metadata that applies to all sites within each country (e.g. time zone).

Figure 1 illustrates the data processing scheme that we follow. Data processing is done using code written in Python that relies heavily on the Pandas time series routines (v.0.15.0; http://pandas.pydata.org) (McKinney, 2010) for efficient file parsing, data gap handling, and time zone conversion, as well as for later quality control and calculating metrics. Simple file-by-file quality control is performed at this stage to collect metadata, and adjust time and units for consistency.

3.2 Metadata

A consistent dictionary of metadata describing the location and characteristics of each site is collected for each site. In general, this metadata includes latitude, longitude, altitude, site name, site abbreviation, a contact person, measurement unit, and time zone. Additional metadata may include the file version or revision date, contributing agency, geopolitical location information (nation, state, city), land use, classifica-

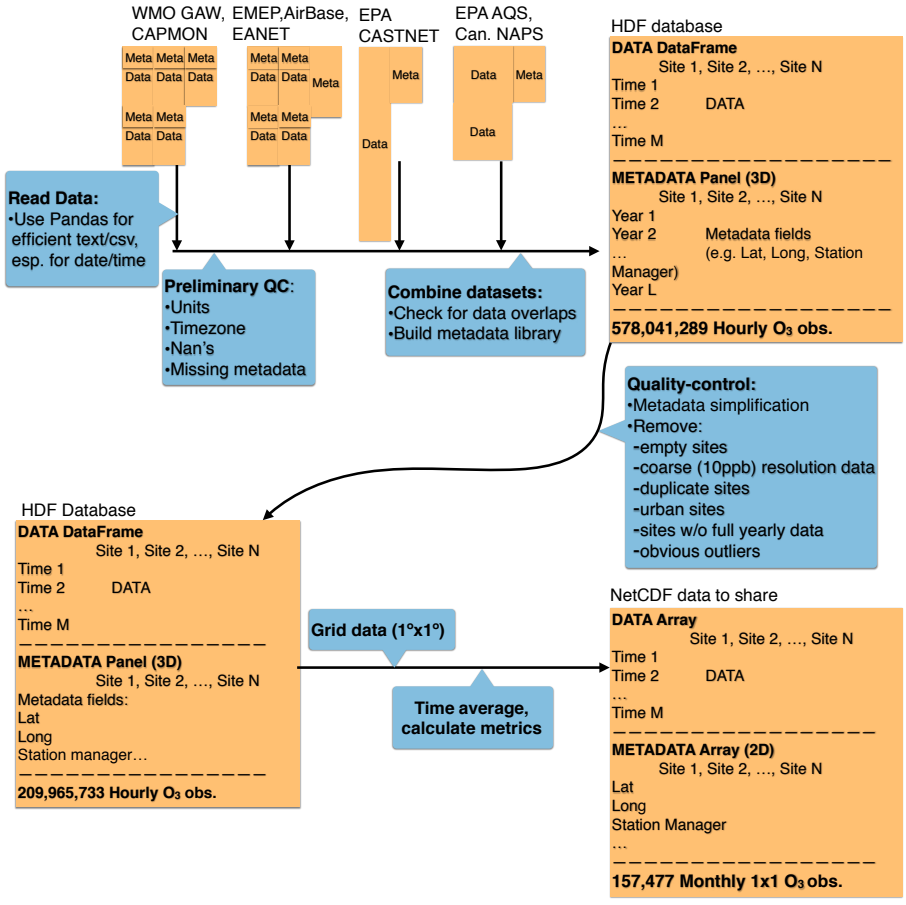

Figure 1. Flow diagram illustrating the data processing strategy. Orange represents file structures. "Meta" and "Data" icons are intended to graphically represent the various structures of the input data sets. Blue represents actions/functions applied to the data.

tion of the observing site (e.g. urban/rural or global/regional), and detection limit.

At sites where location information is provided on a yearly basis, there are cases where the latitude, longitude, and altitude information change from year to year. This makes handling all sites in an automated fashion more complicated. In some cases, these changes in metadata are real, due to a minor change in station location. In many other cases, the changes are due to typographical errors or varying numbers of digits being used to represent the latitude and longitude in decimal notation. In general, the variations in latitude and longitude associated with real site location changes are small enough that they will not impact the ability to use the data for global model evaluation. Differences in metadata for a single site are addressed by finding the most-commonly occurring metadata value and assigning that in the consolidated metadata library (described below).

Data set-specific site identification codes are used throughout the data processing to link data to metadata for a given site. The EPA AQS and EANET data sets do not include such site identification codes. For the EPA AQS data set, site IDs were created based on the state code, county code, and state-specific site-number to produce a unique nine-digit number preceded by "AQS", consistent with the EPA AQS site ID used in the AirData system. For EANET, site IDs were produced from a combination of "EA", the first two letters of the site name, and a three-digit integer.

3.3 Date and time

Next, the series of time stamps and ozone concentration and/or mixing ratio data are read. Date and time information is converted to a Python Numpy "datetime64" object (Walt et al., 2011) that allows for automated handling of many time-related calculations in Python.

Date information in files appears to be generally handled consistently across sites. However, a couple of sites exhibit consistent gaps on 29 February on leap years, indicating issues with how leap years are being handled by data preparation routines. We assume that this is simply a missing day of data and the data labeled 1 March is in fact for 1 March.

Representations of the time of day reported within the files, on the other hand, vary widely and exhibit a number of problems. In order to compare observational time series

Table 1. Number of valid observations after quality control steps.

Quality control step	Number of sites	Number of observations
Original combined data set	7288	578 041 289
Simplify metadata (drop sites w/o data)	7250	578 041 289
Outlier values	7250	578 030 085
Coarse resolution	6962	549 205 906
Duplicate sites	6669	528 557 165
Urban sites	3451	255 367 238
No data in some months	2541	225 143 878
Less than a year of data or gapped series	2531	209 965 733

to models, the observations need to be adjusted to a consistent time frame. In our work, we use UTC. The time coordinate is adjusted to Coordinated Universal Time (UTC) based on time zone metadata. Some WMO GAW stations report data in UTC, others in local time, often expressed in ambiguous terms (e.g. "`Local time UTC + 1`" vs. "`Local time + 1`"). Many GAW sites do not list any time zone information, making it ambiguous whether they are in UTC or local time. A few sites, listed in Table A1 in Appendix A, were found to list time zones that are inconsistent with those listed on GAWSIS (http://gaw.empa.ch/gawsis/). These inconsistencies have been communicated to staff at the WDCGG and the German sites have all been corrected to UTC + 1. Some GAW sites also shift from local time or not specifying a time zone to specifying UTC from one year to the next, leading to potential problems with data gaps or overlaps, as these transitions are not always handled consistently by the data providers. These sites are checked manually, and unless there is compelling evidence to suggest a particular time zone, UTC is assumed. EPA CASTNET data are provided in "local standard time" as constant offsets from UTC (without daylight savings time) although the time zone details specified in the metadata explicitly note daylight savings time, leading to potential confusion. The EANET and NAPS data sets provide integer time zone offsets from UTC. In AirBase, time zone information is provided as fixed offsets from UTC by country. EMEP, CAPMoN, and AQS data are provided in UTC. The databases that use UTC time provide an opportunity to check the time zone conversions applied to duplicate or nearby sites from other databases.

Times are also listed either as 00:00–23:00 or 01:00–24:00. While strictly speaking, ISO-8601 time format specifications allow use of both 0 and 24, date-time routines in Python and most other programming languages prefer 00:00–23:00 and many do not handle 24:00. For sites that run 01:00–24:00, we simply convert 24:00 to 00:00 of the next day. Associated with this difference in notation is an inherent ambiguity as to whether a time value corresponds to the beginning or end of an hour-long period of measuring ozone. That is, a time stamp of 02:00 may refer to the average of measurements from 01:01–02:00 or from 02:00–02:59 and represents an inherent uncertainty in the time of day when a measurement is recorded.

A handful of sites, most notably the Nepal Climate Observatory – Pyramid on Mount Everest record hourly data but not aligned with integer hours of the day. In the case of Pyramid, it is because the local time zone is UTC + 05:45.

While these shifts in timing are likely insignificant for model evaluation metrics such as monthly averages, they represent significant problems for using the data set for other applications. Plant damage due to exposure to ozone is significantly enhanced during the hours of sunshine when stomata are open; photochemical activity is dependent upon the time of day; time series methods such as spectral analysis may give different results if the correct time of day is not used.

Once sites with ambiguous, no data, or data reported off the hour are excluded, the total number of sites is reduced from 6694 to 6446.

3.4 Concentrations and mixing ratios

Most science applications consider the "concentration" of ozone in terms of mixing ratios (ppbv = nmol mol^{-1}), whereas some policy applications consider it in terms of a mass density (μg m^{-3}). Here, we convert observations in μg m^{-3} to mixing ratios in ppbv. Although this may appear to be a trivial exercise, it is not without uncertainties.

Fundamentally the ozone observations used here take advantage of the Beer-Lambert Law to measure ozone through UV absorption spectrophotometry. This relates the absorbance at the 253.65 nm line of mercury to the concentration (in molecules cm^{-3}), the length of the measurement cell (cm), and the ozone absorbance cross section (cm^2 molecule^{-1}). Thus the instrument fundamentally measures the concentration of ozone in molecules cm^{-3}. The conversion to μg m^{-3} is a straight forward constant ($M(O_3)/N_A \times 10^6$), where $M(O_3)$ is the molar mass of ozone and N_A is Avogadro's number. However, if the instrument is reporting in ppbv it must use the atmospheric number density. The conversion from μg m^{-3} to ppbv depends on both pressure and temperature:

$$X(O_3) = C(O_3) \times \frac{R\,T}{M(O_3)\,P}, \quad (1)$$

where $X(O_3)$ refers to the ozone mixing ratio in ppbv, $C(O_3)$ the ozone concentration in $\mu g\,m^{-3}$, R is the gas constant ($8.3144\,J\,mol^{-1}\,K^{-1}$), $M(O_3)$ is the molar mass of ozone ($48\,g\,mol^{-1}$), P is pressure in mPa, and T is temperature in K. The temperature refers to the bench temperature of the instrument and the pressure refers to the internal pressure in the measurement cell, not to ambient conditions. However, when using concentrations they are not reported at the actual temperature and pressure of a given instrument, but are adjusted in-instrument to a user-programmable fixed standard temperature and pressure. This is generally 20 or 25 °C and 1013.25 hPa, but instrument default settings vary and often the assumed temperature is not reported in metadata. We assume a temperature of 20 °C for sites where no other information is specified, based on a survey of instrument default settings and comparison to surrounding sites. AirBase data are converted to ppbv using a conversion factor of 0.501 ppbv $(\mu g\,m^{-3})^{-1}$ based on a temperature of 293 K and pressure of 1013 hPa specified in EU legislation for the AirBase data collection (European Environment Agency, 2002). The conversion factor used may be up to 0.510 ppbv $(\mu g\,m^{-3})^{-1}$ at 25 °C. This ambiguity alone can contribute an uncertainty of up to 0.6 ppbv on a typical ozone concentration of $60\,\mu g\,m^{-3}$ (30 ppbv).

On a Thermo Fisher Scientific Model 49i UV Photometric Ozone Analyzer (Thermo Fisher Scientific Inc., 2011) the instrument outputs ppmv or ppbv by default, with the option to switch to $\mu g\,m^{-3}$ with a specified, fixed temperature and pressure. Ozone observing sites whose primary purpose is regulatory monitoring in the EU report observations as a concentration with units of $\mu g\,m^{-3}$. Often, this conversion factor is taken as 0.5 ppbv $(\mu g\,m^{-3})^{-1}$ for convenience.

While handling the ozone mixing ratio data, the data are checked for negative values, which are set to Not a Number (NaN). Some sites have duplicate values for a single time, usually with no explanation for the repetition. In the absence of further explanation, the first value is kept and all subsequent values for a given hour are dropped.

3.5 Calibration

Compared to many atmospheric composition measurements, calibration plays a relatively small role in the measurement of surface ozone because it relies on a direct spectroscopic technique. Therefore comparing measurements from different networks with different calibration routines is of relatively little concern, e.g. compared to the situation for carbon gases (e.g. Dlugokencky et al., 2005). Ozone monitors are calibrated through linear regression of mole fraction measurements of a synthetic stream of ozone in dry air against the US NIST Standard Reference Photometer (SRP) (Galbally and Schultz, 2013). The NIST SRP relies on a fixed absorption cross section defined as $1.1476 \times 10^{-17}\,cm^{-2}\,molecule^{-1}$ at a wavelength of 253.7 nm. WMO GAW recommends that ozone instruments be calibrated every 3 months with a laboratory standard that is traceable to the NIST SRP. In addition to the calibration of the instrument, there is additional calibration of the entire sampling system to account for losses to the sampling line. Mixing ratios may be reported as wet or dry mole fraction, but practically speaking are measured and reported wet to avoid sample losses associated with drying (Galbally and Schultz, 2013). Finally, concentrations may be reported assuming a variety of different fixed conversion factors to mixing ratios as discussed in Sect. 3.4. Unfortunately, none of these details are regularly reported in ozone time series metadata. While calibration information and traceability would be a useful addition to the ozone data sets, we believe it represents a minor source of uncertainty due to the nature of the measurement.

3.6 Combining time series

Once all time series are individually processed, they are merged together on a single timescale. During the process, the data are checked for data overlaps between sequential files for a site. Overlaps may indicate dating or time zone problems. These overlaps are manually inspected. For overlaps of less than a day, one file is prioritized over the other and the overlap is logged and over-written. Discovery of long overlaps (up to a year) has led to improvements in the date parsing routines, so that long overlaps are eliminated in the final version of the data set. Once the time series are merged into a single data set, additional quality controls can be applied. The initial combined data set represents a total of 578 041 289 hourly observations. Table 1 shows how the number of valid observations is reduced throughout the process of data quality processing.

3.7 Outliers and poor data quality

Upon import of each time series, if data quality flags are provided by a data provider, these are used to consistently remove "questionable" or "bad" values, replacing them with not-a-number (NaN) placeholders. Filler values (e.g. −999, 99 999) are also replaced with NaN and data sets are screened for multiple fillers. Similarly, individual values that are extreme outliers (negative values or values greater than 500 ppbv) are removed from the entire data set.

After all the other quality control screening processes described below, there are still various other sites that exhibit either sporadic (1 hour) or extended periods of high ozone values (> 200 ppbv). We interpret these as being due to instrumental failure. We have implemented a process to screen out extended periods of high ozone, although this process is labor-intensive, as it requires visual inspection of the ozone time series. Inspection usually reveals gross problems and the

data are then removed. We do not believe this removes signals from stratospheric intrusion. From the 146 sites above 1000 m elevation, only 2 hours of data are removed because they exceed 200 ppbv.

For each site, a time series of annual mean values are plotted, as well as plotting the site mean vs. the standard deviation of hourly ozone values. Sites with extremely high or low values in the mean or standard deviation are flagged as suspicious and the time series are inspected individually and removed if they exhibit extended periods of anomalously high/low values or periods of constant values, both of which suggest an instrument failure. These sites are listed in Table A2 in Appendix A.

We now discuss issues relating to specific networks.

3.8 EPA AQS coarse resolution data

Some of the EPA AQS sites, notably some in the early years, report data at a resolution of 10 ppbv, while all other data sets report at 1 ppbv or better. These 10 ppbv sites are detected based on the minimum difference between successive measurements and the sites are removed entirely from the data set. However, this does not catch all coarse resolution sites, as some sites have the occasional pair of values with a small difference for an unknown reason. Additional coarse data are removed from the database on a per-site per-year basis using an automated routine to detect "spiky" histograms when the data are binned at 1 ppbv resolution. Up to 299 sites are removed, with the maximum in the year 1990.

All AQS sites use a minimum level of detection, which is generally (92 % of sites) 5 ppbv with all values below this threshold set to 2 ppbv. This impacts the shape of the probability distribution of ozone concentrations and the mean value for AQS sites, but no correction is applied at this point.

3.9 EPA AQS partial years of data

Approximately 400 of the EPA AQS sites only operate for some months each year, as dictated by the EPA "ozone season". This is most commonly April through October as ozone violations of air quality standards predominantly occur during the summer months. The algorithm to find the partial-year data first finds the number of months that have data for each year for each site. If a year has less than 9 months of data and is followed by another year with less than 9 months of data, the first year is flagged as having a partial year of data. We do this to remove the EPA AQS sites that have a short "ozone season" so as not to introduce bias in the seasonal signal for those gridboxes. The check of the subsequent year is intended to avoid removing the start of a many-year time series simply because the time series starts at some time other than January. Data are removed on a year-by-year basis. A whole site is removed from the data set if all years of data are removed. This removes approximately 95 % of the summer-only sites without removing portions of a year from other time series (e.g. the start of a multi-year continuous time series that starts at a time other than 1 January).

3.10 Removing duplicate sites

A small subset of time series represent duplicate records from the same location that have been contributed to different observing networks. For example, Canadian GAW sites are also reported in CAPMON and NAPS data sets. Duplicates are found based on latitude and longitude, with the shorter time series being excluded. In addition, WMO GAW sites at Cape Point (CPT), Ushuia (USH) and Niwot Ridge (NWT) provide two versions of the data, one unfiltered and one filtered for local influences. We use just the filtered version of the data for these sites.

3.11 Removing urban sites

This data set is designed primarily for the evaluation of models with a horizontal resolution of 10 s of kilometers or coarser. These models do not resolve the nonlinearities of urban roadside chemistry. Therefore, we exclude urban monitoring sites from the database. Urban sites are located using two classification schemes. Some contributing data sets provide a land use classification, in which case sites labeled "urban" are excluded. To screen sites from databases that do not provide their own classification scheme, we use land use data from the Anthropogenic Biomes of the World v2 data set (Ellis et al., 2010, 2013) at $\frac{1}{12}^\circ$ (~ 5 km) resolution. If more than 50 % of the grid boxes in a 4×4 cell area ($\frac{1}{3}^\circ \times \frac{1}{3}^\circ$) around the site are classified as urban, a site is excluded. This results in 1026 stations being excluded, mostly along the East Coast of North America and throughout central Europe. By using the classification of multiple grid boxes to determine if a site is urban, we retain sites that may be located on the edge of an urban area or in small developed areas that the land map classifies as urban.

4 Data set spatial and temporal extent

After data processing is complete, there are 209 965 733 valid hourly observations from 2531 sites.

Figure 2 maps the location of the ozone stations by network. The vast majority (97 %) of the sites are located in the northern midlatitudes between 22 and 69° N, primarily in North America and Western Europe. The WMO GAW sites provide a somewhat uniform distribution across the globe, but are not present at a high enough density to provide global coverage. Southern Hemisphere and East Asian continental regions are still underrepresented in the WMO GAW data set. Swiss alpine sites are arguably over-represented among the WMO GAW Global Stations, as it is unlikely that they represent independent observations.

The number of measurements reported in each hour for each ozone network (after quality control) is shown if Fig. 3.

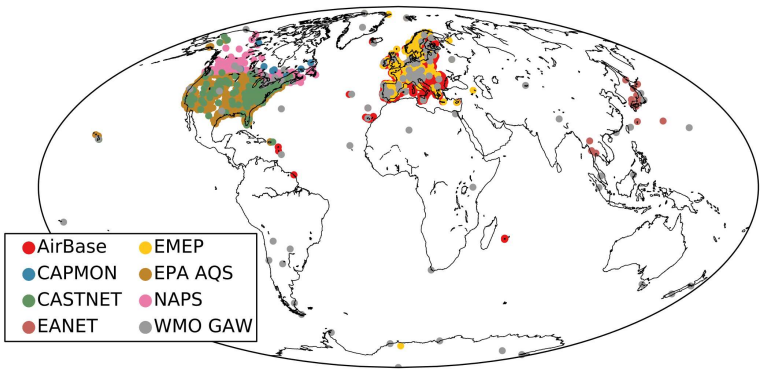

Figure 2. Map of the locations of the publicly available surface ozone observation used to construct the ozone data set. The locations are colored by the observing network.

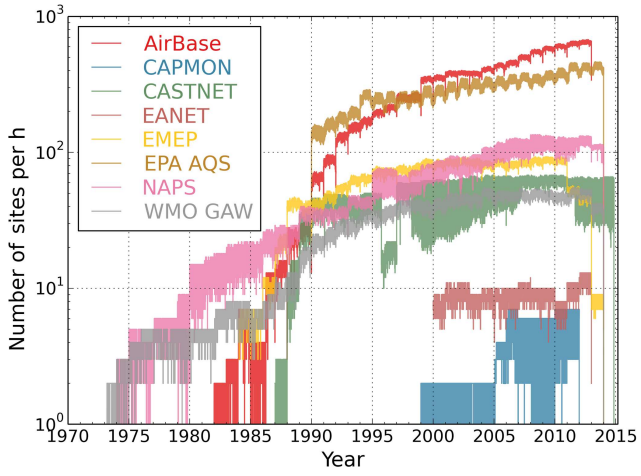

Figure 3. The number of ozone sites in each network reporting data for a given hour, from 1971 to 2015. The count of observations is shown on a logarithmic scale. The color indicates the observing network.

The EU AirBase network represents the largest number of sites contributing to the data set, followed by EPA AQS, EMEP, and NPAS. As the number of measurements in Fig. 3 is shown on an hourly basis, the apparent vertical width of each curve suggests the data quality with regard to missing data for each data set. The WMO GAW data set has notably more continuous ozone time series than the EPA CASTNET data set, based on the black line being thinner than the green line, even though the maximum number of sites reporting at a given time is comparable from 1998–2012. The periodic structure in the EPA AQS data count reflects the remaining 5 % of summertime-only sites that were not screened out by the process described in Sect. 3.9.

Representativeness

These ozone observations would be most useful if they provided, in some sense, a global assessment of the surface ozone concentration. It is impossible to measure the ozone everywhere at infinite resolution but a "reasonable" representation of the surface coverage would be useful. Here we assess the representativeness of the data sets in terms of a real coverage. To do so, we use the surface ozone field from a global model to ascertain the "footprint" of each site and the representativeness of the entire data sets in terms of areal coverage. We deseasonalize monthly surface ozone model output from a 7 year $2° \times 2.5°$ resolution simulation by the GEOS-Chem chemical transport model (version 9-01-03; www.geos-chem.org) (Parrella et al., 2012; Bey et al., 2001). For each surface grid box in the model, we calculate the correlation coefficient between its deseasonalized monthly mean time series and that of every other grid box in turn. Thus, for each model grid box we derive a global one-point correlation map. Grid boxes which show a similar variation in deseasonalized monthly means have a high value and those which do not have a low value. The footprint that is representative of an observation in a grid box is determined by finding those grid boxes with a correlation greater than a fixed value R that are contiguous with the observation site grid box as determined by a "downhill" random walk process. Figure 4 shows the footprint derived for the Cape Verde site for R values ranging from 0.1 to 0.9. We choose a correlation threshold $R = 1/\sqrt{2} = 0.707$ such that $R^2 = 0.5$, meaning that the time series in the observation grid box explains 50 % or more of the non-seasonal variance in each of the other grid boxes in the footprint. The shape of each footprint is typically similar both up- and down-wind. The footprints do not represent back-trajectories, but the areas of similar ozone variation both backward and forward from the observation site. There will be some difference in the footprint between different model simulation and between different models but we do not believe that the general conclusions will be significantly different.

The composite of all footprints with $R \geq 0.707$ provides a mask of the globe that represents the area that is observed by one or more site (Fig. 5). The US, Canada, and Europe

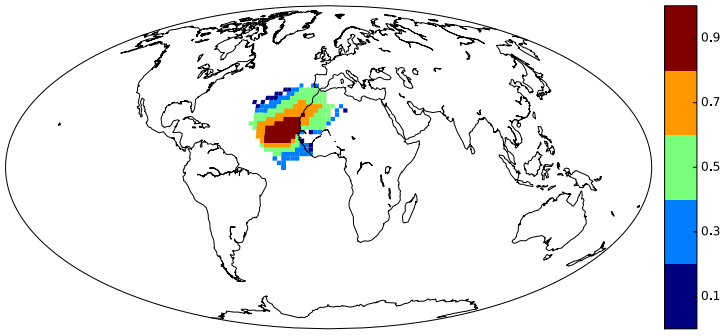

Figure 4. Examples of the ozone site footprint for the Cape Verde Observatory, drawn with thresholds of $R = 0.1$ to 0.9.

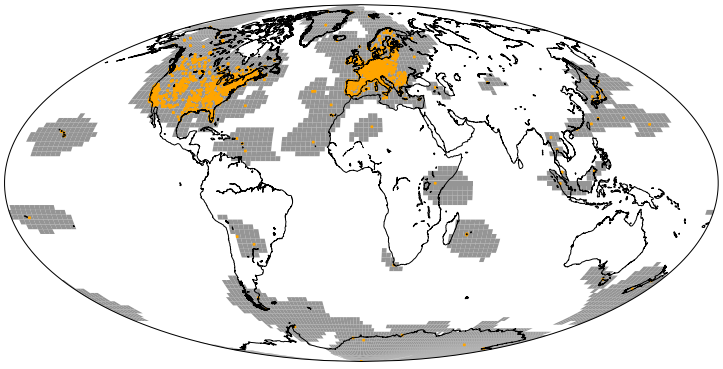

Figure 5. Map of surface ozone observational coverage based on the composite of the footprints with threshold $R>=0.707$ for existing background ozone observing sites. Orange markers indicate site locations.

are fully covered, while large portions of South and Central America, Africa, and Asia are not covered by available ozone observations. In total, 25 % of the Earth's surface area is covered. Due to the spatial scale of the model and strong trends in Asian ozone precursor emissions, the footprints of ozone observations from Japan and Taiwan likely overestimate their ability to capture ozone over China. We find that both tropical regions and the Southern Hemisphere are poorly covered by our current observing capability, meaning that there is not direct observational information about the potential damage to human health and ecosystems in these regions.

5 Statistical overview of observations

With the volume of data available in this database, there are many different statistical aspects that could be explored. Here, we briefly present a few examples. In Fig. 6, we plot the spatial distribution of the mean, standard deviation, skewness, and kurtosis of each time series in the combined data set. The highest mean values are seen across the US "Mountain West", Southern Europe, and Japan. The greatest standard deviation in hourly ozone is seen in the Southeastern US and Southeast Asia. Most sites in the data set exhibit a positive skewness, consistent with ozone concentrations often approximating a log-normal distribution (Denby et al., 2010). The ozone data set exhibits sites with both positive and negative kurtosis, with more remote sites generally exhibiting negative kurtosis (i.e. broader peak and thinner tails), while polluted sites exhibit more positive kurtosis (i.e. narrow peak and fatter tails), consistent with the chemistry of polluted regions leading to more extreme ozone mixing ratios.

6 Gridding data for model evaluation

For the evaluation of ozone in chemical transport models, either model output can be interpolated to the location of the observations, or observations can be averaged to a grid at the resolution of the model. For the sake of sharing data with modeling groups for model evaluation, we have opted to take the latter approach. This approach allows for the redistribution of a derived product without redistributing the raw data, separating it from its original archives and compromising the data ownership rights of the original data contributors. Data are averaged onto a global grid of $1^\circ \times 1^\circ$ or coarser resolution to match the model grids. More complex approaches have been taken (Schnell et al., 2014), but for the sake of simplicity we have opted to calculate the arithmetic mean of the mixing ratios of the representative sites in each gridbox. The data that are incorporated into version 2.7 of the gridded product are restricted to sites below 1500 m altitude

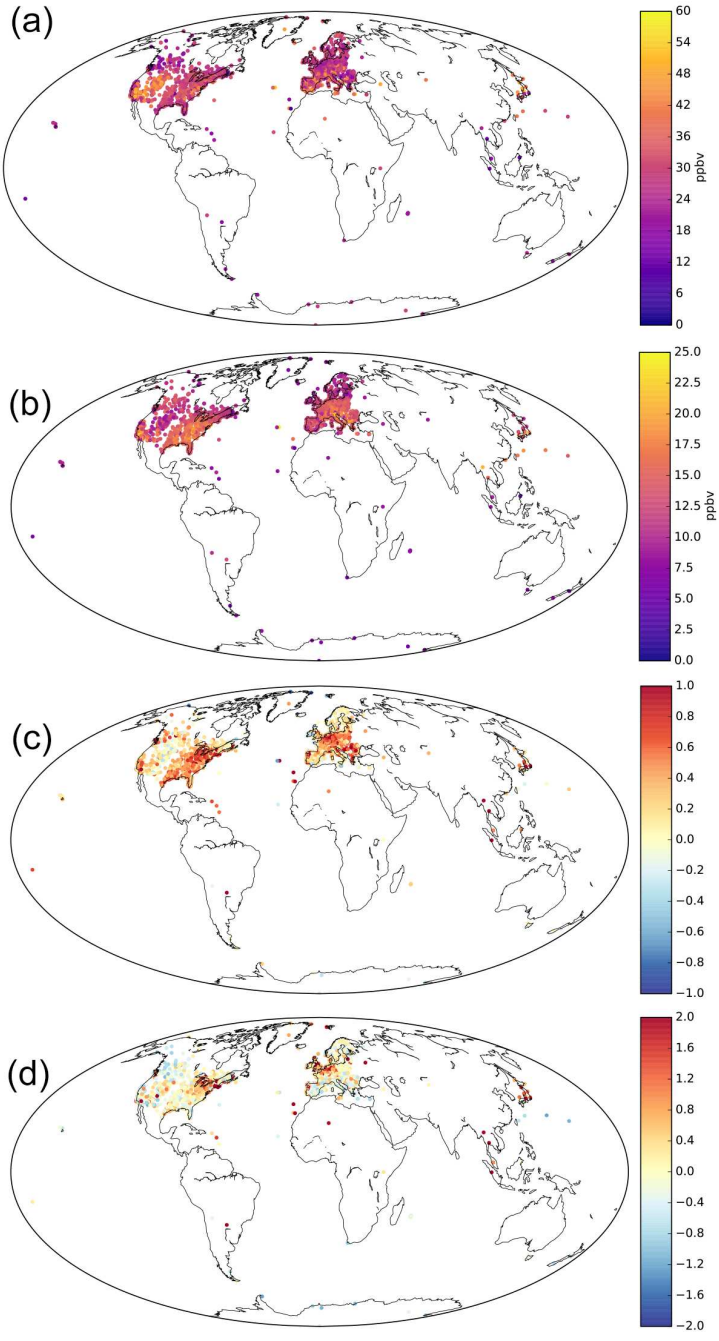

Figure 6. Map of long-term **(a)** mean, **(b)** standard deviation, **(c)** skewness, and **(d)** kurtosis of surface ozone observations.

in order to exclude high mountain-top sites that may sample the free troposphere much of the time. These gridded data are then used to calculate metrics at various timescales that are intended to describe the distribution of ozone concentrations for a grid box and aspects of that distribution that are relevant for air quality policy. Tables 2 and 3 describe the complete suite of resolutions, timescales, and metrics that are available. Time averaging is done with respect to UTC times for consistency with typical model output, with the exception of the AOT40 plant exposure metric. To calculate AOT40, a longitude-based local time is used to determine hours of daylight when plants are susceptible to ozone damage. We calculate AOT40 following EMEP guidelines appropriate for both crops (integrating over 1 May–31 July for Northern Hemisphere extratropics) and forests (integrating over 1 April–30 September for Northern Hemisphere extratropics) (Gauss et al., 2014). Gridded data at 1° resolution, as well as other common model resolutions (e.g. 2° × 2.5°) are

Table 2. Spatial and temporal resolutions available for model evaluation via the BADC.

Resolution	Notes/Model	Bottom left grid edge (°)*	
1° × 1°	centered on degrees	−180.5	−90.5
1° × 1°	centered on 0.5°	−180	−90
2° × 2.5°	GEOS-Chem (Bey et al., 2001)	−181.25	−91
4° × 5°	GEOS-Chem	−182.5	−92
4° × 5°	GEOS-Chem/GCAP (Wu et al., 2008)	−182.5	−90
2° × 2.5°	GISS ModelE (Miller et al., 2014)	−178.75	−90
2.5° × 3.75°	UKCA (O'Connor et al., 2014)	−180	−90
2° × 2°	ACCMIP (Lamarque et al., 2013) common resolution	−180	−90

Timescale	Notes
Monthly	Calendar months; February has 28 or 29 days.
Annual	Calendar year; leap years are 366 days.

*Latitudes less than −90° indicate half-boxes at poles.

made publicly available in netCDF-4 via the British Atmospheric Data Centre (BADC) (doi:10.5285/08fbe63d-fa6d-4a7a-b952-5932e3ab0452) (Evans and Sofen, 2015a, b). The gridded data are freely available, but registration with the Centre for Environmental Data Archival (CEDA) via the "Request Access" link on the data set home page is required to access the download page. Registration requires basic contact and institutional information. Two versions of the data set are now available. The first version (2.4) includes sites from all elevations, while the second version includes two updates. The second version (2.7) includes a restriction to only include sites below 1500 m and includes EPA AQS sites for the years 1980–1989 that were not included in version 2.4.

We provide two error statistics to represent the uncertainty in the calculated gridded mean values. The first is the standard deviation over the hourly data going into the time average ($\sigma_{\bar{x}}$) and represents the temporal variability within the time average (e.g. monthly mean).

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{N-1}\sum_{t}^{N}(\overline{c_x}(t) - \bar{c})^2}, \quad (2)$$

where $c(x,t)$ is the concentration for site x of M sites in the grid box at time t for N times within a time interval (e.g. month), $\overline{c_x}(t)$ is the grid box mean time series and represents the average over the M sites for a given hour t, and $\bar{c}$ represents the grid box mean value over all M sites and N times.

The second error statistic is the mean standard deviation between sites within a grid box over a timestep ($\overline{\sigma_x}$) and indicates the representativeness of the observations contributing to the grid box mean:

$$\overline{\sigma_x} = \frac{1}{N}\sum_{t}^{N}\sqrt{\frac{1}{M-1}\sum_{x}^{M}(c(x,t) - \overline{c_x}(t))^2}. \quad (3)$$

A high value of $\overline{\sigma_x}$ indicates a large degree of variability between the sites going into a grid box mean and would suggest that model-measurement disagreement in this grid box is likely not due to model failure but due to sub-grid scale variability in the observations. The $\overline{\sigma_x}$ is not defined if there is only a single site in the gridbox. A high value of $\sigma_{\bar{x}}$ indicates a large degree of temporal variability within the gridbox-mean time series.

In addition, we provide the number of sites and the fraction of the total possible hours of data contributing to each grid box for each time interval. Finally, we provide the average, minimum, and maximum elevation for site elevation that goes into each grid box, so that models may be sampled at a consistent altitude. For grid boxes that contain only a single site, the average, minimum, and maximum are the same. Only 1608 of the 2389 sites include elevation information. The remaining 781 sites do not contain altitude information and are simply not included in the grid box altitude calculation.

Each netCDF file also includes a complete collection of the site-level metadata for the sites that go into the gridded data set. This includes attributes such as the site name, latitude, longitude, data provider, contact, and additional site description details.

As a simple demonstration of the application of the gridded observations for model evaluation, we compare a number of statistical measures of the ozone data with hourly output from our GEOS-Chem model simulation at 2° × 2.5° resolution. This comparison is shown in Fig. 7. We calculate moments of the ozone distribution (mean, standard devia-

Table 3. Metrics available for model evaluation via the BADC.

Variable Name	Metric and Notes/Definition
`Mean_Gridded_Ozone`	**Mean** $\overline{c} = \sum_t^N \overline{c_x}(t)$; for N hours in averaging period
`Median_Gridded_Ozone`	**Median**
`Std_Dev_Gridded_Ozone`	**Standard deviation in gridded hourly data** $\sigma_{\overline{x}}$ see Eq. (2)
`Skewness_Gridded_Ozone`	**Skewness** pearson skewness
`Kurtosis_Gridded_Ozone`	**Kurtosis** zero-centered
e.g. `25Percentile_Gridded_Ozone`	**25th, 75th, 95th, and 99th percentiles** of gridded hourly ozone.
`MeanMDA8_Gridded_Ozone`	**Maximum daily 8 h average (MDA8)** e.g. monthly mean of daily MDA8. Calculated for 24 periods for each day in local time.
`AOT40crop_Gridded_Ozone`	**AOT40-crop** fixed May–July growing season and 08:00–20:00 h of day (longitudinal local time). Calculated annually only.
`AOT40forest_Gridded_Ozone`	**AOT40-forest** fixed April–September growing season and 08:00–20:00 h of day (longitudinal local time). Calculated annually only.
`DaysMDA8_ge_60ppb_Gridded_Ozone`*	**Number of days with MDA8 > 60 ppbv** relevant to EU air quality standard.
`FourthHighestMDA8_Gridded_Ozone`*	**4th highest MDA8** relevant to US EPA air quality standard.
Auxiliary Metrics	
`Std_Dev_Sites`	**Standard deviation between sites in a gridbox** $\overline{\sigma_x}$ see Eq. (3)
`count_Nsites`	**Sites per grid box per timestep**
`DataFrac`	**Data completeness fraction**

* Although these metrics are useful for air quality policy purposes and regional modeling, they may be of less value for global model evaluation and will require further testing.

tion, skewness and kurtosis), median, and 25th and 75th percentiles from hourly ozone data on an annual basis (e.g. one point per grid box per year) for the years 2005–2012. While there is substantial scatter, the GEOS-Chem model captures the annual mean with little bias. In contrast, GEOS-Chem shows a systematic bias in the annual standard deviation of hourly ozone, with an exaggerated pattern of overestimating high variability and underestimating low variability. The model shows very little skill at capturing variability in the skewness or kurtosis, although there are no substantial systematic biases. The model does a better job at capturing the spatiotemporal variability in the 75th percentile than the 25th percentile. The higher order moments and the low tail of the distribution (e.g. 25th percentile) are not typically considered in model evaluation. In contrast, the mean and high tail of the distribution (e.g. 75th, 90th, or 95th percentile) are commonly evaluated for ozone in atmospheric chemistry models because of their air quality and climate relevance. These figures demonstrate that it may be worth paying additional attention to them, particularly when models are being evaluated for purposes other than predicting mean values.

7 Recommendations for experimentalists and data managers

The process of aggregating and homogenizing the surface ozone data sets has revealed a number of issues with how historical ozone data and metadata archived. These issues dramatically increase the effort involved in working with multiple surface ozone data sets, increase the uncertainty in ozone concentrations, and in some cases make data unusable. In light of this, we provide several recommendations to the ozone measurement community stemming from the quality control processes described in Sect. 3.

Data processing could be greatly simplified if all data providers report hourly mean ozone values indexed to the UTC time zone and used the ISO-8601 standards for formatting data and time information.

A complete set of metadata for an ozone time series should include sufficient information to convert from a number or mass density ($\mu g\,m^{-3}$) to mixing ratio (ppbv). Metadata should also include information on calibration traceability, the absorption cross-section, and whether mole fractions are reported with respect to wet or dry air. These recommendations should significantly reduce uncertainty when comparing ozone observations from different networks. Furthermore, a recent re-measurement of the ozone absorption

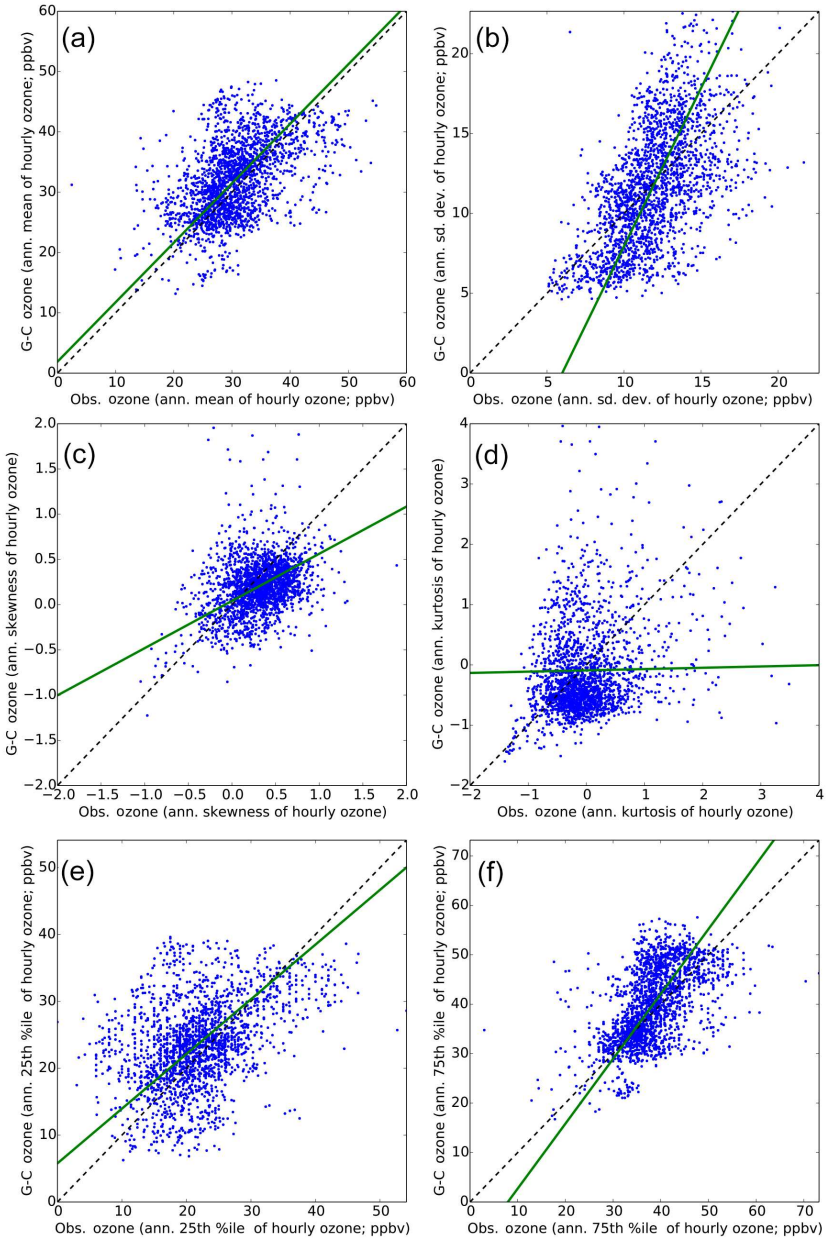

Figure 7. Comparison of annual **(a)** mean, **(b)** standard deviation, **(c)** skewness, **(d)** kurtosis, and **(e)** 25th and **(f)** 75th percentiles of hourly surface ozone between observations (x axis) and the GEOS-Chem global chemical transport model (y axis) for the years 2005–2012. The green best-fit line is the orthogonal least-squares regression.

cross-section has resulted in a 1.8 % decrease in value (Viallon et al., 2015). Should this new cross-section be adopted by the International Union of Pure and Applied Chemistry and be implemented in new ultraviolet absorption spectrophotometers, this update will lead to a 1.8 % increase in the reported ozone values. While a small change that will be difficult to detect in long-term trends, it has significant implications for air quality compliance (Sofen et al., 2015). Metadata structures for many of the networks need to be revised so that the timing of this cross-section update can be recorded.

The EPA AQS and EU AirBase data sets provide metadata that describes the surrounding environment (e.g. urban, rural), but their classifications are not entirely consistent and other data sets do not provide this information. Consistent definitions of the surrounding environment can greatly improve the screening of the data sets for background or regionally representative sites that are appropriate for model evaluation. A valuable contribution from global coordination efforts such as those from the WMO GAW or the ongoing Tropospheric Ozone Assessment Report would be the

development and implementation of a consistent classification scheme based on either site surveys or land cover maps across surface ozone data sets.

8 Conclusions

We have constructed a globally consistent data set of hourly surface ozone observations intended primarily for the evaluation of surface ozone in global atmospheric chemistry models. The vast majority of the observations are made in the Northern Hemisphere extratropics coming from US, Canadian, and EU air quality and regional background networks. The data set is quality controlled to correct time zones and units and to remove outliers, EPA AQS sites that only provide data for a part of each year or at low measurement resolution, and urban sites that are not representative of the large-scale background chemistry that global models simulate. The data are made available to modelers in gridded format with many different metrics. The metrics are intended to provide a description of the probability distribution of ozone concentrations (e.g. mean, median, standard deviation, percentiles), as well as aspects of that distribution that are relevant for air quality policy (e.g. mean maximum daily average 8 h ozone (MDA8), the fourth highest MDA8, and number of days with MDA8 greater than 60 ppbv), or human and ecosystem health (e.g. SOMO35 and AOT40). In addition, we provide auxiliary metrics that are useful for assessing the model-measurement comparison, such as the standard deviation between the sites going into a single grid box.

As new data sets become publicly accessible from regions such as China, this composite data set and gridded data products can be re-visited and expanded. As future expansion of the ozone observing network is planned, consideration should be given to improving the global coverage.

Appendix A

Table A1 records time zone anomalies found in metadata from the WMO GAW data set. Table A2 provides details on other sites with anomalies in their data that were not detected by the primary data quality control routines.

Table A1. Sites with time zone inconsistencies.

Site name	Site ID	TZ given	TZ actual
Issyk-Kul, Kyrgyzstan	isk	UTC + 5	UTC + 6
Cape Grim, Australia	cgo	none (some years)	UTC + 10
Payerne, Switzerland	pay	none (some years)*	UTC + 1
Rucava, Latvia	rcv	none (some years)*	UTC + 2
Jungfraujoch, Switzerland	jfj	none (some years)*	UTC + 1
Rigi, Switzerland	rig	none (some years)*	UTC + 1
K-puszta, Hungary	kps	none (some years)	UTC + 1
Monte Cimone, Italy	cmn	none (some years)*	UTC + 1
Giordan Lighthouse, Malta	glh	none (some years)	UTC + 1, may be UTC based on comparison to EMEP MT0001R.
Eskdalemuir, UK; Vindeln, Sweden; Zeppelinfjellet, Norway; Kosetice, Czech Republic; Hok Tsui, Hong Kong; Puszcza Borecka/Diabla Gora, Poland; Danki, Russia; Shepelevo, Russia; Ähtäri, Finland; Mahon, Spain; Noia, Spain; Oulanka, Finland; Roquetes, Spain; San Pablo de los Montes, Spain Uto, Finland; Virolahti, Finland; Fundata, Romania; Doñana, Spain	edm, vdl, zep, kos, hkg, dig, dak, shp, aht, mhn, nia, oul, roq, spm, uto, vir, fdt, don	none (prior to 2007)	Various (time zone reporting begins in 2007).

* Now fixed on WDCGG database.

Table A2. Sites with data anomalies.

Site Name	Site ID	Note
Funchal, Portugal	fun	Spikes and baseline shifts.
Angra de Heroismo, Portugal	ang	150 ppb annual mean in 1995.
Lazaropole, Macedonia	MK0007R	Baseline shift in 2010.
Topolniky, Slovakia	SK0007R	Repeated biweekly pattern imposed on the data for 2004.
Ventura, CA	AQS061110004	Abnormal nighttime behavior.
Kosetice, Czech Republic	kos	Multiple values per timestamp.

Author contributions. In addition to the coauthors named above, we acknowledge the authorship of all other contributors to WMO GAW, EPA AQS, EPA CASTNET, CAPMON, NAPS, AirBase, EMEP, EANET ozone data sets.

Acknowledgements. We thank the World Data Centre for Greenhouse Gases WDCGG for providing WMO GAW surface ozone data and their staff for their responsiveness in discussing and addressing metadata issues. We thank Barron Henderson (U. Florida) for making available his bpch and pseudonetcdf code for processing GEOS-Chem model output in Python. We acknowledge funding from UK NERC grants NE/K016008/1, NE/L01291X/1. Thanks to the British Atmospheric Data Centre, which is part of the NERC National Centre for Atmospheric Science (NCAS), for their assistance in archiving the gridded ozone data set. We note that named coauthorship has been offered to all data contributors for whom contact information was available, consistent with the WMO GAW data use guidelines.

Edited by: V. Sinha

References

AMEC Environment and Infrastructure, Inc.: Clean Air Status and Trends Network (CASTNET) 2012 Annual Report, Tech. rep., Prepared for U.S. Environmental Protection Agency (EPA), Washington, DC, contract No. EP-W-09-028, 2014.

Amman, M., Bertok, I., Cofala, J., Gyarfas, F., Heyes, C., Klimont, Z., Schöpp, W., and Winiwarter, W.: Baseline Scenarios for the Clean Air for Europe (CAFE) Programme: Final Report, Tech. rep., European Commission Directorate General for Environment, available at: http://ec.europa.eu/environment/archives/cafe/activities/pdf/cafe_scenario_report_1.pdf (last access: 29 July 2015), 2005.

Appel, K. W., Gilliland, A. B., Sarwar, G., and Gilliam, R. C.: Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: sensitivities impacting model performance: Part I – Ozone, Atmos. Environ., 41, 9603–9615, doi:10.1016/j.atmosenv.2007.08.044, 2007.

Ashmore, M. and Wilson, R. (Eds.): Critical Levels for Air Pollutants for Europe (UN-ECE Workshop Report), Department of the Environment, London, UK, 1994.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation, J. Geophys. Res., 106, 23073, doi:10.1029/2001JD000807, 2001.

Brown-Steiner, B., Hess, P., and Lin, M.: On the capabilities and limitations of GCCM simulations of summertime regional air quality: a diagnostic analysis of ozone and temperature simulations in the US using CESM CAM-Chem, Atmos. Environ., 101, 134–148, doi:10.1016/j.atmosenv.2014.11.001, 2015.

Chandra, S., Ziemke, J. R., and Martin, R. V.: Tropospheric ozone at tropical and middle latitudes derived from TOMS/MLS residual: comparison with a global model, J. Geophys. Res., 108, 4291, doi:10.1029/2002JD002912, 2003.

Colette, A., Granier, C., Hodnebrog, Ø., Jakobs, H., Maurizi, A., Nyiri, A., Rao, S., Amann, M., Bessagnet, B., D'Angiola, A., Gauss, M., Heyes, C., Klimont, Z., Meleux, F., Memmesheimer, M., Mieville, A., Rouïl, L., Russo, F., Schucht, S., Simpson, D., Stordal, F., Tampieri, F., and Vrac, M.: Future air quality in Europe: a multi-model assessment of projected exposure to ozone, Atmos. Chem. Phys., 12, 10613–10630, doi:10.5194/acp-12-10613-2012, 2012.

Denby, B., Sundvor, I., Cassiani, M., de Smet, P., de Leeuw, F., and Horálek, J.: Spatial mapping of ozone and SO_2 trends in Europe, Sci. Total Environ., 408, 4795–4806, doi:10.1016/j.scitotenv.2010.06.021, 2010.

Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W., and Steele, L. P.: Conversion of NOAA atmospheric dry air CH_4 mole fractions to a gravimetrically prepared standard scale, J. Geophys. Res., 110, D18306, doi:10.1029/2005JD006035, 2005.

Ellis, E. C., Goldewijk, K. K., Siebert, S., Lightman, D., and Ramankutty, N.: Anthropogenic transformation of the biomes, 1700 to 2000, Global Ecol. Biogeogr., 19, 589–606, doi:10.1111/j.1466-8238.2010.00540.x, 2010.

Ellis, E. C., Goldewijk, K. K., Siebert, S., Lightman, D., and Ramankutty, N.: Anthropogenic Biomes of the World, Version 2: 2000, doi:10.7927/H4D798B9, 2013.

EMEP Steering Body: Revised Strategy for EMEP for 2010–2019, Tech. rep., Economic Commission for Europe: Executive body for the Convention on Long-Range Transboundary Air Pollution, available at: http://www.unece.org/fileadmin/DAM/env/documents/2013/air/emep/Informal_document_no_20_Revised_Strategy_for_EMEP_for_2010-2019_clean_text.pdf (last access: 29 July 2015), 2012.

Emery, C., Jung, J., Downey, N., Johnson, J., Jimenez, M., Yarwood, G., and Morris, R.: Regional and global modeling estimates of policy relevant background ozone over the United States, Atmos. Environ., 47, 206–217, doi:10.1016/j.atmosenv.2011.11.012, 2012.

European Environment Agency: Directive 2002/3/EC of the European Parliament and of the Council of 12 February 2002 relating to ozone in ambient air, Official Journal of the European Communities, L67/14, L067, 0014–0030, 2002.

European Topic Centre on Air Pollution and Climate Change Mitigation: AirBase – the European air quality database, available at: http://www.eea.europa.eu/data-and-maps/data/ds_resolveuid/3c756b2021754f6bba40447397d67fdf (last updated: 14 March 2014), 2015.

Evans, M. J. and Sofen, E. D.: Gridded Global Surface Ozone Metrics data (1971–2015) for Atmospheric Chemistry Model Evaluation – version 2.4, doi:10.5285/08fbe63d-fa6d-4a7a-b952-5932e3ab0452, 2015a.

Evans, M. J. and Sofen, E. D.: Gridded Global Surface Ozone Metrics data (1971–2015) for Atmospheric Chemistry Model Evaluation – version 2.7, doi:10.5285/897a3958-5bfc-4311-9bb0-01134bf6aefa, 2015b.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Dun-

can, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S., and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res., 114, D04301, doi:10.1029/2008JD010816, 2009.

Galbally, I. E. and Schultz, M. G.: Guidelines for Continuous Measurements of Ozone in the Troposphere, Tech. Rep. 209, World Meteorological Organization Global Atmospheric Watch, Geneva, 2013.

Gauss, M., Semeena, V. S., Benedictow, A., and Klein, H.: Transboundary air pollution by main pollutants (S, N, O_3) and PM in 2012, Tech. Rep. MSC-W Data Note 1/2014, European Union, available at: http://emep.int/publ/reports/2014/Country_Reports/report_EU.pdf (last access: 29 July 2015), 2014.

Jacob, D. J. and Winner, D. A.: Effect of climate change on air quality, Atmos. Environ., 43, 51–63, doi:10.1016/j.atmosenv.2008.09.051, 2009.

Katragkou, E., Zanis, P., Tsikerdekis, A., Kapsomenakis, J., Melas, D., Eskes, H., Flemming, J., Huijnen, V., Inness, A., Schultz, M. G., Stein, O., and Zerefos, C. S.: Evaluation of near-surface ozone over Europe from the MACC reanalysis, Geosci. Model Dev., 8, 2299–2314, doi:10.5194/gmd-8-2299-2015, 2015.

Kim, H. C., Lee, P., Ngan, F., Tang, Y., Yoo, H. L., and Pan, L.: Evaluation of modeled surface ozone biases as a function of cloud cover fraction, Geosci. Model Dev., 8, 2959–2965, doi:10.5194/gmd-8-2959-2015, 2015.

Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., and Zeng, G.: The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics, Geosci. Model Dev., 6, 179–206, doi:10.5194/gmd-6-179-2013, 2013.

Lapina, K., Henze, D. K., Milford, J. B., Huang, M., Lin, M., Fiore, A. M., Carmichael, G., Pfister, G. G., and Bowman, K.: Assessment of source contributions to seasonal vegetative exposure to ozone in the U.S., J. Geophys. Res., 119, 324–340, doi:10.1002/2013JD020905, 2014.

Lefohn, A. S. and Runeckles, V. C.: Establishing standards to protect vegetation – ozone exposure/dose considerations, Atmos. Environ., 21, 561–568, doi:10.1016/0004-6981(87)90038-2, 1987.

Liu, G., Liu, J., Tarasick, D. W., Fioletov, V. E., Jin, J. J., Moeini, O., Liu, X., Sioris, C. E., and Osman, M.: A global tropospheric ozone climatology from trajectory-mapped ozone soundings, Atmos. Chem. Phys., 13, 10659–10675, doi:10.5194/acp-13-10659-2013, 2013.

Logan, J. A.: An analysis of ozonesonde data for the troposphere: recommendations for testing 3-D models and development of a gridded climatology for tropospheric ozone, J. Geophys. Res., 104, 16115–16149, doi:10.1029/1998JD100096, 1999.

McKinney, W.: Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, edited by: van der Walt, S. and Millman, J., SciPy, Austin, Texas, 51–56, 2010.

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., DelGenio, A. D., Kelley, M., Lo, K. K., Ruedy, R., Shindell, D. T., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Menon, S., Oinas, V., Pérez García-Pando, C., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Sun, S., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: CMIP5 historical simulations (1850–2012) with GISS ModelE2, J. Adv. Model. Earth Syst., 6, 441–477, doi:10.1002/2013MS000266, 2014.

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G. E., Stevenson, D. S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., and Williams, M. L.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmos. Chem. Phys., 15, 8889–8973, doi:10.5194/acp-15-8889-2015, 2015.

Müller, G., Artz, R., Baltensperger, U., Carmichael, G., Dlugokencky, E., Penkett, S., Stähelin, J., Webb, A., Hov, Ø., Klausen, J., Sturges, B., Barrie, L., Braathen, G., Jalkanen, L., and Nickovic, S.: WMO Global Atmosphere Watch (GAW) Strategic Plan: 2008–2015, Tech. rep., World Meteorological Organization, available at: ftp://ftp.wmo.int/Documents/PublicWeb/arep/gaw/gaw172-26sept07.pdf (last access: 29 July 2015), 2007.

O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., Folberth, G. A., Sanderson, M. G., Telford, P. J., Voulgarakis, A., Young, P. J., Zeng, G., Collins, W. J., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 2: The Troposphere, Geosci. Model Dev., 7, 41–91, doi:10.5194/gmd-7-41-2014, 2014.

Parrella, J. P., Jacob, D. J., Liang, Q., Zhang, Y., Mickley, L. J., Miller, B., Evans, M. J., Yang, X., Pyle, J. A., Theys, N., and Van Roozendael, M.: Tropospheric bromine chemistry: implications for present and pre-industrial ozone and mercury, Atmos. Chem. Phys., 12, 6723–6740, doi:10.5194/acp-12-6723-2012, 2012.

Parrington, M., Palmer, P. I., Lewis, A. C., Lee, J. D., Rickard, A. R., Di Carlo, P., Taylor, J. W., Hopkins, J. R., Punjabi, S., Oram, D. E., Forster, G., Aruffo, E., Moller, S. J., Bauguitte, S. J.-B., Allan, J. D., Coe, H., and Leigh, R. J.: Ozone photochemistry in boreal biomass burning plumes, Atmos. Chem. Phys., 13, 7321–7341, doi:10.5194/acp-13-7321-2013, 2013.

Pfister, G. G., Walters, S., Lamarque, J.-F., Fast, J., Barth, M. C., Wong, J., Done, J., Holland, G., and Bruyère, C. L.: Projections of future summertime ozone over the U.S., J. Geophys. Res., 119, 5559–5582, doi:10.1002/2013jd020932, 2014.

Reidmiller, D. R., Fiore, A. M., Jaffe, D. A., Bergmann, D., Cuvelier, C., Dentener, F. J., Duncan, B. N., Folberth, G., Gauss, M., Gong, S., Hess, P., Jonson, J. E., Keating, T., Lupu, A., Marmer, E., Park, R., Schultz, M. G., Shindell, D. T., Szopa, S., Vivanco, M. G., Wild, O., and Zuber, A.: The influence of foreign vs. North American emissions on surface ozone in the US, Atmos. Chem. Phys., 9, 5027–5042, doi:10.5194/acp-9-5027-2009, 2009.

Schnell, J. L., Holmes, C. D., Jangam, A., and Prather, M. J.: Skill in forecasting extreme ozone pollution episodes with a global atmospheric chemistry model, Atmos. Chem. Phys., 14, 7721–7739, doi:10.5194/acp-14-7721-2014, 2014.

Shim, C., Wang, Y., and Yoshida, Y.: Evaluation of model-simulated source contributions to tropospheric ozone with aircraft observations in the factor-projected space, Atmos. Chem. Phys., 8, 1751–1761, doi:10.5194/acp-8-1751-2008, 2008.

Simon, H., Reff, A., Wells, B., Xing, J., and Frank, N.: Ozone trends across the United States over a period of decreasing NO_x and VOC emissions, Environ. Sci. Technol., 49, 186–195, doi:10.1021/es504514z, 2015.

Sofen, E. D., Evans, M. J., and Lewis, A. C.: Updated ozone absorption cross section will reduce air quality compliance, Atmos. Chem. Phys., 15, 13627–13632, doi:10.5194/acp-15-13627-2015, 2015.

Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Denier van der Gon, H., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Jeričević, A., Kraljević, L., Miranda, A. I., Nopmongcol, U., Pirovano, G., Prank, M., Riccio, A., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., and Galmarini, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII, Atmos. Environ., 53, 60–74, doi:10.1016/j.atmosenv.2012.01.003, 2012.

Thermo Fisher Scientific Inc.: Model 49i Instruction Manual, Franklin, MA, 2011.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Conley, A., Schultz, M. G., Saunois, M., Thouret, V., Thompson, A. M., Oltmans, S. J., Johnson, B., and Tarasick, D.: Technical Note: Ozonesonde climatology between 1995 and 2011: description, evaluation and applications, Atmos. Chem. Phys., 12, 7475–7497, doi:10.5194/acp-12-7475-2012, 2012.

Tørseth, K., Aas, W., Breivik, K., Fjæraa, A. M., Fiebig, M., Hjellbrekke, A. G., Lund Myhre, C., Solberg, S., and Yttri, K. E.: Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972–2009, Atmos. Chem. Phys., 12, 5447–5481, doi:10.5194/acp-12-5447-2012, 2012.

Viallon, J., Lee, S., Moussay, P., Tworek, K., Petersen, M., and Wielgosz, R. I.: Accurate measurements of ozone absorption cross-sections in the Hartley band, Atmos. Meas. Tech., 8, 1245–1257, doi:10.5194/amt-8-1245-2015, 2015.

Voulgarakis, A., Savage, N. H., Wild, O., Carver, G. D., Clemitshaw, K. C., and Pyle, J. A.: Upgrading photolysis in the p-TOMCAT CTM: model evaluation and assessment of the role of clouds, Geosci. Model Dev., 2, 59–72, doi:10.5194/gmd-2-59-2009, 2009.

Walt, S. v. d., Colbert, S. C., and Varoquaux, G.: The NumPy array: a structure for efficient numerical computation, Comput. Sci. Eng., 13, 22–30, doi:10.1109/MCSE.2011.37, 2011.

Wang, Y., Jacob, D. J., and Logan, J. A.: Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry: 3. Origin of tropospheric ozone and effects of nonmethane hydrocarbons, J. Geophys. Res., 103, 10757–10767, doi:10.1029/98JD00156, 1998.

Wu, S., Mickley, L. J., Jacob, D. J., Rind, D., and Streets, D. G.: Effects of 2000–2050 changes in climate and emissions on global tropospheric ozone and the policy-relevant background surface ozone in the United States, J. Geophys. Res., 113, D18312, doi:10.1029/2007JD009639, 2008.

Zhang, L., Jacob, D. J., Liu, X., Logan, J. A., Chance, K., Eldering, A., and Bojkov, B. R.: Intercomparison methods for satellite measurements of atmospheric composition: application to tropospheric ozone from TES and OMI, Atmos. Chem. Phys., 10, 4725–4739, doi:10.5194/acp-10-4725-2010, 2010.

Ziemke, J. R., Chandra, S., Duncan, B. N., Froidevaux, L., Bhartia, P. K., Levelt, P. F., and Waters, J. W.: Tropospheric ozone determined from Aura OMI and MLS: evaluation of measurements and comparison with the Global Modeling Initiative's Chemical Transport Model, J. Geophys. Res., 111, D19303, doi:10.1029/2006JD007089, 2006.