Supplemental Material

Tropospheric Ozone Assessment Report: Tropospheric ozone from 1877 to 2016, observed levels, trends and uncertainties.

David Tarasick[1*], Ian E. Galbally[2*], Owen R. Cooper[3], Martin G. Schultz[4], Gerard Ancellet[5], Thierry Leblanc[6], Timothy J. Wallington[7], Jerry Ziemke[8], Xiong Liu[9], Martin Steinbacher[10], Johannes Staehelin[11], Corinne Vigouroux[12], James W. Hannigan[13], Omaira García[14], Gilles Foret[15], Prodromos Zanis[16], Elizabeth Weatherhead[3], Irina Petropavlovskikh[3], Helen Worden[13], Mohammed Osman[17], Jane Liu[18], Kai-Lan Chang[3], Audrey Gaudel[3], Meiyun Lin[19], Maria Granados-Muñoz[20], Anne M. Thompson[8], Samuel J. Oltmans[21], Juan Cuesta[15], Gaelle Dufour[15], Valerie Thouret[22], Birgit Hassler[23], Thomas Trickl[24] and Jessica L. Neu[25]

[1]Environment and Climate Change Canada, 4905 Dufferin Street, Downsview, ON, M3H 5T4 Canada

[2]Climate Science Centre, CSIRO Oceans and Atmosphere, Aspendale, Australia

[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, USA; NOAA Earth System Research Laboratory, Boulder, Colorado

[4]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

[5]LATMOS/IPSL, UPMC Univ. Paris 06 Sorbonne Universités, UVSQ, CNRS, Paris, France

[6]Jet Propulsion Laboratory, California Institute of Technology, Table Mountain Facility, Wrightwood, CA, USA

[7]Research and Advanced Engineering, Ford Motor Company, Dearborn, Michigan, USA

[8]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[9]Harvard‐Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA.

[10]Empa, Swiss Federal Laboratories for Materials Science and Technology, Duebendorf, Switzerland

[11]Dep. of Environmental Systems Science, Zürich, Switzerland

[12]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

[13]National Center for Atmospheric Research, Boulder, CO, USA

[14]Agencia Estatal de Meteorología, Izana Atmospheric Research Centre, Santa Cruz de Tenerife, Spain

[15]Laboratoire Inter-universitaire des Systèmes Atmosphériques (LISA), UMR7583, Universités Paris-Est Créteil et Paris, Diderot, CNRS, Créteil, France

[16] Department of Meteorology and Climatology, School of Geology, Aristotle University of Thessaloniki, Thessaloniki, Greece

[17]Cooperative Institute for Mesoscale Meteorological Studies, The University of Oklahoma, and NOAA/National Severe Storms Laboratory, Norman, OK, USA; now at Enable Midstream Partners, Headquarters: Oklahoma City, USA

[18]Department of Geography and Planning, University of Toronto, Canada, and School of Atmospheric Sciences, Nanjing University, Nanjing, China

[19]Atmospheric and Oceanic Sciences, Princeton University and NOAA Geophysical Fluid Dynamics Lab, Princeton, New Jersey, USA

[20]Remote Sensing Laboratory (RSLAB), Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), Barcelona, 08034, Spain

[21]Global Monitoring Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA (retired)

[22]Laboratoire d'aérologie, CNRS UMR 5560, Observatoire Midi-Pyrénée, Université de Toulouse III, Toulouse, France

[23]Institute of Atmospheric Physics, Earth System Modelling, Münchener Straße 20, 82234 Oberpfaffenhofen-Wessling

[24]Karlsruher Institut für Technologie, Garmisch-Partenkirchen, Germany

[25]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

*Co-ordinating Lead Authors and Corresponding Authors
(david.tarasick@canada.ca; ian.galbally@csiro.au)

List of Contents:

Text S-1. Instrumental method: Gas phase titration (GPT)

Gas phase titration (GPT) is a method for establishing a primary standard for ozone analysis, an alternative to UV absorption coefficients. GPT is not and has not been used for ambient air analyses.

Although used since the mid-20th century (Saltzman and Gilbert, 1959a,b; Hodgeson et al., 1971), the uncertainty of GPT was large because of the technology available at that time. Recent technical advances in gas standards and flow measurements have greatly reduced the uncertainty in GPT measurements. With optimal design of the GPT system, particularly the flow control system and the reaction vessel, the titration reaction

$$NO + O_3 \rightarrow NO_2 + O_2$$

goes virtually to completion. The ozone mole fraction is quantified through the change of mole fraction of either NO or NO_2, traceable to high-precision gravimetrically produced standards. Both NO and NO_2 are stable when isolated in cylinders. Differences between this method and standard UV photometry have been investigated by Tanimoto et al. **(**2006) and Viallon et al. (2006b; 2016). They are very small (~0.3%) when newer measurements of the ozone absorption cross-sections (see Section 2) are used (Viallon et al. 2016).

Text S-2. Comparison of modern surface ozone observations with historical records

Tropospheric ozone measurements before 1970 have been previously reviewed a number of times (e.g. Fabry, 1950; Junge, 1962; 1963; Galbally, 1971; Oltmans and Komhyr, 1976; Crutzen, 1987; Marenco et al., 1994; Staehelin et al., 1994; Calvert et al., 2015).

When compiling data for the analysis of long-term trends, scientists often ask if it is possible to compare individual current records with historical records for the same or nearby stations, e.g. central European stations (historical and current). Except in special cases, each historical set of observations of surface ozone should be seen as having a substantial unknown bias and the historical observations at a particular location will not necessarily sensibly relate to current measurements at the same location. However, when averaged across multiple stations, the normalised biases of sets of observations, if random, will tend to cancel out. Thus the average of multiple sets of these past observations for multiple years and a particular geographic region are more suited for comparison with current observations. This must be borne in mind when examining past observations that do not match current observations at the same location.

Following is a description of the logic for comparing current and historical surface ozone observations from the same or nearby stations, taking into account differences in (a) measurement technique and (b) atmospheric variability. For simplicity, a mean bias or difference between the measurement techniques used for the two sets of observations (historical and current) in ambient air is assumed. (Laboratory comparisons do not necessarily reflect the same results as field comparisons.) This analysis can be extended to a proportional difference between measurement techniques and a combination of mean and proportional difference.

For a station or nearby paired stations with surface ozone observations, we define:

t_1 and t_2 indicate the two time periods of the ozone observations: t_1 historical, and t_2 current.

$\mu(t)$ = observed yearly or long term mean surface ozone value (mole fraction) for period t.

$\mu(t_2) - \mu(t_1)$ = difference in observed surface ozone between the two time periods at this station, that has been observed, generally measured with two different types of analyser.

$d(t_2-t_1)$ = difference in surface ozone between the two time periods at this station, that would have been observed if measured by a modern analyser in a continuous long term record.

Note $\mu(t_2) - \mu(t_1)$ can be determined from the available records but $d(t_2-t_1)$ is needed for comparison against atmospheric chemical models.

$\delta\mu(t_1)$ = for a short sampling period, the deviation of the observed surface ozone record from what would have been the long term average for that station at that time.

$I(t_1)$ = the mean difference between the standard method used in period t_2 and the measurement technique used in period t_1 for simultaneous ambient measurements. This is the simplified representation of instrument bias as a difference in this analysis.

$\delta I(t_1)$ is the uncertainty in the measured mean value of $I(t_1)$.

$Is(t_1,t_2)$ = the uncertainty in difference between the measurements made with the different measurement techniques used in period t_1 and t_2 respectively when only one or a few measurements are made. This is distinct from the uncertainty associated with the mean difference.

For yearly average historical surface ozone records, primarily available since 1950, we assume the yearly averaging reduces to a minimum the magnitude of $\delta\mu(t_1)$ and $ls(t_1,t_2)$, and:

$\mu(t_2) - \mu(t_1) = d(t_2-t_1) + I(t_1) + \delta l(t_1)$

For single station comparisons, whether $I(t_1) + \delta l(t_1) << d(t_2-t_1)$ or not determines whether $\mu(t_2) - \mu(t_1)$ and $d(t_2-t_1)$ are approximately equal or not.

For historical surface ozone records which are of a single measurement or a few days, which are the most common cases, and dominate the pre 1950 records:

$\mu(t_2) - \{\mu(t_1) + \delta\mu(t_1)\} = d(t_2-t_1) + I(t_1) + \delta l(t_1) + ls(t_1,t_2)$

For the case of one or few historic observations, the analysis is complicated by three factors. Firstly, the results of the limited number of historical observations will differ from the long term mean surface ozone at the station by an amount influenced by diurnal, synoptic and seasonal influences that occurred when the observations were made. Furthermore the difference between the two measurement techniques now includes the uncertainty involved with one or a few measurements, as well as the uncertainty associated with the multi-measurement average.

It is possible, from the characteristics of the modern observations at the station, and knowledge of the date, time of day, weather etc. of the historical measurements to make an estimate of $\delta\mu(t_1)$ with an associated uncertainty.

With regard to differences and uncertainty associated with comparing two measurement techniques there are some issues:

- most of the instrument comparisons available (Table 2 in the main text) are historical and no detailed information on the protocol is available. The results are typically expressed as a single value, either the difference or ratio of the paired observations from the two techniques. Hence, little or no information is available about $\delta l(t_1)$ and $ls(t_1,t_2)$. Regretfully, the historical ozone instrument comparisons presented are the only ones identified for these historical measurement techniques applied in the regional or background atmosphere.
- there is a recognised tendency for experimenters to consciously or unconsciously bias the data towards agreement, (Morgan and Henrion 1990), which is the reason for blind and double blind comparison protocols today. Thus, estimates of $I(t_1)$, $\delta l(t_1)$, and $ls(t_1,t_2)$ are likely to be biased, the first towards 1 and the latter two minimised.
- measurements made away from the intense scrutiny that occurs during comparisons (i.e. under normal operating conditions) can have additional influences that affect the quality of observations and increase the measurement uncertainty, including technician availability, environmental conditions of the monitoring instrument etc.

At this time there is inadequate information to quantify the values and uncertainties in $I(t_1)$ and $ls(t_1,t_2)$ for ambient instrument comparisons of historical surface ozone measurement techniques. What is required is some critical evaluation of the contribution of measurement uncertainties to the observed differences between individual historical and modern measurements. A rough guide to these uncertainties may be obtained from knowledge of the bias on BM ozonesondes in the lower troposphere in the period 1970 – 2000, (Section 4.3 of the main text) which may give a lower limit to the uncertainties associated with historical surface ozone instrument biases. The

lower limit is assumed because knowledge of ozone measurements and scientific and technical resources available were far superior in 1970 to 2000 compared with 1950 – 1970 or pre 1950. The biases of BM relative to ECC sondes from 11 separate studies during 1970 to 2000 vary over time from -20% ± 30% to + 10% ± 15%. These values are an estimate of the lower limit of uncertainties in $I(t_1)$ and $ls(t_1,t_2)$ for ambient instrument comparisons of historical surface ozone measurement techniques, and should be taken into consideration when assessing the uncertainties in individual ozone instrument records, unless better information is available.

In summary, the advantage of grouping multiple stations over multiyear time periods is based on the likelihood that there are some random aspects to these uncertainties and that they cancel out.

Text S-3. Data selection identified as questionable under Criterion 4

For surface ozone, a set of four criteria for selecting data for the historical reconstruction were developed and applied: (1) the measurement methods used should be related, through intercomparisions, to the current standard UV absorption photometer method; (2) the likelihood for significant contamination of the ozone measurements due to interfering pollutants in the sampled atmosphere should be low; (3) the air sampled should be representative of the well-mixed boundary layer, and (4) recognizing the uncertainties associated with all of the historical data sets the measurements should be free from extreme artifacts or inconsistencies. Considerations regarding Criterion 4 are examined here in greater detail: published data may be suspect for a variety of reasons, including inconsistency with other observations, or artifacts or outliers that suggest instrument problems. The outstanding example is Pring (1914) where the measurements appear to be of excellent quality and well documented, except that they differ by a factor of 100 from current measurements. As Fabry (1950) wrote, the results are absurd. Further discussion of credibility follows.

There are five sets of data in this review that in part or whole are marked as questionable in the context of Criterion 4. Three are from the Northern High Latitudes and show anomalously high surface ozone values inconsistent with current observations (Helmig et al. 2007) in the Arctic. The fourth is from Antarctica and the fifth from a 16-station network stretching from northern Europe to southern Africa. In each case the effect of including or omitting these data from the analysis is considered, and the results are found to be not sensitive to that choice.1) Dauvillier (1934) used the KI-arsenite method at Scorsby Sund, Greenland, from November 1932 to August 1933. His daily data can be notionally divided into two sets, 225 days of data show background values of about 5 mg 100 m^{-3}, with less than 5% of values over 10 mg 100 m^{-3}. In contrast, there is a 45-day period including 4 weeks in December-January and another 2 weeks in February which shows several major events with ozone reaching as high as 57 mg 100 m^{-3} (approximately 270 nmol mol^{-1}). Dauvillier (1934) identified these events as anomalous but could not identify any instrumental fault. This work is discussed in detail in the main text. The conclusion is that these measurements during the six weeks of major events are not considered credible and they are omitted from the subsequent analysis, while the other 225 days of ozone measurements are accepted.

2) At College, Alaska Wilson et al. (1952) used the same KI-arsenite ozone analysis method as Dauvillier (1934, 1935). The observations are divided into two periods. For the first 7 months the inlet was 3 ft (0.9 m) above the ground, protruding from the side of an unidentified building. Subsequently it was moved to 25 ft (7.6 m) level on the new Sydney Chapman Building on the campus of the University of Alaska Fairbanks. The maximum ozone observed during the first period was 75 mg 100 m^{-3} (~ 345 nmol mol^{-1}) based on 24-hour means, and during the second period the maximum value was 14 mg 100 m^{-3} (~64 nmol mol^{-1}). Similarly the mean and standard deviations for the two periods were 54 ± 31 nmol mol^{-1} and 27 ± 12 nmol mol^{-1}. In the 29 data sets of historic surface ozone observations where there is a year or more of observations, the largest other standard deviation is 16 nmol mol^{-1}. Given the discussion in the main text about the credibility of the same analysis method at Scoresby Sund by Dauvillier (1934), and the anomalously high variance in the first data period, the considered judgement here is to omit the initial 7 months of data as not credible and to accept the data from the second period.

3) At North Lake Meadow near Barrow, Alaska, Kelley (1967, 1970, 1973) undertook a carefully documented set of ozone measurements including laboratory calibrations of the field

instruments against the NBKI method. The ambient observations were made with a 2 m inlet height during 1965-67. Six years later, a continuous and ongoing ozone time series, based on the UV absorption method, was initiated at Barrow by NOAA's Global Monitoirng Division. The ten greatest 24-hr ozone values measured by UV instruments at Barrow are in the range 47.8 – 52.8 nmol mol^{-1}, and all occurred in spring. The highest value observed in the 1970s was 42 ppb. In contrast, just 6-8 years earlier, 15% of the Kelley data exceeded 53 nmol mol^{-1} (based on 24-hr means). The most extreme anomaly in the Kelley record is the near constant occurrence of daily values above 60 nmol mol^{-1} during November 1966, with a monthly mean of approximately 65 nmol mol^{-1}. The considered judgement here is that the Kelley data set is likely biased high and has questionable credibility in providing a continuous ozone record at an individual site, but in the context of utilizing a multiple sets of observations with varying biases to define the average ozone levels in the historic atmosphere, it is included in the historic data.

4) The fourth questionable data set is the Halley Bay, Antarctica time series from 1958, presented by MacDowall (1962). The ozone levels observed are very low compared with other stations operating in Antarctica about this time. The two possibilities of either (a) instrument malfunction or (b) ozone depletion events were proposed by Roscoe and Roscoe (2006). MacDowall (1962) wrote that "The main trouble with the instrument was blockage of the air-flow capillary." This happened 25 times in March reducing to 5 times in August. Given that this blockage effectively reduced the output reading of the instrument, then, without the instrument logs, it is impossible to fully evaluate the relative roles of the two explanations above in contributing to the lower ozone levels in this set. Hence the considered judgement here is that the data set has questionable credibility, but in the context of utilizing multiple sets of observations with varying biases to define the average ozone levels in the historic atmosphere, it is included in the historic data.

5) The fifth questionable data set is comprised of the 16 stations in the TROZ network (Fabian and Pruchniewicz 1977; hereafter F&P (1977)). The TROZ data set poses a major challenge for this historic analysis due to the following four issues: (a) the traceability of the method to the modern UV standard; (b) absence of interfering or ozone-destroying pollutants (c) representative sampling and the ozone metric chosen to represent the TROZ network; (d) the high number of TROZ sites compared to the total number of sites compiled for TOAR-Observations with respect to the statistical arguments.

As discussed in Section 2.1.3, the MPI Pruch method is traceable to the modern UV standard, see Table 2, (a) indirectly via ambient measurements with the HP-KI method at Hohenpeissenberg, which gives a ratio of 0.5, and (b) indirectly via the Ehmert method which gives a ratio of 1.0. As stated, there is insufficient evidence available to explain this difference. Thus, with regard to traceability, the entire TROZ ozone data should be either accepted as is or multiplied by a factor of two.

The sites selected by F&P (1977) are not described in detail: no specific locations or map references are given, nor inlet heights, surrounding buildings or nearby sources of pollutants. For contrast see Warmbt (1964). Fabian and Pruchniewicz (1975) state that at Lindau (52^{o}N) there were relatively high pollution levels that affected the ozone results and therefore that site is not included in the subsequent analysis. A visual inspection of the records from other sites suggests the presence of interfering or ozone-destroying pollutants; see the Zugspitze record for April and May 1969 (Pruchniewicz 1973). The question is, in the absence of any other information, can the whole data set be accepted as being free of the negative influence of interfering or ozone destroying pollutants? Probably not.

Throughout the analysis in this paper the criteria for representative sampling are a well exposed sampling point, a well exposed site and averaging over the hours of the day when boundary layer mixing is greatest. F&P (1977) put particular emphasis on reporting ozone data from well mixed conditions. First, the maximum 5 minute ozone value for each day was determined and then the monthly mean of the daily maxima. Next F&P (1977) limited the reported monthly value to the average of all daily maxima exceeding the respective monthly mean of daily maxima based on all days in the month. An analysis using this procedure and the time series of modern UV observations at Hohenpeissenberg, Germany indicated that these final values obtained using the F&P (1977) procedure are similar to the 75^{th} percentiles of all daily maxima or the 95^{th} percentile of the daytime means in a given month. Thus to make the F&P (1977) data comparable with the other historic and modern ozone data used here, a substantial negative correction varying with site and in the range 10 to 40% is required. This correction has not been applied by TOAR-Observations.

Finally, this analysis has utilized the premise that the grouping of multiple stations over multiyear time periods, run by different organisations making measurements by different methods is consistent with the assumption that there are some random aspects to these uncertainties and that they cancel out. However for the 1950s-1970s between 90^{o}N and 60^{o}S F&P (1977) contribute 15 stations which, for the reasons described in the other three issues above, will have correlated errors. This constitutes more than half the number of stations in these regions for this period. Thus the assumption of cancelling errors is most likely invalid.

In summary, on all 4 issues the data of Fabian and Pruchniewicz (1977) appears to have substantial errors and could potentially bias the 1950s-1970s historic analysis. The data are not corrected because the two main corrections applied would be large, uncertain and in opposite directions. For robust results the analysis is performed with and without the Fabian and Pruchniewicz (1977) data.

Text S-4. A note on units and nomenclature

The unit of ozone measurement appropriate to surface air and in the troposphere is mole fraction, based on the International System of Units, SI, (http://www.bipm.org). Under tropospheric conditions of temperature, pressure and ozone amount, the mole fraction is indistinguishable from volumetric mixing ratio, of which parts per billion, ppb is commonly used and equivalent to nanomole per mole (Galbally et al. 2013).

A key concern when using older data is that before the 1970's it was quite common for the ozone concentration to be presented in units of $\mu g\ m^{-3}$. But the volume m^{-3} is ambiguous: is it at ambient conditions or at STP? In many cases this information is not supplied. If the site is near sea level, then the ambient conditions and STP can be assumed to be the same, within the uncertainty of this historical reconstruction. However for elevated or mountain top sites, the appropriate correction must be identified. For example the 16 station network TROZ (Fabian and Pruchniewicz 1977) reported ozone concentrations in $\mu g\ m^{-3}$ without more detailed specification. Another report (Fabian and Pruchniewicz 1975) clarifies that the ozone concentrations were reported under conditions of STP, which is taken into account when converting their reported values to units of $nmol\ mol^{-1}$

When referring to integrated ozone amounts, the nomenclature used in this paper is: "total ozone" = total column amount of ozone in the atmosphere, while SCO = stratospheric column amount of ozone, and TCO = tropospheric column amount of ozone.

Text S-5. Standard deviations used in representing data sets

The standard deviations differ for the periods 1890-1950 and 1950-1970 because of the paucity of data in the earlier period. In the period 1890-1950 the standard deviation is calculated from daily ozone values when there are 7 or more days of data. The standard deviation represents the dispersion of individual days of measurements. Where only the range is available from the original paper, the standard deviation is estimated as a quarter of the range. In the period 1950-1970 the standard deviation is calculated from the monthly mean ozone values for an annual cycle or longer when there is at least one year of data. The standard deviation represents the dispersion of monthly values from the yearly mean. In the case of Fabian and Pruchniewicz (1977) the standard deviation is calculated using the uncertainty in the annual mean value plus the amplitude of the annual cycle. Where only the range is available from the original paper, the standard deviation is estimated as a quarter of the range.

Text S-6. Historical measurements of urban and suburban ozone in northern Europe

In addition to the long record at Montsouris, there are occasional measurements of surface ozone in urban or suburban environments. It is interesting to compare these with the Montsouris average. Götz et al. (1935) report a measurement in the city of Zurich, giving 12 nmol mol^{-1} after correction for the absorption coefficients of Lauchli (1929). Edgar and Paneth (1941b) present measurements of ozone in the air in a street of South Kensington, London, UK, made from February to July 1938. They used a cryogenic trapping and purification method followed by both UV and KI analysis (Paneth and Edgar (1938); Edgar and Paneth, 1941a; 1941b).

Rassool et al. (1956) and Vassy (1958) present an interesting comparison of UV measurements made at the same time as coulometric KI measurements, at their site at Val Joyeux, about 30 km from Paris. On several days with lower wind speeds the KI measurements were much lower (4.4 nmol mol^{-1}) than the long-path spectroscopic UV measurements (13.6 nmol mol^{-1}), although on days with more wind the results from the two methods were identical within experimental error.

The historical urban or semi-urban measurements, including the measurements from the Observatoire Municipale de Montsouris in Paris discussed in Section 3.1, are listed in Table S-1 and shown in Figure S-1. These values are much lower than the rural observations in Figure S-2, as are the corresponding averages of modern UV measurements at urban and suburban sites in the TOAR database. Like the Montsouris measurements, the 20th century measurements also likely represent local conditions, subject to NO titration, and in the case of the KI measurements, possible interference from SO_2. The four data selection criteria have not been applied to these data.

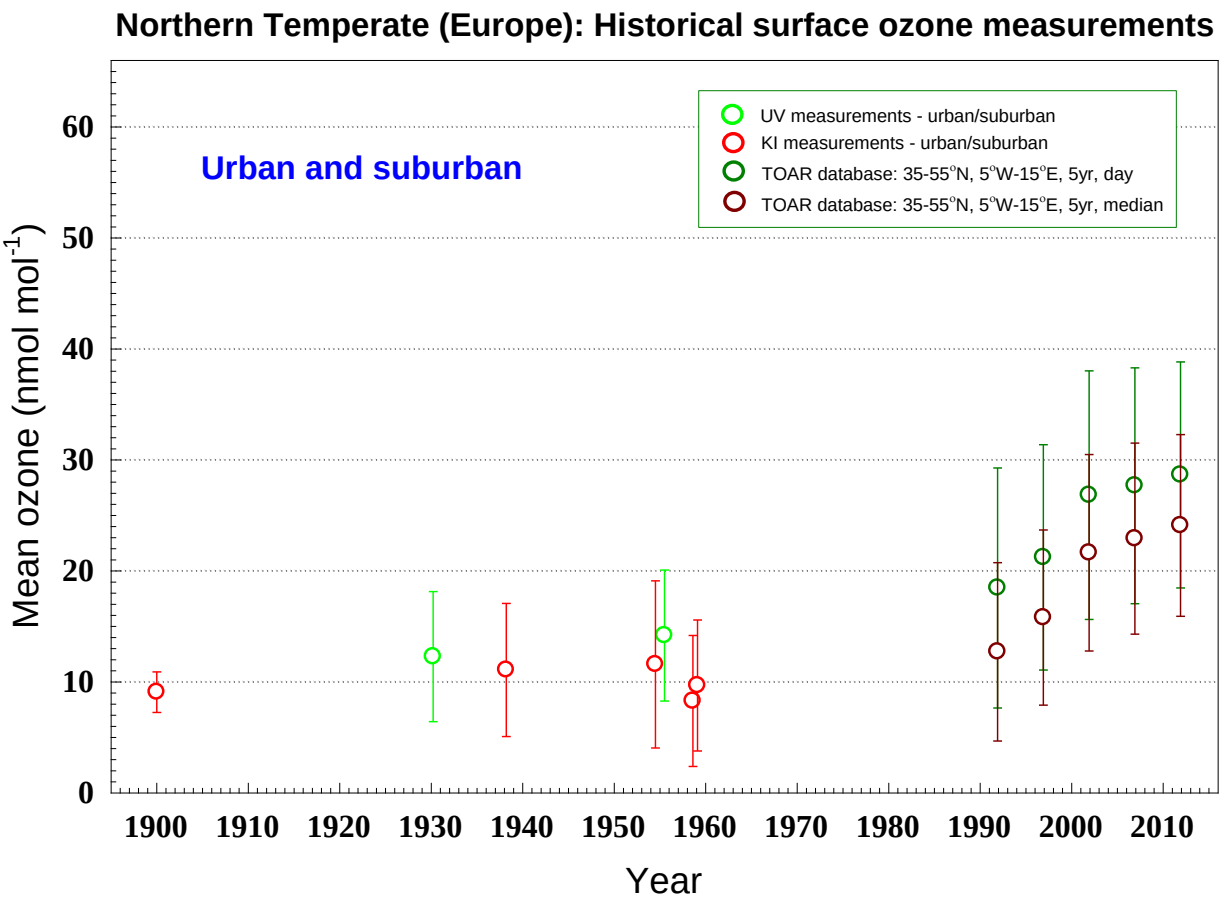

Figure S-1: Some historical measurements of urban and suburban ozone in northern Europe. Five-year averages of modern UV measurements at sites classified as "urban" in the TOAR database are also shown (Schultz et al., 2017).

Table S-1: Some historical measurements of urban and suburban ozone in northern Europe. Mean values and standard deviations in nmol mol^{-1}. See Text S-7 for Notes.

Location	Period	Method	Reference	Range	Days	Mean	SD	Notes
1877-1955 Northern Temperate – Europe								
Montsouris 48.8N, 2.3E, 75m	1877-1903	KI-arsenite	Albert-Lévy (1903); Volz and Kley (1988)	1.8 mg/100m^3	9000	9	3	
Zurich 47.4N, 8.5E, 500m	April 30, 1930	UV	Götz et al. (1935)	10 nmol mol^{-1}	1	12.2	n/a	2
London 51.5N, 0.1W, 35m	Feb. 7 - May 10, 1938	Cryogenic, KI and UV analysis	Paneth and Edgar (1938); Edgar and Paneth (1941b)	5-23 nmol mol^{-1}	19	11.0	6.0	
Val-Joyeux 48.8N, 2.0E, 114m	1954	KI coulometric	Vassey (1958)	11.5 nmol mol^{-1} (daytime mean)	180	12.4	n/a	31
Val-Joyeux	May-July 1955	UV	Rasool et al. (1956)	15.3 nmol mol^{-1}	8	14.1	5.9	23
Tokyo 35.7N, 139.7E, 40m	Aug.-Sept. 1958	Ehmert	Kawamura and Sakurai (1966)	16 μg/m^3	21	8.2	n/a	
Tokyo	Feb. 1958 and 1959	Ehmert	Kawamura and Sakurai (1966)	20 μg/m^3	21	9.6	n/a	

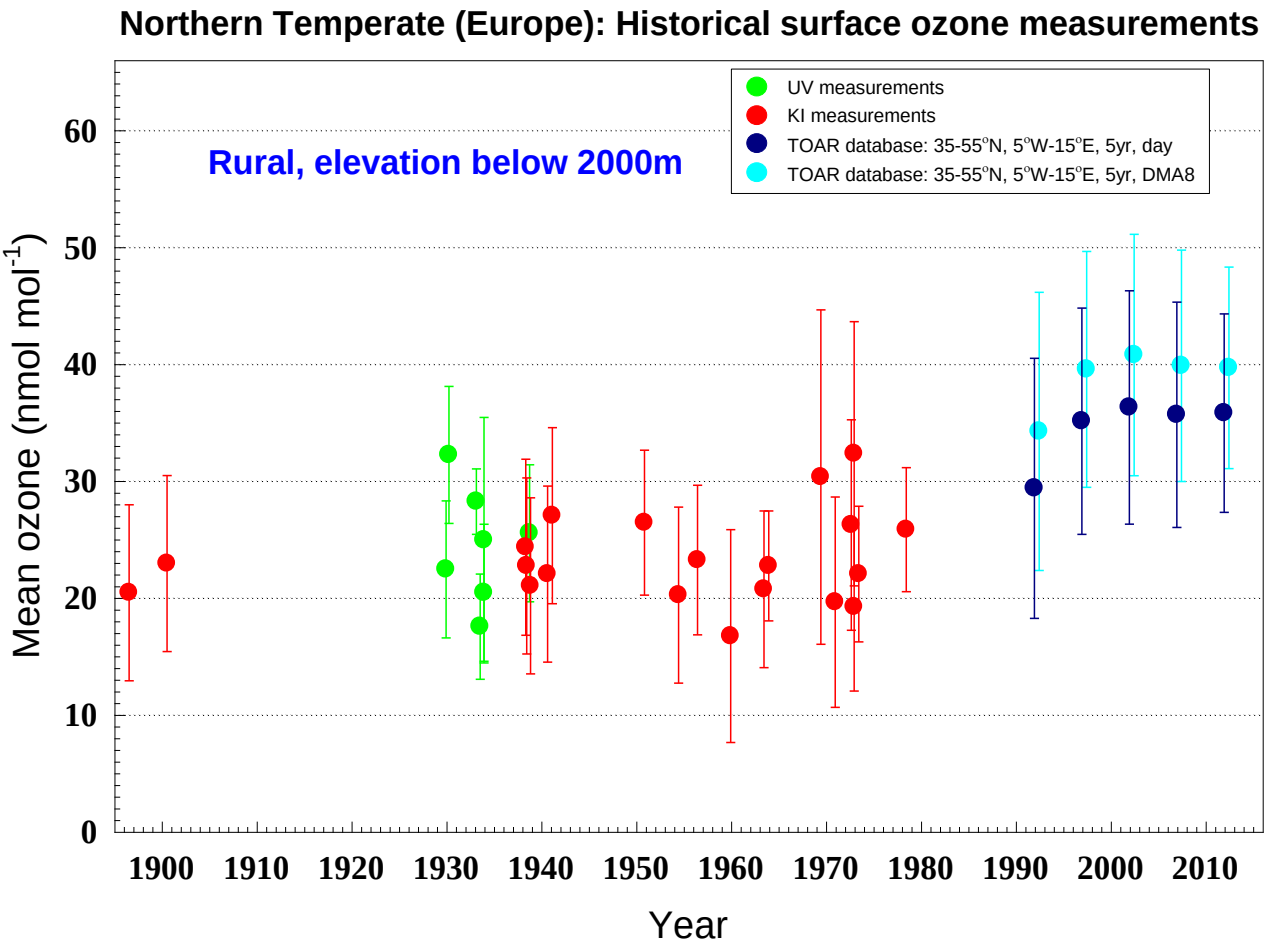

Figure S-2: Historical measurements of rural background surface ozone at low altitudes in northern temperature regions (almost exclusively Europe). As Figure 2 in the main text, except that all measurements at sites above 2000 m (Grands-Mulets, Mont Blanc, Jungfraujoch, Zugspitze, Mt. Norikura and Capillo Peak in Table 4) have been excluded.

Text S-7. Notes on Tables 4-6, 8, S-1

1. Corrected for coefficients of Fabry and Buisson (1931), at $\lambda \geq 270$ nm (+2%).
2. Corrected for coefficients of Lauchli (1929), at $\lambda > 254$ nm (+11%).
3. Corrected for coefficients of Lauchli (1929), at $\lambda > 260$ nm (+10%), and Fabry and Buisson (1931), (-5%). The different ozone values given for the two sets of coefficients agree when converted to modern (Hearn) values.
4. Wavelengths from 310 to 257.6 nm. April 29 (34.4 nmol mol^{-1}) doubtful, not included. +6.5% correction for Ny and Choong (1933) coefficients.
5. Details in Chalonge and Vassy (1934).
6. Corrected to NTP by Fabry (1950).
7. Values corrected for stoichiometry by Volz.
8. Estimated number of days from observations presented. Excluded values with wind from Durham city (ozone values drop to 0).
9. Mean and standard deviation estimated from time series plot; standard deviation estimated; one quarter of range.
10. Monthly means estimated from time series plot; standard deviation of monthly means.
11. Tower 80m, afternoon averages. Combined 3 periods. Value increased by 11% based on Table 2.
12. Added 1.5 nmol mol^{-1} for SO_2 interference (Feister and Warmbt, 1987). The degree to which SO_2 interferes with the Cauer method is not clear (Warmbt, 1964; Feister and Warmbt, 1987). Value increased by 11% based on Table 2. Note that this correction is conservative, based on Table 2.
13. No correction -- SO_2 filter installed. Value increased by 11% based on Table 2.
14. Means of daily maxima. Standard deviation calculated from fitted data from Fabian & Pruchniewicz (1977). For low bias at Hohenpeissenberg and Zugspitze and other issues with this data set, see Section 2.1.3 in the main text and the section above: "Data selection identified as questionable under Criterion 4"
15. Background value 5 mg/$100m^3$; low 2; values of 10 or more common; the maximum value of 57 mg/$100m^3$ corresponds to a mixing ratio of 262 nmol mol^{-1} which greatly exceeds any 24-hour mixing ratio in the TOAR database. Anomalously high values were deemed unreliable and were omitted from the average value. Median estimated from time series plot. For further details see section above: "Data selection identified as questionable under Criterion 4"
16. Standard deviation estimated as one quarter of range.
17. Some very high values (over 300 nmol mol^{-1}) in April and May. As for the Dauvillier (1934) data, these outliers are excluded, by omitting the first 7 months, and taking the median of the remaining 5 months. For further details see section above: "Data selection identified as questionable under Criterion 4"

17a. Likely biased high. For further details see section above: "Data selection identified as questionable under Criterion 4"

18. Soot in air intake noted with concern by Wexler et al., but ozone values did not change when this was (periodically) cleaned. Data accepted.
19. Mean and standard deviation are for composite year; standard deviation of monthly means.
20. Values increased by 20% based on Table 2. For further details see section above: "Data selection identified as questionable under Criterion 4"
21. Value increased by 11% based on Table 2. Note that this correction is conservative, based on Table 2.
22. Afternoon averages.
23. Corrected for coefficients of Vigroux (1953), (-8%).

24. Estimated from data plot; assumed coefficients of Lauchli (1929), at λ>260 nm, corrected (+10%).
25. Estimated from data plots in Fabry (1950); Possibly only 5 DU
26. Estimated from data plots; Corrected for coefficients of Lauchli (1929), at λ>260 nm (+10%)
27. Estimated from data plots; wavelengths from 290 to 330 nm. +13% correction for Ny and Choong (1933), Fabry and Buisson (1931) coefficients. A total of 19 flights were made, but data were published for only 11 of them.
28. Estimated from data plots; apparatus based on Regener (1938b), so no thiosulphate stabilization; titrations done after flight in lab.
29. Corrected for absorption coefficients quoted at 306 and 318 nm (+24%).
30. Estimated from data plots.
31. Value increased by 7% based on Table 2.

Table S-2 (next page): Change of ozone from the historical period to the modern period for 3 categories of observations: 1) all available historical observations below 2000 m elevation (blue rows); 2) same as for Category 1, but with all observations from Fabian and Pruchniewicz (1977) omitted (gray rows); 3) same as for Category 2, but with three questionable high latitude data sets omitted (Barrow, College and Halley Bay) and with the Northern Temperate sites limited to just those that measured ozone with UV methods (white rows). The modern period is based on rural ozone observations below 2000 m for the years 1990-2014. ***The metric for the modern data is the 12-hr daytime average.*** The modern data were extracted from the TOAR Surface Ozone Database and reduced to monthly means across 5°x5° grid cells. In addition, modern data for the Northern Temperate zone is provided for a large region covering western Europe (16 5°x5° grid cells), and also provided for a single 5°x5° grid cell which encompasses most of the historical observations from western Europe. Within this grid cell monthly means from the individual rural sites were used, rather than using the multi-site average across the grid cell.

Region	Sites included[3]	Historical[1]	Modern[1] (daytime)	Wilcoxon test[2]	p	*t*-test[2]	p
Northern High Latitude	36-43	23.1 (8)	30.5 (21)	**38.7% (8.4%, 53.9%)**	**0.01**	**31.8% (2.9%, 60.7%)**	**0.05**
	36-38, 42, 43	24.8 (5)	30.5 (21)	26.7% (-5.2%, 54.3%)	0.20	23.1% (-15.3%, 61.4%)	0.30
	36, 37, 38	19.4 (3)	30.5 (21)	**55.7% (30.3%, 89.1%)**	**0.01**	**57.1% (15.2%, 99.0%)**	**0.04**
Northern Temperate (modern gridded data, using 16 5°x5° grid cells across western Europe: 35-55°N and 5°W-15°E)	1, 3, 6-11, 13-14, 16-19, 21-28, 30-32, 34-35	23.6 (27)	36.1 (16)	**51.1% (38.5%, 67.3%)**	**<0.01**	**53.1% (29.8%, 76.4%)**	**<0.01**
	1, 3, 6-11, 13, 14, 16-19, 21-26, 34-35	23.6 (22)	36.1 (16)	**51.0% (36.8%, 67.9%)**	**<0.01**	**52.9% (29.4%, 76.4%)**	**<0.01**
	6-11, 13	24.4 (7)	36.1 (16)	**46.9% (25.6%, 70.2%)**	**<0.01**	**47.7% (16.9%, 78.5%)**	**0.01**
Northern Temperate (modern data individual sites in the 5°x5° region, 45-50°N and 5-10°E)	1, 3, 6-11, 13-14, 16-19, 21-28, 30-32, 34-35	23.6 (27)	32.5 (23)	**40.8% (25.1%, 51.8%)**	**<0.01**	**37.8% (22.8%, 52.8%)**	**<0.01**
	1, 3, 6-11, 13, 14, 16-19, 21-26, 34-35	23.6 (22)	32.5 (23)	**40.7% (25.1%, 51.1%)**	**<0.01**	**37.6% (22.4%, 52.9%)**	**<0.01**
	6-11, 13	24.4 (7)	32.5 (23)	**32.0% (9.8%, 52.8%)**	**0.01**	**33.0% (9.3%, 56.6%)**	**0.03**
Northern Low Latitude	50, 52, 53	22.2 (3)	31.8 (12)	44.2% (-3.1%, 90.6%)	0.07	**43.4% (9.9%, 76.9%)**	**0.04**
	50, 52	25.5 (2)	31.8 (12)	21.5% (-25.1%, 73.3%)	0.26	24.8% (-10.2%, 60.0%)	0.30
	50, 52	25.5 (2)	31.8 (12)	21.5% (-25.1%, 73.3%)	0.26	24.8% (-10.2%, 60.0%)	0.30
Southern Low Latitude	54-57	17.5 (4)	16.3 (5)	-9.8% (-58.0%, 56.4%)	0.73	-7.0% (-54.8%, 40.8%)	0.78
	None available						
	None available						
Southern Temperate	58-60	23.3 (3)	23.5 (10)	1.8% (-13.6%, 17.0%)	0.94	0.8% (-34.3%, 35.8%)	0.97
	59, 60	24.6 (2)	23.5 (10)	-8.1% (-17.3%, 11.7%)	0.61	-4.6% (-46.5%, 37.4%)	0.86
	59, 60	24.6 (2)	23.5 (10)	-8.1% (-17.3%, 11.7%)	0.61	-4.6% (-46.5%, 37.4%)	0.86
Southern High Latitude	44, 45, 47-49	20.5 (5)	24.1 (6)	22.4% (-8.8%, 42.5%)	0.33	17.9% (-17.2%, 52.9%)	0.34
	44, 45, 47-49	20.5 (5)	24.1 (6)	22.4% (-8.8% 42.5%)	0.33	17.9% (-17.2%, 52.9%)	0.34
	44, 47-49	21.6 (4)	24.1 (6)	16.7% (-9.5%, 31.5%)	0.61	11.6% (-24.9%, 48.0%)	0.56

[1]Values indicate the mean ozone in nmol mol^{-1}, with sample size in parentheses
[2]Values indicate the percent change in mean ozone from the historical to the modern period with the confidence interval in parentheses.
[3]Site numbers correspond to the historical sites listed in Tables 4, 5 and 6.

Table S-3 (next page): Change of ozone from the historical period to the modern period for 3 categories of observations: 1) all available historical observations below 2000 m elevation (blue rows); 2) same as for Category 1, but with all observations from Fabian and Pruchniewicz (1977) omitted (gray rows); 3) same as for Category 2, but with three questionable high latitude data sets omitted (Barrow, College and Halley Bay) and with the Northern Temperate sites limited to just those that measured ozone with UV methods (white rows). The modern period is based on rural ozone observations below 2000 m for the years 1990-2014. ***The metric for the modern data is the daily maximum 8-hour average (DMA8)***, in order to focus on well-mixed conditions. The modern data were extracted from the TOAR Surface Ozone Database and reduced to monthly means across 5°x5° grid cells. In addition, modern data for the Northern Temperate zone is provided for a large region covering western Europe (16 5°x5° grid cells), and also provided for a single 5°x5° grid cell which encompasses most of the historical observations from western Europe. Within this single grid cell monthly means from the individual rural sites were used, rather than using the multi-site average across the grid cell.

Region	Sites included[3]	Historical[1]	Modern[1] (DMA8)	Wilcoxon test[2]	p	*t*-test[2]	p
Northern High Latitude	36-43	23.1 (8)	33.1 (21)	**49.8% (21.4%, 66.0%)**	**<0.01**	**43.3% (14.5%, 72.0%)**	**0.02**
	36, 37, 38, 42, 43	24.8 (5)	33.1 (21)	**35.4% (4.7%, 64.1%)**	**0.02**	33.8% (-4.4%, 72.0%)	0.15
	36, 37, 38	19.4 (3)	33.1 (21)	**69.9% (39.0%, 102.2%)**	**<0.01**	**70.8% (29.0%, 112.5%)**	**0.04**
Northern Temperate (modern gridded data, using 16 5°x5° grid cells across western Europe: 35-55°N and 5°W-15°E)	1, 3, 6-11, 13-14, 16-19, 21-28, 30-32, 34-35	23.6 (27)	40.3 (16)	**69.5% (57.2%, 84.1%)**	**<0.01**	**70.8% (46.9%, 94.6%)**	**<0.01**
	1, 3, 6-11, 13-14, 16-19, 21-26, 34-35	23.6 (22)	40.3 (16)	**69.8% (55.9%, 85.3%)**	**<0.01**	**70.6% (46.5%, 94.7%)**	**<0.01**
	6-11, 13	24.4 (7)	40.3 (16)	**65.0% (44.3%, 84.8%)**	**<0.01**	**64.8% (33.6%, 95.9%)**	**<0.01**
Northern Temperate (modern data using individual sites in the 5°x5° region, 45-50°N and 5-10°E)	1, 3, 6-11, 13-14, 16-19, 21-28, 30-32, 34-35	23.6 (27)	37.5 (25)	**62.8% (47.3%, 72.7%)**	**<0.01**	**59.0% (44.3%, 73.7%)**	**<0.01**
	1, 3, 6-11, 13-14, 16-19, 21-26, 34-35	23.6 (22)	37.5 (25)	**62.2% (47.2%, 72.5%)**	**<0.01**	**58.8% (43.8%, 73.9%)**	**<0.01**
	6-11, 13	24.4 (7)	37.5 (25)	**52.4% (32.7%, 73.3%)**	**<0.01**	**53.4% (29.9%, 76.9%)**	**<0.01**
Northern Low Latitude	50, 52, 53	22.2 (3)	35.1 (12)	**58.9% (12.9%, 105.7%)**	**0.03**	**58.5% (24.3%, 92.6%)**	**0.02**
	50, 52	25.5 (2)	35.1 (12)	35.3% (-17.2%, 91.4%)	0.13	38.0% (-3.7%, 79.7%)	0.16
	50, 52	25.5 (2)	35.1 (12)	35.3% (-17.2%, 91.4%)	0.13	38.0% (-3.7%, 79.7%)	0.16
Southern Low Latitude	54-57	17.5 (4)	19.0 (5)	5.6% (-48.6%, 78.8%)	0.90	8.6% (-41.9%, 58.9%)	0.75
	None available						
	None available						
Southern Temperate	58-60	23.3 (3)	25.7 (10)	9.4% (-6.5%, 27.1%)	0.61	10.2% (-25.1, 45.4%)	0.57
	59, 60	24.6 (2)	25.7 (10)	0.9% (-8.1%, 26.9%)	0.91	4.4% (-37.7%, 46.4%)	0.87
	59, 60	24.6 (2)	25.7 (10)	0.9% (-8.1%, 26.9%)	0.91	4.4% (-37.7%, 46.4%)	0.87
Southern High Latitude	44, 45, 47-49	20.5 (5)	25.9 (6)	29.6% (-0.5%, 47.2%)	0.05	26.6% (-9.1%, 62.3%)	0.18
	44, 45, 47-49	20.5 (5)	25.9 (6)	29.6% (-0.5%, 47.2%)	0.05	26.6% (-9.1%, 62.3%)	0.18
	44, 47-49	21.6 (4)	25.9 (6)	24.9% (-3.9%, 34.9%)	0.11	19.8% (-17.2%, 56.8%)	0.33

Text S-8. Analysis of changes in surface ozone

Figure S-2 indicates an increase of about 12-16 nmol mol-1 between the historical record and the modern period, depending on whether the 12-hr daytime average or the DMA8 metric is used for the modern data. While afternoon averages were chosen for the historical data where available (see Tables 4-6 and notes above), in a number of cases only daytime or 24-hour averages were available, and the UV measurements were all made at night. Tables S-2 and S-3 therefore show analyses of the estimated increases based on each metric.

The historical datasets in each are treated as one sample, with equal weights and a simple mean not weighted by the size of individual datasets (unlike the weighted means in Table 4-6). The sample variance is estimated is estimated as the average of the individual variances; where the latter is not available a standard deviation of 0.3 of the mean is assumed. For the modern data, means and variances are calculated from (up to) 300 monthly means of all available station data for 1990-2014, weighted by the number of stations reporting in each month. For each region, means are calculated for each 5°x5° grid cell with at least one modern station, and the mean and variance for the region is calculated from the pooled grid-cell means and variances. For the Northern Temperate region, where there are the largest numbers both of historical and modern datasets, more detailed estimates are feasible, and two are presented: the former uses a mean over 16 5°x5° grid cells across western Europe (35-55°N and 5°W-15°E), with the mean and variance calculated from the 16 grid-cell means and pooled variances, and the latter uses data from up to 32 modern sites within the 5°x5° region (45-50°N and 5-10°E) where most of the early measurements, and in particular all of the long-path UV measurements, were taken. The purpose of these different estimates is to explore the sensitivity of the calculated change in ozone to the modern data used for comparison, as contemporary measurements of surface ozone appear to vary more than historical estimates.

All samples, of both the historical and modern data, pass the Shapiro-Wilk and Kolmogorov-Smirnov tests for normality, so the *t*-test for unequal variances (Welch, 1947) is valid. Nevertheless calculations are also made using the non-parametric Wilcoxon rank sum test (Mann and Whitney, 1947)), which does not assume that the sample data are normally distributed.

Text S-9. Umkehr retrieval algorithms

The Dobson Umkehr measurements are composed of the ratio of the intensities of scattered light detected at the C-pair wavelengths that are centred at 311.5 nm (strong ozone absorption) and 332.4 nm (weak ozone absorption). The Umkehr intensity curve is recorded while the SZA changes between 60 degrees and approximately 90 degrees. Dobson Umkehr retrieval algorithms have evolved over the last 20 years to derive ozone profiles from morning and afternoon measurements depending on clear sky conditions (Mateer and DeLuisi, 1992; Petropavlovskikh et al., 2005, Petropavlovskikh et al., 2011).

The Brewer instrument makes zenith sky measurement by recording the intensities of polarized zenith-sky light at five wavelengths nearly simultaneously in two, partially overlapping, wavelength bands (Kerr et al. 1985; Kerr and McElroy, 1995; Kerr, 2002; Cede et al., 2003; 2006). Like the Dobson Umkehr algorithm, the ratio of measured intensities centred at ~310 nm (strong ozone absorption) and ~326 nm (weak ozone absorption) can be combined to create an Umkehr curve. A somewhat shorter range of SZAs (70 - 92.5 degrees) is selected in the Brewer measurement schedule.

Like FTIR measurements, the information content of the measured Umkehr curve is analyzed by the optimal estimation approach (Rodgers, 2000) where a priori information about the vertical distribution of ozone and its natural variability are used for the iterative solution expected under pre-determined conditions. Figure S-3 shows an example of the AK for the Dobson (black) and Brewer (red) middle latitude climatological ozone profile with 325 DU column value. The AK shows the sensitivity of the retrieval to ozone vertical information. The AK plot shows that information in the lowermost layers is sensed at ~60-70 %, while the rest of the information comes from the adjacent layers. It also explains how the algorithm applies vertical smoothing to the retrieved ozone profile. Unfortunately this implies that for the tropospheric retrieval (Layers 1+0), 30-40% of the information comes from the stratosphere.

The errors of ozone retrieval can be attributed to the accuracy of the forward model (i.e. multiple scattering tables, temperature profile, spectroscopic dataset, SO_2 interference, quadratic spline-fit of measurement, aerosol interference), accuracy of the measurement and smoothing mechanism of the optimal statistical retrieval method. An example of Umkehr ozone profile errors (Figure S-3) is shown as function of altitude for Dobson (green solid line) and Brewer (red solid line). Also shown separately are errors associated with the measurement (dotted) and smoothing (dashed) uncertainties that contribute to the ozone profile retrieval error. Climatological ozone variability is shown as black line and is taken from the MLL climatology (McPeters et al, 2007) for Northern middle latitude (40-45 N) and for the month of January. The covariance matrix used in the smoothing error calculations of the Umkehr retrieval is representative of the ozone variability over the middle Northern latitudes based on observations by SAGE II (ozone ~above 25 km) satellite instrument and ozone sonde (ozone below ~25 km) profiles collected between 1988 and 2002.

Recent analyses of errors (Orphal, 2003; Evans et al, 2009; Petropavlovskikh et al, 2009; 2011) suggest that mean changes in both the Umkehr Dobson and Brewer retrieved profiles due to spectroscopic dataset choice are within ±2 %. The largest difference in the retrieved ozone profile is associated with uncertainties in the stray light (out of band) contribution to Umkehr measurements. It can cause ozone to be overestimated by 5% in the troposphere.

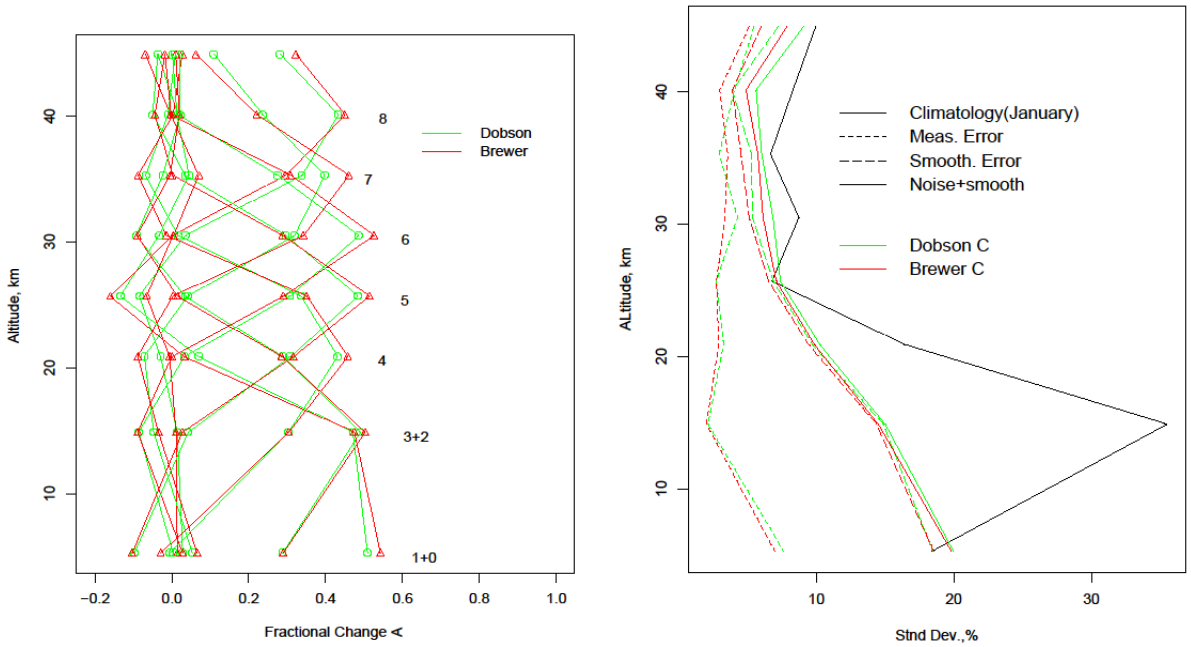

Figure S-3: (left) Umkehr averaging kernels. (right) Uncertainties for standard Umkehr levels.

Text S-10. Analysis of ozonesonde intercomparison results

ECC ozonesondes have a precision of 3-5% and a total (random + systematic) uncertainty of about 10% throughout most of the profile below ~28 km (*Smit et al.*, 2007; *Kerr et al*, 1994; *Deshler et al.*, 2008a; *Liu et al.*, 2009). The precision of the Brewer-Mast sonde is somewhat poorer, at about 10%, while the Japanese KC series sondes show a precision of about 5% (*Kerr et al.*, 1994; *Smit and Kley*, 1998). The Indian sonde and the now-obsolete GDR sonde show significantly larger variability in tropospheric response than other sonde types [*Kerr et al*, 1994; *Smit and Kley*, 1998; *Attmannspacher and Dütsch*, 1970; 1981; *Feister et al.*, 1985]. Figure S-4 shows estimated and measured error characteristics of ECC and BM sondes (Tarasick et al., 2016).

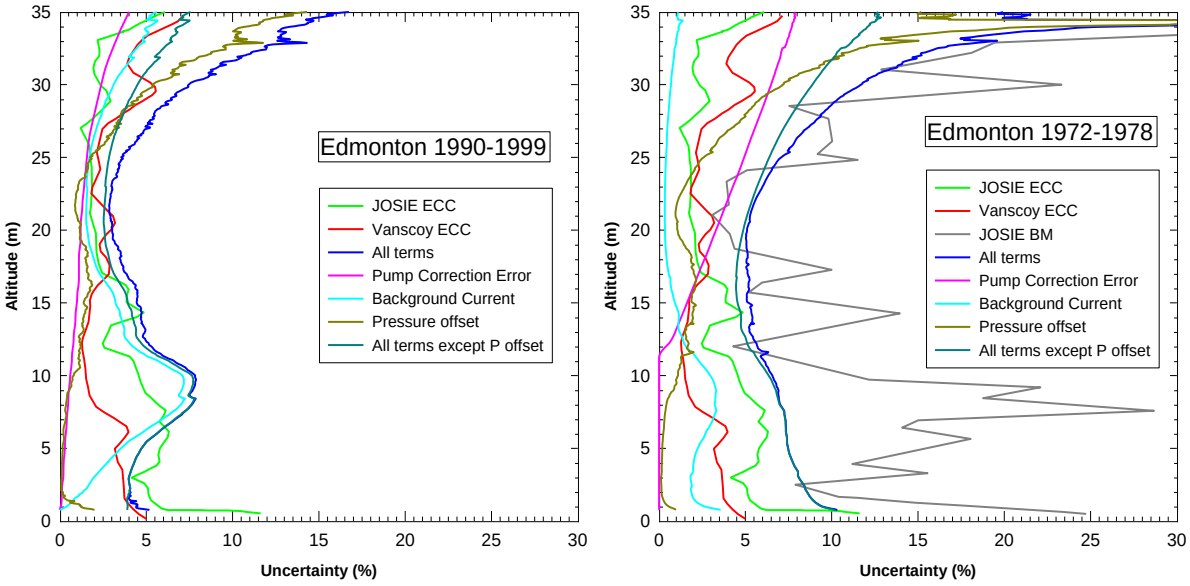

Figure S-4: Left: Average estimated uncertainty of ECC (4A and 5A) soundings in the 1990s at a Canadian site, showing contributions from selected sources. The total uncertainty without the contribution from radiosonde pressure offsets is also shown, to facilitate comparison with the JOSIE and Vanscoy intercomparison uncertainty estimates, which were referenced to a common pressure measurement. The uncertainty in the troposphere for ECC sondes is dominated by the uncertainty in the background current, which is related to uncertainty in the stoichiometry of the KI reaction. Right: Average estimated uncertainty of Brewer-Mast soundings at a Canadian site in the 1970s, showing contributions from selected sources. The Brewer-Mast sondes showed somewhat variable response in the troposphere.

Table S-4: Ozonesonde Intercomparison results 1960-2015. BM sondes. "Bias" is the mean difference from the reference (typically the multi-sonde mean in early intercomparisons), and σ is the precision (one standard deviation) of the measurement.

Brewer-Mast

Name	Date	LT			UT			Ref.	Citation	Notes
		Bias		σ	Bias		σ			
WMO1	1970	-13%	±22%	±15%	-22%	±11%	±27%	Sonde mean	Attmannspacher and Dütsch (1970)	
WMO2	1978	-3%	±6%		-4%	±10%		Sonde mean	Attmannspacher and Dütsch (1980)	
BOIC	1983-84	-20%	±10%	±10%	-10%	±10%	±10%	ECC	Hilsenrath et al. (1986)	
	1965-91	-22%	±5%		-18%	±8%		SAGE, ECC	Lehmann (2005)	Relevant to Australian & Canadian SOPs
WMO3	1991	12%	±10%	±15%	20%	±10%	±20%	ECC	Kerr et al. (1994)	
OHP1	1989	-7%*	±3%*	±3%	-11%	±9%	±3%	Lidar	Beekmann et al. (1994)	
OHP2	1991				2%	±3%	±9%	UV	Beekmann et al. (1995)	
JOSIE	1996	-3%	±10%	±10%	-3%	±8%	±8%	UV	Smit and Kley (1998)	Chamber expt.
	1996-98	12%	±16%		11%	±16%		EN-SCI, 0.5% KI	De Backer et al. (1998)	26 dual flights
OZEX SONDEX	1998-02 1996	-2% -5%	±6% ±6%		-4% -7%	±6% ±8%		ENSCI, 1% KI EN/SP, 1% KI	Stübi et al. (2008)	111 dual flights 29 dual flights

***relative to ECC sonde**

Table S-5: Ozonesonde Intercomparison results 1960-2015. ECC sondes. "Bias" is the mean difference from the reference (typically the multi-sonde mean in early intercomparisons), and σ is the precision (one standard deviation) of the measurement.

ECC

Name	Date	LT			UT			Ref.	Citation	Notes
		Bias		σ	Bias		σ			
WMO1	1970	7%	±20%	±16%	7%	±22%	±12%	Sonde mean	Attmannspacher and Dütsch (1970)	
WMO2	1978	10%	±10%		7%	±12%		Sonde mean	Attmannspacher and Dütsch (1980)	
	1975-77	-4%	±6%					Dasibi	Torres and Bandy (1978)	100 lab calibrations
	1984	8-14%		6-10%				UV	Barnes et al. (1985)	chamber expt. 2 sondes 1.5% KI soln +6-8% bias?
	1984	21%		2%	3%		5%	UV	Komhyr et al. (1985)	BOIC data
BOIC	1983-84	20%	±10%	±7%	20%	±10%	±7%	UV	Hilsenrath et al. (1986)	
WMO3		n/a		±6%	n/a		±6%	ECC	Kerr et al. (1994)	
OHP1	1989			±3%	12%	±8%	±3%	Lidar	Beekmann et al. (1994)	
OHP2	1991				28%	±4%	±11%	UV	Beekmann et al. (1995)	
JOSIE	1996	0%	±4%	±4%	0%	±3%	±3%	UV	Smit and Kley (1998)	Chamber expt.
JOSIE	2000	-1%	±4%	±4%	2%	±6	±6%	UV	Smit and Sträter (2000)	Chamber expt.
STOIC	1989	6%	±2%	±8% (both)	3%	±3%	±8% (both)	5A 1% KI	Komhyr et al. (1995)	NASA (4A,1.5% KI) vs NOAA (5A,1%)
BESOS	2004	5%	±10%	±4%	14%	±15%	±2%	UV	Deshler et al. (2008)	ENSCI 0.5% KI/SP 1% KI UV photometer problem
	1996-97	8%	±3%		6%	±3%		EN-SCI 0.5% KI	Boyd et al. (1998)	Dual ECC flights of ENSCI @1% KI
	1996	4%	±3%		2%	±3%		EN-SCI 0.5% KI	Johnson et al. (2002)	Dual ECC flights of ENSCI @1% KI
	1993			±4%			±4%		Reid *et al.* (1996)	1 flight, Bendix comparison

Table S-6: Ozonesonde intercomparison results 1960-2015. KC sondes. "Bias" is the mean difference from the reference (typically the multi-sonde mean in early intercomparisons), and σ is the precision (one standard deviation) of the measurement.

KC

Name	Date	LT			UT			Ref.	Citation	Notes
		Bias		σ	Bias		σ			
WMO1		-3%	±17%	±43%	-5%	±18%	±18%	Sonde mean	Attmannspacher and Dütsch (1970)	KC-68
WMO2	1978	-3%	±10%		-2%	±7%		Sonde mean	Attmannspacher and Dütsch (1980)	KC-68
WMO3		-5%	±6%	±6%	-4%	±6%	±7%		Kerr et al. (1994)	KC-79
JOSIE	1996	-7%	±6%	±6%	-2%	±4%	±4%	UV	Smit and Kley (1998)	Chamber expt. KC-79
JOSIE	2000	-9%	±4%	±5%	-4%	±5	±5%	UV	Smit and Sträter (2000)	Chamber expt. KC-96
	2004	~0 (KC79) -6% (KC96)	±6% ±3%		3% (KC79) -3% (KC96)	±5% ±5%			Fujimoto et al. (2004)	Revised algorithm
BESOS	2004	-10%	±5%	±2%	-10%	±5%	±5%		Deshler et al. (2008)	KC-96 Problem UV photometer
	2004-09	-10%	±10%		-8%	±10%		ECC	Morris et al. (2013)	7 sondes

Table S-7: Ozonesonde intercomparison results 1960-2015. Indian sondes. "Bias" is the mean difference from the reference (typically the multi-sonde mean in early intercomparisons), and σ is the precision (one standard deviation) of the measurement.

Indian

Name	Date	LT			UT			Ref.	Citation	Notes
		Bias		σ	Bias		σ			
WMO1		-13%	±18%		-25%	±31%		Sonde mean	Attmannspacher and Dütsch (1970)	
WMO2	1978	~10%	±10%		~10%	±10%		BM	Attmannspacher and Dütsch (1980)	pump problems
WMO3	1991	-2%	±8%	±8%	-6%	±8%	±8%		Kerr et al. (1994)	
JOSIE	1996	8%	±6%	±6%	30%	±50%	±50%	UV	Smit and Kley (1998)	Chamber expt.

Table S-8: Ozonesonde intercomparison results 1960-2015. GDR sondes. "Bias" is the mean difference from the reference (typically the multi-sonde mean in early intercomparisons), and σ is the precision (one standard deviation) of the measurement.

GDR

Name	Date	LT			UT			Ref.	Citation	Notes
		Bias		σ	Bias		σ			
WMO1		-7%	±27%		7%	±32%		Sonde mean	Attmannspacher and Dütsch (1970)	Numerous failures
WMO2	1978	-4%	±10%		2%	±12%		Sonde mean	Attmannspacher and Dütsch (1980)	Numerous failures
	1982			12%			11%		Feister et al. (1985)	15 dual flights

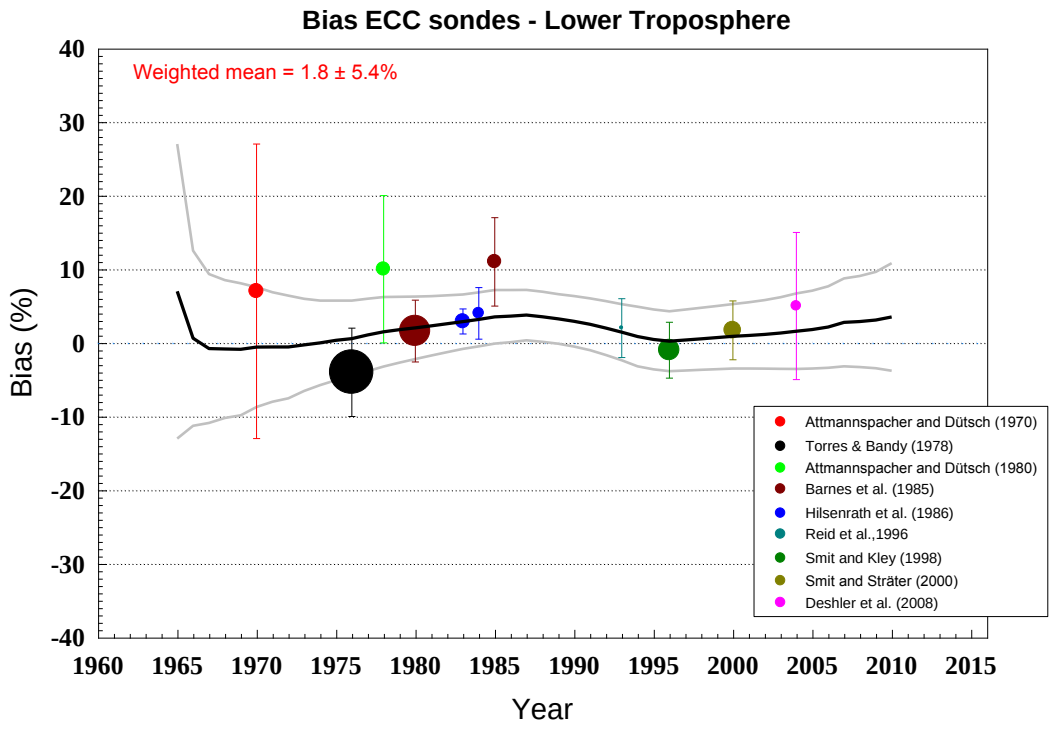

Figure S-5: Published intercomparison results of ECC sondes without adjustments. The first two WMO-sponsored field intercomparisons (Attmannspacher and Dütsch, 1970; 1978) show a high bias, but these results are relative to an average profile, as a UV reference photometer was not available. The symbol size is proportional to the standard error of the mean difference, while the error bars show standard deviations (1σ). The curves are running means (weighted by the standard error) of the mean difference and the ±1σ variability.

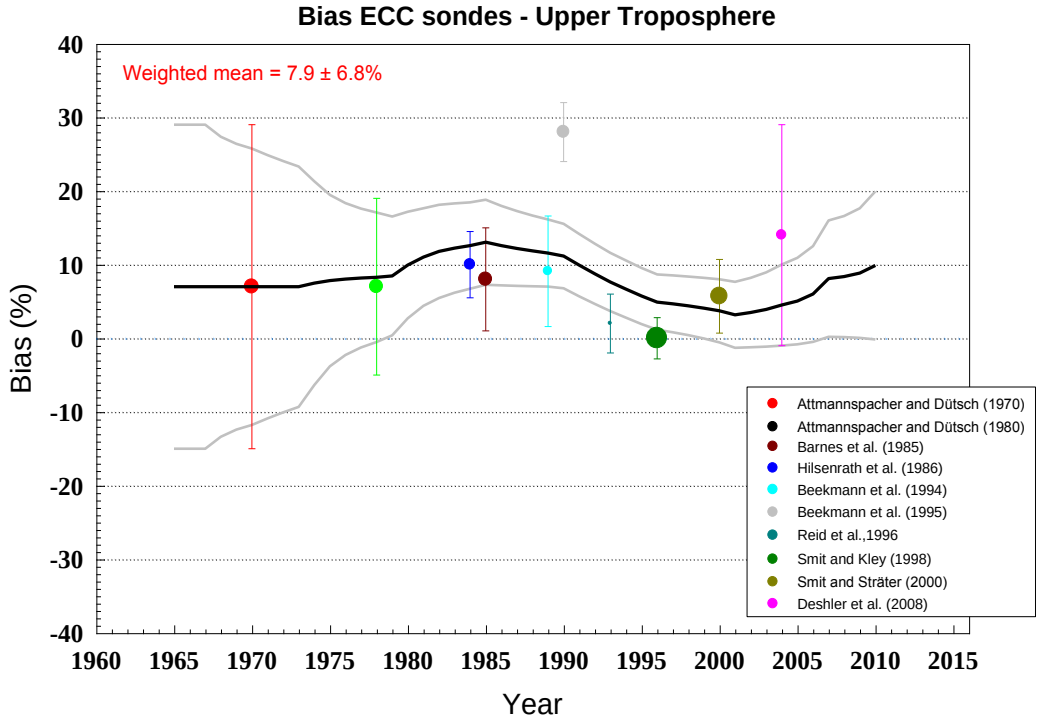

Figure S-6: As Figure S-5, for the upper troposphere.

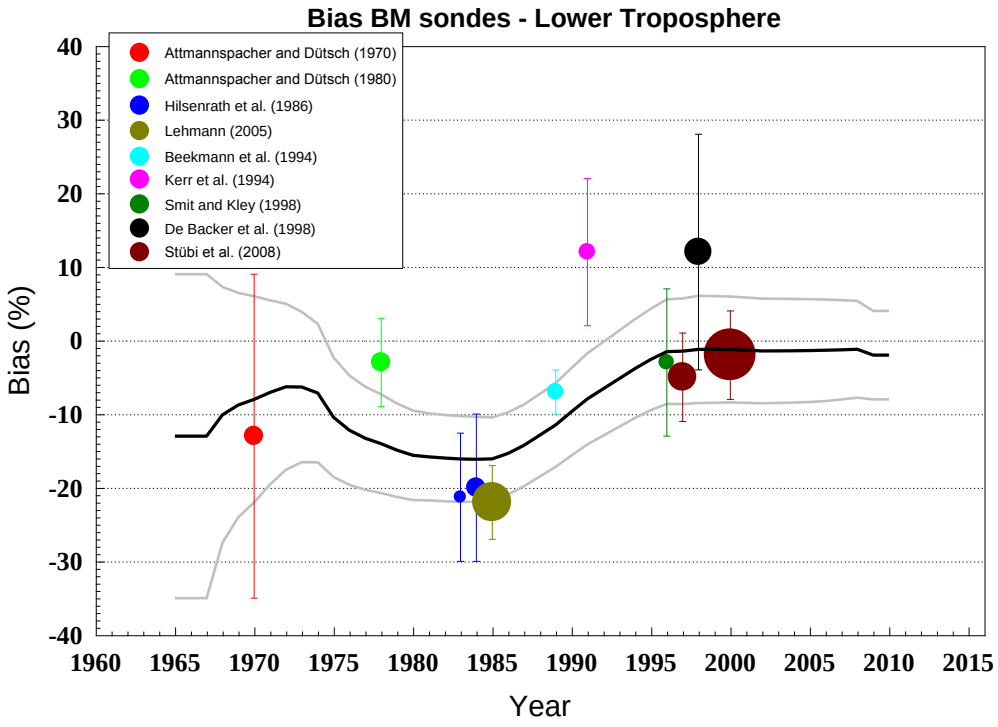

Figure S-7: Intercomparison results of BM sondes, without adjustments. Bias results are variable, but several field intercomparisons lack a UV photometer (Attmannspacher and Dütsch, 1970; 1978; Kerr et al., 1994).

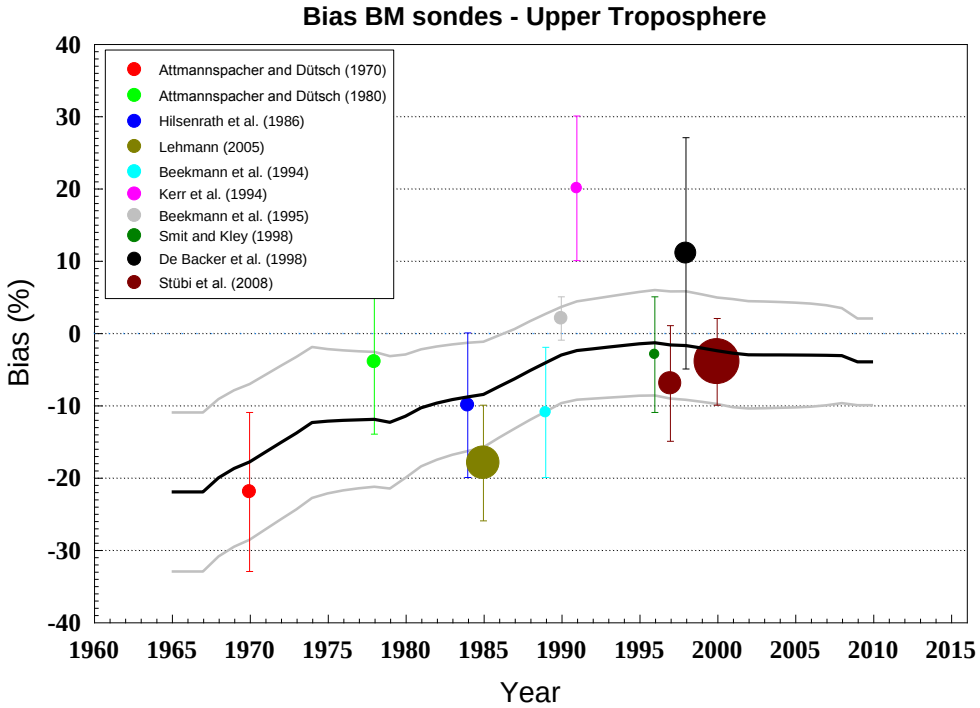

Figure S-8: As Figure S-7, for the upper troposphere.

Table S-9. Sites and characteristics existing of tropospheric ozone lidar systems.

Location	Wavelength Pair (nm)	Altitude range (km)	Regular monitoring	Time period	Reference
OHP, Southern France 44°N, 6°E	289-316	2-14	yes	1990-present	Ancellet et al., 1991
Table Mountain Observatory, CA USA 34°N, 118°W	289-299	4-23	yes	1989-present	McDermid et al., 2002
Garmisch, Germany 47°N, 11°E	277, 292, 313	0.2-15	yes	1991; 1996-2001; 2007-present	Carnuth et al., 2002; Trickl et al., 2015; 2019.
Huntsville, USA 35°N, 87°W	285-291	1-10	yes	2008-present	Kuang et al., 2011
Reunion 21°S, 56°E	289-316	3-17	yes	2013-present	
Thessaloniki, Greece	289-316	2-12	no		Papayannis et al., 2005
Boulder, USA	289-300	0.4-3	no		Alvarez et al. 2011
Greenbelt, USA	289-299	0.4-12	no		Sullivan et al. 2014
Saga, Japan	276, 287, 292	1-17	no		Uchino et al., 2014

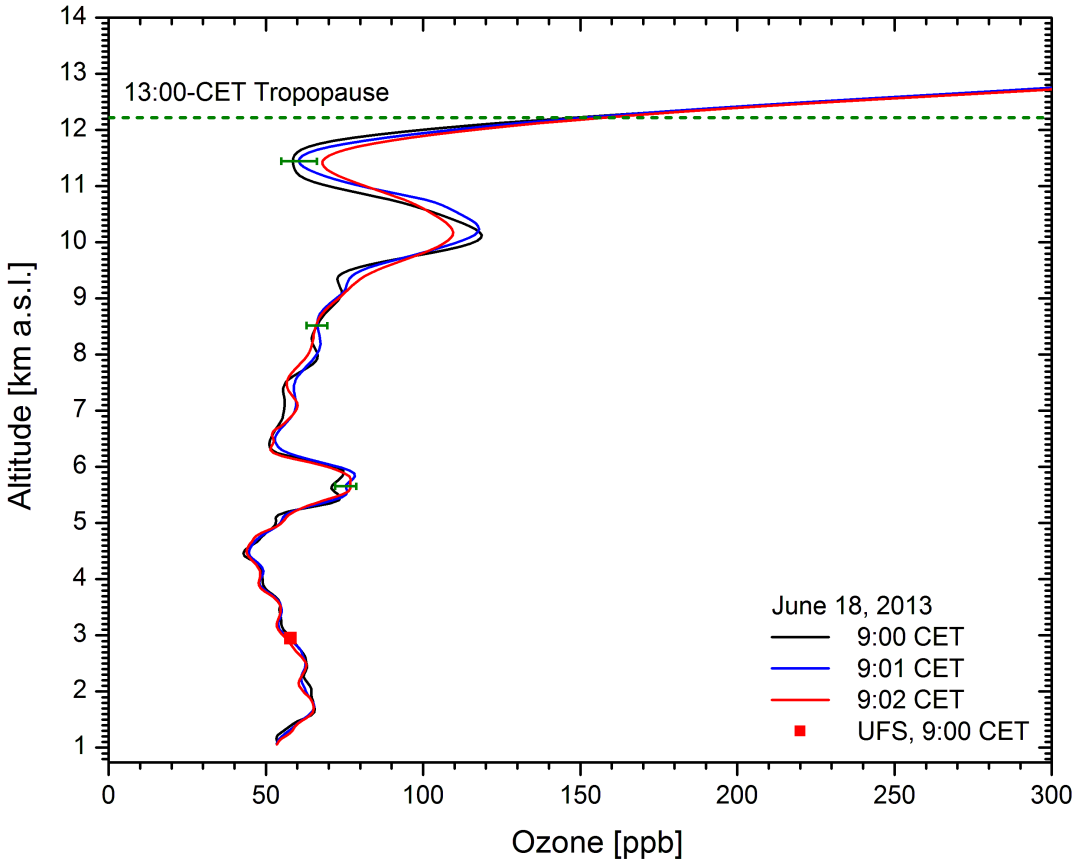

Figure S-9: Reproducibility test of the ozone DIAL at Garmisch-Partenkirchen (German Alps, at 740 m) at intervals of 1 min, under conditions of moderately elevated ozone. Due to the short period the influence of atmospheric variability can be excluded. Up to 5.8 km the ozone profile for the 277-nm − 313-nm wavelength pair was taken, above this that for 292 nm − 313 nm. For comparison we show the 9:00-CET half-hour average at the Schneefernerhaus high-altitude station (UFS) (Ludwig Ries, German Federal Environmental Agency, personal communication), but at the altitude of the Zugspitze ridge above the station that lifts the air towards the lidar in the valley behind the mountain. Until 2012 in-situ measurements were also carried out at the Zuspitze and Wank summits (at 2962 m and 1780 m, respectively) and used for comparisons with the lidar. The thermal tropopause is taken from the Munich radiosonde that is launched at a distance of about 100 km from the lidar.

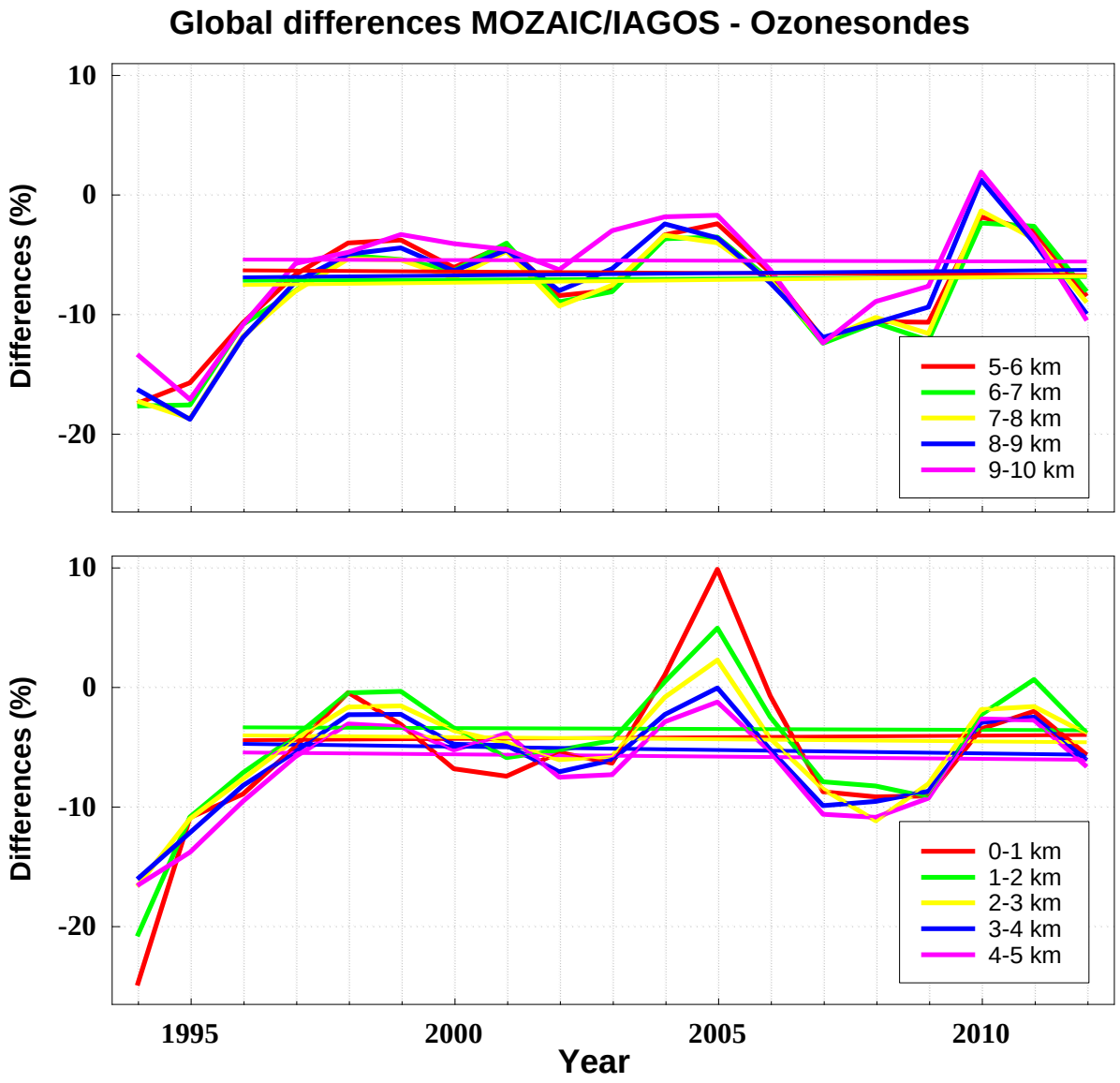

Figure S-10: Average (1994-2012) relative differences (%) between trajectory-mapped MOZAIC/IAGOS profile data and trajectory-mapped ozonesonde data, averaged over latitude. The year-by-year comparison shows considerable variability (almost certainly due to sampling differences) but no overall trend if the first two years, 1994-95, are excluded. The apparent high bias of the sondes is reduced to ~4% and ~7% if 1994-95 are excluded. These large differences in the first two years of the MOZAIC/IAGOS record may be due to sampling differences as well. They are consistent with results by Logan et al. (2012) and Staufer et al. (2014), the latter also employing a trajectory-mapping analysis.

Text S-11. FTIR retrieval and information content

Early ground-based observations of the infrared terrestrial atmosphere were published in solar spectral surveys such as Migeotte et al. (1956), and the earliest attempt to determine the vertical distribution of ozone from IR measurements was made by Walshaw and Goody (1956). Later quantitative studies of ozone from IR measurements and inter-comparisons focusing on ozone with various methods including Dobson and Brewer spectrometers, satellite measurements and ECC (Electro Chemical Cell) sondes were conducted at several stations: Mauna Loa Observatory (David et al., 1993), Kitt Peak (Pougatchev et al., 1995) and Lauder (Rinsland et al., 1996; Pougatchev et al., 1996). These reflect measurements that predate or as part of the early years of the NDACC (Network for the Detection of Atmospheric Composition change, www.ndacc.org, Kurylo, 1991; Kurylo and Zander, 2001) observational record.

Within the NDACC every attempt is made to maintain a consistent data product across all stations. This is done by direct open and blind inter-comparison of multiple instruments (eg. Goldman et al., 1999, Paton-Walsh et al., 1997), inter-comparison of the two commonly used retrieval code sets SFIT2/4 and PROFFIT (Hase et al., 2004), exchanges of calibration cells (Coffey et al., 1998, Hase et al., 1999, Hase et al., 2012) among different groups.

Tropospheric FTIR ozone measurements that are described and used for trends in this report are derived from ground based mid-infrared spectra of the sun. Current (since ~ early-mid 1990's) instrumentation are both high resolution (0.004 cm^{-1}) and broad band (750 - 4500 cm^{-1}) hence retrieval of ozone from these spectra hold a lot in common with the retrieval of many other species. The instruments are solar viewing and susceptible to poor weather conditions. Consequently, limited seasonal observations are made at high latitudes. At most sites, several observations are taken per day given the weather. A single observation for ozone takes between 2-6 minutes with a typical resultant signal to noise ratio (SNR) of between 300 and 500 in the 10µm spectral region. These observations may be performed manually on site, remotely via internet connection or by autonomous operation (Neefs et al., 2007; Hannigan et al., 2009) depending on logistical access to the observation site.

The FTIR measurement records the spectrum along the line of sight (LOS) to the sun. In the retrieval process, mass paths on a fine altitude-pressure grid are raytraced along the LOS and zenith. The modeled spectrum is varied by adjusting the ozone volume mixing ratios (VMR) during the process of fitting the pressure-broadened and temperature sensitive ozone spectral absorption features. The zenith mass paths are integrated with the retrieved VMR profile to determine a total and partial (tropospheric and stratospheric) ozone column amount.

The retrieval methodology is an implementation of the Optimal Estimation (OE) theory (Rodgers, 2000) and requires a priori data for the atmospheric state and other forward model parameters. Indeed in OE, the retrieved state vector $\hat{x}$ (in our case the ozone vertical mixing ratio (VMR) and other fitted parameters e.g. scaling factors for interfering species VMR profiles or the baseline of the spectral window) is linked to the real state $\boldsymbol{x}$ by:

$$\hat{x} = \boldsymbol{x_a} + \mathbf{A}\,(\boldsymbol{x} - \boldsymbol{x_a}) + \text{error terms}, \qquad (1)$$

with $\boldsymbol{x_a}$ the a priori state vector, and $\mathbf{A}$ the averaging kernel matrix which represents the sensitivity of the measured state of the atmosphere to the real state. The trace of A is the degrees of freedom for signal (DOFS) which is the number of independent pieces of information available from the measurements on the retrieved state. While the retrieved state vector is

calculated on the fine grid of about 40 to 50 layers, the typical DOFS of retrieved ozone profiles are between 4 - 5 depending on the site location, the SNR of the spectra, and the solar zenith angle. Figure S-11 shows an example of averaging kernels (rows of **A**) obtained in the retrieval grid (left), and of the partial column averaging kernels obtained for four layers containing at least one DOFS (right). They correspond roughly to one layer in the whole troposphere (red curve), and three in the stratosphere. Further details about the information content on ozone can be found in Schneider et al. (2008) and Vigouroux et al. (2015), and in Wespes et al. (2012) focusing on tropospheric ozone.

Among the FTIR stations measuring ozone (a complete list can be found at https://www2.acom.ucar.edu/irwg), some can only provide total columns or profiles with too few DOFS in the troposphere to be informative in this report. We summarize in Table S-10 the stations that obtain at least 0.8 DOFS in the ground - 8 km layer. Among these stations, some have long enough time-series to be relevant for ozone trends studies as done in Vigouroux et al. (2008, 2015) or García et al. (2012), and they are used in *TOAR-Climate* of the present report. Except for the two Japanese stations, the measurements are still ongoing and being archived in NDACC within one year at the latest. Data quality cut off is determined by site and is informed by a goodness of fit criterion, spectral SNR, minimum DOFS or maximum total error.

The theoretical uncertainty estimation on the FTIR ozone data is also based on Rodgers (2000), where the difference between the retrieved state $\hat{x}$ and the real state $\boldsymbol{x}$ can be written as a linear combination of the a priori state $\boldsymbol{x_a}$, the real and estimated model parameters b and $\hat{b}$, respectively, and the measurement noise ε:

$$(\hat{x}-x) = (A-I)(x-x_a) + GK_b(b-\hat{b}) + G\varepsilon \qquad (2)$$

where $\boldsymbol{I}$ represents the Identity matrix, $\boldsymbol{G}$ the gain matrix, and K_b the sensitivity matrix to model parameters. Thereby, equation (2) defines three types of error. The first term is the smoothing error associated with the limited vertical sensitivity of the FTIR remote sensing technique, and is calculated as $(A-I)S_a(A-I)^T$, where S_a is the assumed a priori covariance matrix of the ozone vertical profile (i.e. its natural variability). The second one represents the errors due to uncertainties in the forward model parameters (temperature, instrumental characteristics, spectroscopic data, …) and is calculated as $(GK_b)S_b(GK_b)^T$, where S_b is the covariance matrix of parameter *b*. The third one corresponds to the measurement noise $GS_\varepsilon G^T$, where S_ε is the measurement noise covariance matrix, with diagonal terms equal to $1/\mathrm{SNR}^2$.

The uncertainties budget depends strongly on the S_a, S_b and S_ε matrices, therefore on station to station ozone natural variability (S_a), goodness of the spectra (S_ε), but also on the estimation of the S_b matrix. On-going projects are currently taking place on a harmonization of these uncertainties among NDACC stations.

The most recent and thorough ozone uncertainty budget and inter-comparisons were performed at the Izaña Atmospheric Observatory (IZO, www.izana.aemet.es). We give here as an example that uncertainty budget as it has been very-well documented and validated with correlated data available at IZO (Schneider and Hase (2008), Schneider et al. (2008) and García et al. (2012)). Details on the determination of S_a, S_b and S_ε matrices at IZO can be found in Schneider and Hase (2008) and García et al. (2012).

The straightforward comparison of ozone partial columns from FTIR and ECC sondes is rather satisfactory, as shown by Schneider et al. (2008) and García et al. (2012). We found a mean bias (FTIR-ECC) of about 4% in troposphere (2.37–13 km), while the scatter (one standard deviation, 1σ, of the relative differences) was about 9% when comparing the coincident FTIR and ECC time series between 1999 and 2010. These results rather agree with the expected uncertainty for the ECC sondes given by Smit et al. (2007) (±5–10%) and with the FTIR theoretical error estimation previously shown. Furthermore, they are in good agreement with other inter-comparison studies using ECC sonde data (Clerbaux et al., 2009; Nair et al., 2011 and references therein). Part of the discrepancy observed between FTIR and ECC sonde measurements is due to the smoothing error (recall Table 10 in the main text). Smoothing the ECC sonde profiles with the FTIR averaging kernels improves the agreement by reducing the scatter, on average, to 7 % for the tropospheric 2.37–13 km partial column. Other sources of discrepancy might be errors in the FTIR and/or ECC data and the observation of different air masses by the FTIR experiment, on the one hand, and ECC experiment, on the other (Schneider et al., 2008). The systematic discrepancies might be attributable to systematic ECC and FTIR errors caused, for instance, by uncertainties in the spectroscopic line parameters.

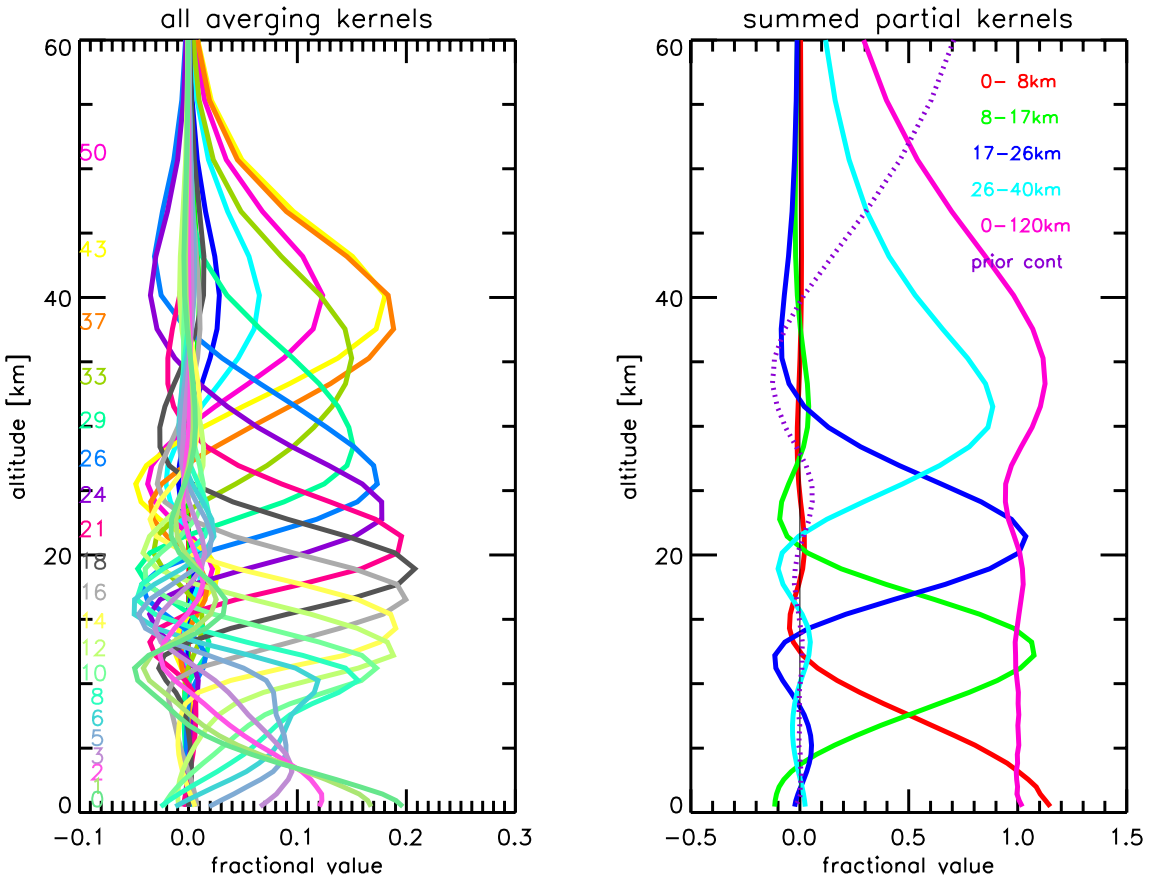

Figure S-11: Ozone averaging kernels at a typical NDACC site, Thule, GR, on the retrieval grid (left), and for layers containing at least one DOFS (red, green, blue, cyan). The total column averaging kernel is also shown (magenta) and a priori content in dotted magenta.

Table S-10. Sites and characteristics existing of tropospheric FTIR stations.

Stations	Time-series period with relevant DOFS	Ground to top of tropospheric layer (8km)	Mean DOFS in the layer	Used in TOAR-Climate
Eureka 80°N, 86°W	2006-2016	0.6-8.0 km	0.95	No
Ny-Alesund 79°N, 12°E	1995-2012	0.0-8.0 km	1.00	Yes
Thule 77°N, 69°W	1999-2014	0.2-8.0 km	1.28	Yes
Kiruna 68°N, 20°E	1996-2012	0.4-8.0 km	0.92	Yes
Harestua 60°N, 11°E	1995-2012	0.6-8.6 km	0.82	No
St Petersburg 60ºN 30ºE	2009-2016	0.0-8.0 km	0.83	No
Bremen 53ºN 9ºE	2004-2016	0.0-8.0 km	0.8	No
Zugspitze 47ºN 11ºW	1995-2016	2.95-8.0 km	0.61	No
Jungfraujoch 47°N, 8°E	1995-2014	3.6-8.0 km	0.87	Yes
Moshiri 44°N, 142°E	1996-2007	0.3-8.0 km	0.78	No
Rikubetsu 44°N, 144°E	1995-2009	0.4-8.0 km	0.78	No
Toronto 44ºB 281ºE	2002-2016	0.2-8.0 km	0.83	No
Izaña 28°N, 16°E	1999-2014	2.4-8 km	0.78	Yes
Mauna-Loa 20°N, 156°W	1995-2014	3.4-8.0 km	1.01	No
Altzomoni 19ºN 261ºE	2012-2016	4.0-8.0 km	0.67	No
Reunion Island 21°S, 55°E	2004-2011 (Saint-Denis)	0.1-8.0 km	0.97	No
Reunion Island 21°S, 55°E	2013-2014 (Maïdo)	2.2-8.0 km	0.91	No
Wollongong 34°S, 151°E	1996-2014	0.0-8.0 km	0.98	Yes
Lauder 45°S, 170°E	2001-2012	0.4-8.0 km	1.04	Yes
Arrival Heights 78°S, 167°E	1997-2012	0.2-8.0 km	0.94	Yes

Text S-12. Tropospheric ozone satellite and residual measurements: Instrument and method details

OMI: The Ozone Monitoring Instrument on Aura is a UV/VIS nadir-viewing solar backscatter spectrometer that makes daily measurements of Earth radiances and solar irradiances from 270 to 500 nm with a spectral resolution of approximately 0.5 nm (Levelt et al., 2006). The UV channel consists of two sub-channels: The UV-1 with the range 270–310 nm and UV-2 that covers 310–365 nm. Retrievals of column amounts of ozone and other trace gases as well as cloud and aerosol properties are provided by observations in these channels. The OMI measurements provide daily (~1:45 pm local time) near-global coverage with a footprint of 13 km×24 km. The OMI data record begins October 2004 and extends to present time. However, since January 2009 OMI data have been affected by a serious "row anomaly", which has affected the quality of retrieved profiles (Huang et al., 2017; 2018). OMI data products including ozone along with documentation can be obtained from the NASA Goddard website http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/.

MLS: The Aura Microwave Limb Sounder instrument is a thermal-emission limb sounder that obtains profiles by scanning vertically. As MLS looks forward along the orbital path, an MLS profile is obtained about 7 min before OMI. Ozone profiles from the MLS v3.3/3.4 and v4.2 retrievals are given on pressure levels between 261 hPa and 0.0215 hPa. Horizontal resolution is ~300 km. The instrument provides ozone profiles during both daytime (ascending orbit) and nighttime (descending orbit). An extensive description of the v3.3 ozone profile product is found in Livesey et al. (2011). MLS data are available at http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/.

TOMS and SBUV: The Total Ozone Mapping Spectrometer (1978-2005) and Solar Backscatter UltraViolet radiometer are UV nadir-measuring instrument series using similar algorithms. TOMS scanned horizontally and provided near-global coverage of total column ozone each day. SBUV measures both total ozone and stratospheric ozone profiles along the nadir orbital track. As a result, the SBUV measurements are separated along longitude by about 25^{o}. The footprint of the TOMS scanning measurements is about 50×50 km. The SBUV nadir footprint is about 200×200 km. The TOMS record extends to 2005. The long SBUV record continues to the present time, from SBUV instruments onboard multiple NOAA satellites. TOMS and SBUV merged data products are available at http://acd-ext.gsfc.nasa.gov/Data_services/merged/.

SAGE: The Stratosphere Aerosols and Gas Experiment (1979-2006) is a solar occultation grating spectrometer measuring UV, VIS, and IR radiances. As an occultation instrument daily coverage is limited to ~30 profiles in narrow latitude bands. Time coverage of the existing SAGE ozone profiles extends from 1979 to 2006 from combined SAGE, SAGEII, and Meteor-3M SAGE III measurements. SAGE data are available at http://disc.sci.gsfc.nasa.gov/.

GOME: The Global Ozone Monitoring Experiment (1995-2011) on board the ESA ERS-2 satellite measured backscattered radiance from 240–790 nm with spectral resolution of 0.2–0.4 nm in the Hartley (200–320 nm) and Huggins (320–350 nm) bands (Burrows et al., 1999) and a footprint of 40x320 km^2 (but 80x960 km^2 in the shorter UV region due to weaker signal and longer interation time). GOME data are available at https://earth.esa.int/web/guest/-/gome-total-colomn-amount-of-trace-gases-product-1484.

GOME-2 is a similar instrument on both MetOp-A and MetOp-B, with a spectral resolution of 0.26-0.51 nm, spectral sampling of 0.11–0.22 nm, a 40x40 km footprint (but 40x640 km^2 in the

shorter UV region due to weaker signal and longer interation time) and ~1 DFS in the troposphere (Miles et al., 2015; van Oss et al., 2015). Near real time and off-line operational GOME-2 data are available at http://acsaf.org/offline_access.html, https://acsaf.org/products/nhp.html.

TES: The Tropospheric Emission Spectrometer (Beer et al., 2001) on EOS/Aura is an FTIR interferometer that uses the 9.6 micron ozone absorption band to retrieve ozone concentrations. Its 0.1 cm-1 spectral resolution is sufficiently fine to distinguish the pressure-broadening of ozone absorption lines at atmospheric pressures in the lower troposphere, giving vertical information to discriminate tropospheric and stratospheric ozone. Like ground-based FTIR, it uses OE algorithms to retrieve vertical profiles of ozone concentration (Bowman et al., 2006). It is mostly sensitive to tropospheric ozone between 700 and 300 hPa, owing to the higher thermal contrast there with respect to the surface. Relative to the UV instruments discussed above it has a small footprint for nadir observations (5x8 km). It also has a better ability see through clouds (Kulawik et al., 2006). TES data are available at https://tes.jpl.nasa.gov/data/.

IASI: the Infrared Atmospheric Sounding Interferometer (Clerbaux et al., 2009) on MetOP-A & B is an FTIR interferometer with 0.5 cm-1 spectral resolution in the 9.6 micron band. While TES a priori are from a climatology that varies seasonally and spatially, IASI (processed via FORLI) uses a single global average (Hurtmans et al., 2012). This means that IASI profiles may have too little tropospheric ozone in the Northern Hemisphere and too much ozone in the Southern Hemisphere, but allows more direct comparisons of ozone distributions. Table S-12 shows a comparison of TES and IASI specifications. IASI data are available at http://www.eumetsat.int/website/home/Data/DataDelivery/EUMETSATDataCentre/index.html. IASI+GOME2 daily global observations are available at (http://www.aeris-data.fr).

TOR residual methods: The original TOMS/SAGE/SBUV tropospheric residual techniques used daily stratospheric column ozone maps estimated from either SAGE or SBUV stratospheric ozone profiles. Later versions of the TOR technique used ozonesonde measurements to adjust the SBUV retrievals [Fishman et al., 2003]. The technique suffers to some extent from the necessity of assuming that the stratospheric ozone column has a simple, well-defined zonal structure. For this reason its usefulness is in general restricted to tropical latitudes where stratospheric variability is typically small. At mid-latitudes, the TOR method suffers from uncertainty in the lower-stratospheric ozone amounts resulting from limited sampling and limited accuracy of the lower stratospheric interpolation. Temporal coverage is 1979-2005.

OMI/MLS residual: The OMI/MLS residual method (Ziemke et al., 2006) subtracts MLS SCO from OMI total ozone for near clear-sky scenes (i.e., OMI radiative cloud fractions less than 30%) to derive gridded tropospheric ozone residuals at 1°×1.25° resolution. SCO is first calculated along orbit paths using standard vertical pressure integration of MLS ozone VMR profiles from tropopause pressure to 0.0215 hPa. Tropopause pressure is determined from NCEP analyses using the WMO 2K-km^{-1} lapse-rate definition. The SCO measurements for each day are subsequently interpolated horizontally (Gaussian + linear) between orbit paths to obtain gridded SCO fields at the 1°×1.25° horizontal resolution of OMI. These SCO fields are then subtracted from the gridded OMI total ozone to derive daily gridded TCO fields. An additional tropospheric ozone data product, of mean VMR (in ppbv), is calculated each day from TCO (in DU) by the relation 1270×TCO/ΔP, where ΔP is the difference of surface pressure minus tropopause pressure in units hPa. Both TCO and VMR daily measurements have been averaged monthly for October 2004 through December 2014. On average, 30 DU in TCO corresponds to

about 42 ppbv in VMR. Precision uncertainty for daily TCO and VMR at 1°×1.25° gridding is approximately 5 DU and 7 ppbv, respectively. For monthly means at the same gridding the precision numbers are about 1 DU and 1.3 ppbv, respectively. Ozone mass in the troposphere can be also calculated by area integration of these gridded TCO measurements. The OMI/MLS TCO and mean VMR products are obtainable at http://acd-ext.gsfc.nasa.gov/Data_services/cloud_slice/.

OMI/MLS Trajectory mapping: The TRAJ method (Schoeberl et al., 2007) uses a trajectory-mapping algorithm to advect v3.3 ozone from MLS two days forward and backward to produce detailed fields of stratospheric ozone profiles. These daily ozone profiles are then vertically integrated to produce high-resolution (1° × 1.25°) gridded maps of SCO. (The current TRAJ data product includes gridded SCO but does not include gridded stratospheric ozone profiles.) The gridded SCO is then subtracted from coincident total ozone from OMI. This method produces daily global maps that include regions associated with the jet streams. Vertical resolution of MLS-retrieved ozone near the tropopause is about 3 km which may impact the accuracy of derived column amounts by several DU. The fields produced using this method incorporate NASA's Modern Era Retrospective ReAnalysis (MERRA) assimilated winds. TRAJ data are available at ftp://jwocky.gsfc.nasa.gov/pub/ccd/traj/.

Data assimilation of OMI/MLS: The Global Modeling and Assimilation Office (GMAO) assimilated ozone product is generated by ingesting OMI v8.5 total column ozone and MLS v3.3 ozone profiles into the Goddard Earth Observing System version 5.7.2 (GEOS-5.7.2) assimilation system with several modifications listed below. GEOS-5.7.2 is an updated version of the global data assimilation system used to produce the MERRA reanalysis, described by Rienecker et al. [2011]. It consists of an Atmospheric General Circulation Model (AGCM) and a statistical analysis module that combines 6-hourly meteorological and chemical forecasts from the AGCM with observational data from meteorology and satellite-borne sounders to produce an analysis state. The assimilation system does not include ozonesonde measurements. The AGCM is a version of the Fortuna general circulation model described in Molod et al. [2012]. The statistical analysis is done using the Gridpoint Statistical Interpolation (GSI) approach [Wu et al., 2002; Purser et al., 2003a, 2003b]. The assimilated product was generated at a horizontal resolution of 2° latitude by 2.5° longitude and at 72 vertical layers between the surface and 0.01 hPa. Depending on the height of the tropopause, about 25 to 35 of those layers are in the troposphere. The assimilated tropospheric ozone product can be obtained from the NASA website http://avdc.gsfc.nasa.gov/index.php?site=1093970695.

Cloud slicing methods: The convective-cloud differential (CCD) method subtracts SCO over deep convective clouds from nearby clear-sky total ozone (Ziemke et al., 1998). The profile cloud slicing method combines above-cloud column ozone with in situ cloud pressures to measure ozone concentrations in clouds (Ziemke et al., 2001; 2009). Profile cloud slicing has been used only for the Aura time record (2004-present), while the CCD product begins in 1979. Horizontal resolution of the ozone measurements is $5^{o}\times5^{o}$ and limited primarily to tropical latitudes. Temporal data sampling is monthly averages. The cloud slicing products are obtainable at http://acdext.gsfc.nasa.gov/Data_services/cloud_slice/.

The GOME/GOME-2/SCIAMACHY CCD tropospheric ozone data are available via the GOME portal at Institute of Environmental Physics (IUP) at the University of Bremen (UB) and the Ozone Climate Change Initiative (CCI) Climate Research Data Package (CRDP; http://www.esa-ozone-cci.org/?q=node/160).

OMI/GOME profile retrieval: In this method (Liu et al., 2005; 2010a,b) a profile is retrieved at 24 layers from the surface to ~60 km from backscattered UV at 270–330 nm using the optimal estimation technique [Rodgers, 2000]. Direct retrieval of tropospheric information is possible from the wavelength-dependent photon penetration and temperature-dependent spectral structure in the Huggins bands (Chance et al., 1997). Daily NCEP tropopause pressures are used to separate stratospheric ozone column from tropospheric ozone column; total, stratospheric, and tropospheric ozone columns and their retrieval errors can be derived from the retrieved profiles and corresponding retrieval error covariance matrices. The a priori is the monthly zonal-mean Labow-Logan-McPeters ozone profile climatology (McPeters et al., 2007). The retrieval has four to seven layers in the troposphere, although typically <1.2 DFS, with a vertical resolution generally about 10-14 km. Retrieval errors are dominated by smoothing errors, and vary from 6-35% in the troposphere, with a contribution of as much as 10% from random-noise errors. Total tropospheric column retrieval uncertainty is ~2-5 DU. The main sources of systematic errors are in temperature and cloud-top pressure. This algorithm has been applied to GOME and OMI data to produce the entire global data record.

The GOME and OMI algorithms differ slightly as described in Table S-11. Monthly mean GOME profile data are available at https://www.cfa.harvard.edu/~xliu/GMLV3/, and the daily OMI PROFOZ data product are available at Aura validation data center (AVDC, https://avdc.gsfc.nasa.gov/index.php?site=1389025893&id=74). Validation of the 10-year PROFOZ product during 2004-2014 is presented in Huang et al. (2017, 2018).

Table S-11. Comparison of retrieval algorithms and datasets.

	GOME 07/1995-06/2003	**OMI** From 10/2004	**OMI-RAL** From 01/2007	**OMI/MLS** 10/2004-09/2015	**IASI-FORLI** From 10/2007	**IASI-SOFRID** From 01/2007	**IASI-LISA** From 01/2007	**IASI + GOME2** From 01/2007
Spatial resolution	960 x 80 km^2	52 x 48 km^2 at nadir	50x50km (sampling 100x100km)	1°x1°	12 km pixels at nadir 50 km apart	12 km at nadir	12 km at nadir	12 km pixels at nadir spaced by 50 km
Global coverage	3-day	daily*	yes	Global minus polar night latitudes	Global twice daily	Global twice daily	Europe and Asia	daily, except for polar night
Spectral region	289-307 nm (band 1) 325-340 nm (band 2b)	269-309 (UV1) 312-330 nm (UV2)	266-307 nm (UV1), 321-338 nm (UV2)	OMI: 270-314 nm (UV1), 306-380 nm (UV2), 350-500 (VIS); MLS: mm & sub-mm wavelengths	1035-1078 cm^{-1}	980-1100 cm^{-1}	7 windows, range 975-1100 cm^{-1}	7 spectral micro-windows, range 980-1070 cm^{-1} & 2 windows, range 290-345 nm
Solar irradiance	daily	2005-2007 mean adjusted for sun-earth distance	Monthly median from same month in 2005	N/A	N/A	N/A	N/A	daily
Slit function	Derived variable asymmetric Gaussian slit f'n	Derived Gaussian slit function	Asymmetric super-Gaussian fit to solar ref. spectrum	N/A	N/A	N/A	N/A	Derived variable asymmetric Gaussian slit f'n
Wavelength calibration	Pre-calibration of radiance/ irradiance, fit 3rd order polynomial shift, each band	No pre-calibration, fit zero-order shift in each band	Solar calibrated as part of slit-function fit. Spectral mis-registration fitted in every scene.	N/A	N/A	N/A	N/A	Pre-calibration. Zero-order fit of spectral shift for each band.
Soft calibration (wavelength and scan angle dependent)	Derived from average radiance spectrum over 60°S-60°N relative to that at the beginning as function of time	Derived from comparison of observation/ simulation over tropics in 2 days and applied independent of space and time	Fixed time, scan + λ-dependent bias in UV1 range derived by comparison to simulated spectra. UV2 no soft correction	N/A	N/A	N/A	N/A	Derived from comparison observation/ simulation over the Pacific

On-line correction	Fit λ-dependent 2nd order correction in band 1	Fit zero order offset in UV1	N/A	N/A	N/A	surface temperature retrieval	N/A	No
Undersampling correction	Yes	No	Using high-res solar reference spectrum	N/A	N/A	N/A	N/A	No
Radiative transfer model	LIDORT with look-up table correction for polarization. Effective cross section (pre-convolved with slit functions) with solar-I_0 effect	VLIDORT with on-line polarization correction Interpolate radiative calculation, selected wavelengths to high-res grid & then convolve with slit functions	Modified GOMETRAN++	TOMRAD	FORLI	SOFRID/ RTTOV	KOPRA-KOPRAFIT	KOPRA for IASI and VLIDORT for GOME-2
Measurement error	Random-noise error	Random-noise error noise floor 0.4% in UV1 and 0.2% in UV2	Random based on L1, noise floor 1% (UV1), ~0.04% UV2	1.3 DU precision	Random spectral noise	Retrieval error 10-15 %	Random-noise error	Random-noise error
Other trace gases retrieved	SO_2, BrO	BrO	HCHO, BrO, NO_2	SO_2, NO_2, BrO, HCHO, OCLO, + gases from OMI)	H_2O total column	H_2O profile	H_2O profile	NO_2, SO_2 and H_2O
Ring effect scaling	2nd-order polynomial	Zero-order	Yes	N/A	N/A	N/A	N/A	Zero-order
Surface albedo	Zero-order in band 1, 2nd-order polynomial	Zero-order in band 1, 1st-order polynomial	Yes	N/A	N/A	UWIREMIS climatology	N/A	2nd order polynomial fit, Huggins band
Surface pressure	NCEP reanalysis pressure	NCEP pressure scaled with OMI terrain height	No	N/A	Numerical Weather Prediction	climatological	ERAI reanalysis	ERAI reanalysis
Aerosol fields	SAGE/ GEOS-Chem climatology	Not used	Not modelled	N/A	N/A	N/A	N/A	not directly accounted
Cloud fraction initialization	368-372 nm	345-349 nm	OMI L2	N/A	N/A	N/A	N/A	FRESCO, using O_2 A-band data (Koelemeijer et al., 2001)

Albedo database	GOME [Koelemeijer et al., 2003]	OMI [Kleipool et al., 2008]	Not used: effective scene albedo retrieved	N/A	N/A	N/A	N/A	from OMI at 347 nm (Kleipool et al. 2008)
A priori	McPeters et al. (2007)	McPeters et al. (2007)	McPeters et al. (2007)	OMI: McPeters et al. (2007)	McPeters et al. (2007)	2008 MLS, ozonesonde, MOZAIC-IAGOS	McPeters et al. (2007)	McPeters et al. (2007)
Data version used in TOAR	V2.1	V1.0	Lv2: RAL OMI fv0214	OMI: v8.5, MLS: v4.2	FORLI v20100815	SOFRID v1.5	v2.0	V2

* Lost global coverage since the occurrence of serious row anomaly in January 2009.

Table S-12. Comparison of GOME, OMI, TES and IASI specifications.

	GOME	OMI	TES	IASI
Launch Date	1996	2004	20040715	2007 (A); 2012 (B)
Orbital Elevation	780 km	705 km	705 km	817 km
Daytime overpass time	10:30	13:45	13:45	9:30
Spatial Sampling	40 km x 320 km	13 km x 25 km	8 km x 5 km	12 km at nadir
Spatial Coverage	3-day global	Global, daily	Sparse (nadir ~180 km apart), 16-day global repeat	±48° off-nadir swath, 2 x global per day
Cloud threshold	Cloud < 30%	Cloud < 30%	Retrieves both clear/cloudy	Cloud < 13%
Spectral resolution (apodized)	0.2-0.4 nm	0.4-0.6 nm	0.1 cm^{-1}	0.5 cm^{-1}
NESR (9.6 micron)	N/A	N/A	50 $nW/cm^2/sr/cm^{-1}$	20 $nW/cm^2/sr/cm^{-1}$
A priori	McPeters et al. (2007)	McPeters et al. (2007)	MOZART-4 climatology (Emmons et al., 2010)	McPeters et al. (2007)
Data version used in TOAR	V2.1	V1.0	TES L2 v006	FORLI v20100815 SOFRID v1.5 LISA v2.0

References

Alvarez, RJ, Senff, CJ, Langford, O, Weickmann, AM, Law, C, Machol, JL, Merritt, DA, Marchbanks, RD, Sandberg, P, Brewer, WA, -Hardesty, RM and **Banta, RM.** 2011. Development and application of a compact, tunable, solid-state airborne ozone lidar system for boundary layer profiling. *J. Atmos. Ocean. Technol*. **28**: 1258–1271. DOI: https://doi.org/10.1175/JTECH-D-10-05044.1

Ancellet, G, Beekmann, M and **Papayannis, A.** 1991. Impact of a cutoff low development on downward transport of ozone in the troposphere. *J. Geophys. Res.* **99**: 3451–3468. DOI: https://doi.org/10.1029/93JD02551

Attmannspacher, A and **Dütsch, HU.** 1981. Second international ozone sonde intercomparison at the Observatory Hohenpeissenberg, *Ber. Dtsch. Wetterdienstes.* **157**: 1–64.

Boyd, I, Bodeker, G, Connor, B, Swart, D and **Brinksma, E.** 1998. An assessment of ECC ozone sondes operated using 1% and 0.5% KI cathode solutions at Lauder, New Zealand, *Geophys. Res. Lett*. **25**: 2409–2412. DOI: https://doi.org/10.1029/98GL01814

Calisesi, Y, Soebijanta, V and **van Oss, R.** 2005. Regridding of remote soundings: Formulation and application to ozone profile comparison. *J. Geophys. Res.* **110**: D23306. DOI: https://doi.org/10.1029/2005JD006122

Carnuth, W, Kempfer, U and **Trickl, T.** 2002. Highlights of the Tropospheric Lidar Studies at IFU within the TOR Project. *Tellus B.* **54**: 163–185. DOI: https://doi.org/10.1034/j.1600-0889.2002.00245.x

Clerbaux, C, Boynard, A, Clarisse, L, George, M, -Hadji-Lazaro, J, Herbin, H, Hurtmans, D, -Pommier, M, Razavi, A, Turquety, S, Wespes, C and **Coheur, P-F.** 2009. Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder. *Atmos. Chem. Phys*. **9**: 6041–6054. DOI: https://doi.org/10.5194/acp-9-6041-2009

Coffey, MT, Goldman, A, Hannigan, JW, Mankin, WG, Schoenfeld, WG, Rinsland, CP, Bernardo, C and **Griffith, DWT.** 1998. Improved Vibration-Rotation (0-1) HBr Line Patrameters For Validating High Resolution Infrared Atmospheric Spectra Measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*. **60**(5): 863–867. DOI: https://doi.org/10.1016/S0022-4073(98)00088-0

Crutzen, PJ. 1988. Tropospheric Ozone: An Overview, In Tropospheric Ozone. *NATO ASI Series*. Isaksen, ISA (ed.). D. Reidel, Norwell, Mass **227**: 3–32. DOI: https://doi.org/10.1007/978-94-009-2913-5_1

David, SJ, Beaton, SA, Anderberg, MH and **Murcray, FJ.** 1993. Determination of total ozone over Mauna Loa using very high resolution infrared solar spectra. *Geophysical Research Letters*. **20**(19): 2055–2058. DOI: https://doi.org/10.1029/93GL02470

De Backer, H, De Muer, D and **De Sadelaer, G.** 1998. Comparison of ozone profiles obtained with Brewer-Mast and Z-ECC sensors during simultaneous ascents, *J. Geoph. Res.* **103**(D16): 19641–19648. DOI: https://doi.org/10.1029/98JD01711

Deshler, T, Mercer, J, Smit, HGJ, Stuebi, R, Levrat, G, Johnson, BJ, Oltmans, SJ, Kivi, R, Thompson, AM, Witte, J, Davies, J, Schmidlin, FJ, Brothers, G and **Sasaki, T.** 2008. Atmospheric comparison of electrochemical cell ozonesondes from different manufacturers, and with different cathode solution strengths: The Balloon Experiment on Standards for Ozonesondes. *J. Geophys. Res.* **113**: D04307. DOI: https://doi.org/10.1029/2007JD008975

Evans, R, McConville, G, Oltmans, S, -Petropavlovskikh, I and **Quincy, D.** 2009. Measurement of internal stray light within Dobson ozone spectrophotometers. *International Journal of Remote Sensing.* **30**(15): 4247–4258. DOI: https://doi.org/10.1080/01431160902825057

Fabian, P and **Pruchniewicz, PG.** 1975. Surface Ozone Daily Maxima, Project "Troposphärisches Ozon" Contract No Fa 62/1 Deutsche Foraschungagamsinschaft, multiple tables, un-numbered.

Feister, U, Grasnick, KH and **Peters, G.** 1985. Performance of the electrochemical ozone sonde OSR. *Pure Appl. Geophys.* **123**(3): 422–440. DOI: https://doi.org/10.1007/BF00880741

Fujimoto, T, Sato, T, Nagai, K, Nakano, T, Shitamichi, M, Kamata, Y, Miyauchi, S, Akagi, K and **Sasaki, T.** 2004. Further evaluation and improvements of Japanese KC-Ozonesonde through JOSIE-2000. *Proc. XX Quadrennial Ozone Symposium*, 1–8 June 2004, Kos, Greece, International Ozone Commission, Athens, 540–541.

Goldman, A, Paton-Walsh, C, Bell, W, Toon, G, Blavier, J, Sen, B, Coffey, M, Hannigan, J and **Mankin, W.** 1999. Network for the Detection of Stratospheric Change Fourier Transform Infrared Intercomparison at Table Mountain Facility, November 1996. *J. Geophys. Res*. **104**(D23): 30481–30503. DOI: https://doi.org/10.1029/1999JD900879

Hannigan, JW, Coffey, MT and **Goldman, A.** 2009. Semiautonomous FTS Observation System for Remote Sensing of Stratospheric and Tropospheric Gases. *Journal of Atmospheric and Oceanic Technology*. **26**: 1814–1828. DOI: https://doi.org/10.1175/2009JTECHA1230.1

Hase, F. 2012. Improved instrumental line shape monitoring for the ground-based, high-resolution FTIR spectrometers of the Network for the Detection of Atmospheric Composition Change. *Atmos. Meas. Tech*. **5**: 603–610. DOI: https://doi.org/10.5194/amt-5-603-2012

Hase, F, Blumenstock, T and **Paton-Walsh, C.** 1999. Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software. *Appl. Opt*. **38**(15). DOI: https://doi.org/10.1364/AO.38.003417

Hase, F, Hannigan, JW, Coffey, MT, Goldman, A, -Höpfner, M, Jones, NB, Rinsland, CP and **Wood, SW.** 2004. Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*. **87**: 24–52. DOI: https://doi.org/10.1016/j.jqsrt.2003.12.008

Helmig, D, Ganzeveld, L, Butler, T and **Oltmans, SJ.** 2007. The role of ozone atmosphere-snow gas exchange on polar, boundary-layer tropospheric ozone – a review and sensitivity analysis. *Atmos. Chem. Phys*. **7**: 15–30. DOI: https://doi.org/10.5194/acp-7-15-2007

Hurtmans, D, Coheur, PF, Wespes, C, Clarisse, L, Scharf, O, Clerbaux, C, Hadji-Lazaro, J, George, M and **Turquety, S.** 2012. FORLI radiative transfer and retrieval code for IASI. *J. Quant. Spectrosc. Ra.* **113**: 1391–1408. DOI: https://doi.org/10.1016/j.jqsrt.2012.02.036

Johnson, B, Helmig, D and **Oltmans, SJ.** 2008. Evaluation of ozone measurements from a tethered balloon-sampling platform at South Pole Station in -December 2003. *Atmos. Environ.* **42**: 2780–2787. DOI: https://doi.org/10.1016/j.atmosenv.2007.03.043

Junge, CE. 1962. Global ozone budget and exchange between stratosphere and troposphere. *Tellus.* **14**: 363−377. DOI: https://doi.org/10.3402/tellusa.v14i4.9563

Junge, CE. 1963. *Air Chemistry and Radioactivity*. New York: Academic Press. 382 pp.

Kleipool, QL, Dobber, MR, de Haan, JF and **Levelt, PF.** 2008. Earth surface reflectance climatology from 3 years of OMI data. *J. Geophys. Res*. **113**: D18308. DOI: https://doi.org/10.1029/2008JD010290

Koelemeijer, RBA, Haan, JFD and **Stammes, P.** 2003. A database of spectral surface reflectivity in the range 335–772 nm derived from 5.5 years of GOME observations. *J. Geophys. Res.* **108**(D2): 4070. DOI: https://doi.org/10.1029/2002JD002429

Koelemeijer, RBA, Stammes, P, Hovenier, JW and **Haan, JF**. 2001. A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment, *J. Geophys. Res*. **106**(D4), 3475–3490. DOI: https://doi.org/10.1029/2000JD900657.

Komhyr, WD, Oltmans, SJ, Chopra, AN and **-Franchois, PR.** 1985. Performance characteristics of high-altitude ECC ozonesondes. *Atmospheric ozone; Proceedings of the Quadrennial Ozone Symposium, Halkidiki, Greece, September 3–7, 1984*, Zerefos, CS and Ghazi, A (eds.), Dordrecht: D. -Reidel Publishing Co, p. 499. DOI: https://doi.org/10.1007/978-94-009-5313-0_100

Kulawik, SS, Worden, J, Eldering, A, Bowman, K, Gunson, M, Osterman, GB, Zhang, L, Clough, S, Shephard, MW and **Beer, R.** 2006. Implementation of cloud retrievals for Tropospheric Emission Spectrometer (TES) atmospheric retrievals: part 1. Description and characterization of errors on trace gas retrievals. *J. Geophys. Res*. **111**: D24204. DOI: https://doi.org/10.1029/2005JD006733

Kurylo, MJ. 1991. Network for the detection of stratospheric change (NDSC). *Proc. Soc. Photo-Opt. Instrum. Eng*. **1491**: 168–174. DOI: https://doi.org/10.1117/12.46658

Kurylo, MJ and **Zander, R.** 2001. The NDSC—Its status after 10 years of operation. Proc. 19th Quadrennial Ozone Symp, Sapporo, Japan, Hokkaido University, 167–168.

Levelt, PF, van den Oord, GHJ, Dobber, MR, -Malkki, A, Visser, H, de Vries, J, Stammes, -Lundell, JOV and **Saari, H.** 2006. The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Remote Sens.* **44**(5): 1093–1101. DOI: https://doi.org/10.1109/TGRS.2006.872333

Livesey, NJ, Read, WG, Froidevaux, L, Lambert, A, -Manney, GL, Pumphrey, HC, Santee, ML, Schwartz, MJ, Wang, S, Cofield, RE, Cuddy, DT, Fuller, RA, Jarnot, RF, Jiang, JH, Knosp, BW, Stek, PC, Wagner, PA and **Wu, DL.** 2011. EOS MLS -Version 3.3 Level 2 data quality and description document. *Tech. rep*. Jet Propulsion Laboratory, available from http://mls.jpl.nasa.gov.

Mateer, CL, DeLuisi, JJ. 1992. A new Umkehr inversion algorithm. *J. Atmos. Terr. Phys*. **54**: 537–556. DOI: https://doi.org/10.1016/0021-9169(92)90095-3

McDermid, IS, Beyerle, G, Haner, DA and **Leblanc, T.** 2002. Redesign and improved performance of the JPL-TMF tropospheric ozone lidar. *Appl. Opt*. **41**: 7550–7555. DOI: https://doi.org/10.1364/AO.41.007550

Migeotte, M, Neven, L and **Swensson, J.** 1956. The solar spectrum from 2.8 to 23.7 microns', Technical report. *Memoirs of the Royal Society of Sciences of Liege*.

Molod, A, Takacs, L, Suarez, M, Bacmeister, J, Song, IS and **Eichmann, A.** 2012. The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna. *NASA Technical Report Series on Global Modeling and Data Assimilation, NASA TM-2012-104606*. **28**: 117. http://gmao.gsfc.nasa.gov/pubs/tm/archive/tm_2012.php.

Morgan, MG and **Henrion, M.** 1990. *Uncertainty*. -Cambridge University Press, Cambridge, UK. 332 pp.

Morris, GA, Labow, G, Akimoto, H, Takigawa, M, Fujiwara, M, Hasebe, F, Hirokawa, J and **Koide, T.** 2013. On the use of the correction factor with Japanese ozonesonde data. *Atmos. Chem. Phys.* **13**: 1243–1260. DOI: https://doi.org/10.5194/acp-13-1243-2013

Nair, PJ, Godin-Beekmann, S, Pazmiño, A, -Hauchecorne, A, Ancellet, G, Petropavlovskikh, I, Flynn, LE and **Froidevaux, L.** 2011. Coherence of long-term stratospheric ozone vertical distribution time series used for the study of ozone recovery at a northern mid-latitude station. *Atmos. Chem. Phys.* **11**: 4957–4975. DOI: https://doi.org/10.5194/acp-11-4957-2011

Neefs, E, De Mazière, M, Scolas, F, Hermans, C and **Hawat, T.** 2007. BARCOS, an automation and remote control system for atmospheric observations with a Bruker interferometer. *Rev. Sci. Instrum*. **78**: 035109-1–0035109-9. DOI: https://doi.org/10.1063/1.2437144

Orphal, J. 2003. A critical review of the absorption cross-sections of O_3 and NO_2 in the ultraviolet and visible. *J. Photochem. Photobiol, A.* **157**: 185–209. DOI: https://doi.org/10.1016/S1010-6030(03)00061-3

Paneth, FA and **Edgar, JL.** 1938. Concentration and measurement of atmospheric ozone. *Nature* **142**: 112–13. DOI: https://doi.org/10.1038/142112a0

Papayannis, A, Balis, D, Zanis, P, Galani, E, Wernli, H, Zerefos, C, Stohl, A, Eckhardt, S and **Amiridis, V.** 2005. Sampling of an STT event over the Eastern Mediterranean region by lidar and electrochemical sonde. *Ann. Geophys*. **23**: 2039–2050. DOI: https://doi.org/10.5194/angeo-23-2039-2005

Paton-Walsh, C, Bell, W, Gardiner, T, Swann, N, Woods, P, Notholt, J, Schiitt, H, Gallo, B, Ariander, W and **Mcllqvist, J.** 1997. An uncertainty budget for ground-based Fourier transform infrared column measurement of HCL, HF, N_2O, and HNO_3 deduced from results of side-by-side instrument intercomparisons. *J. Geophys. Res* **102**(D7): 8867–8873. DOI: https://doi.org/10.1029/97JD00133

Petropavlovskikh, I, Bhartia, PK, DeLuisi, J. 2005. New Umkehr ozone profile retrieval algorithm optimized for climatological studies. *Geophys. Res. Lett* **32**: L16808. DOI: https://doi.org/10.1029/2005GL023323

Petropavlovskikh, I, Evans, RB, McConville, G, -Miyagawa, K and **Oltmans, S.** 2009. Effect of the out-of-band stray light on the retrieval of the Umkehr Dobson ozone profiles. *International Journal of Remote Sensing* **30**: 6461–6482. DOI: https://doi.org/10.1080/01431160902865806

Petropavlovskikh, I, Evans, R, McConville, G, Oltmans, S, Quincy, D, Lantz, K, Disterhoft, P, Stanek, M and **Flynn, L.** 2011. Sensitivity of Dobson and Brewer Umkehr ozone profile retrievals to ozone cross-sections and stray light effects. *Atmos. Meas. Tech* **4**: 1841–1853. DOI: https://doi.org/10.5194/amt-4-1841-2011

Pougatchev, NS, Connor, BJ, Jones, NB and **Rinsland, CP.** 1996. Validation of ozone profile retrievals from infrared ground-based solar spectra. *Geophysical Research Letters* **23**(13): 1637–1640. DOI: https://doi.org/10.1029/96GL01501

Pougatchev, NS, Connor, BJ and **Rinsland, CP.** 1995. Infrared measurements of the ozone vertical distribution above kitt peak. *J. Geophys. Res.* **100**(D8): 16689–16697. DOI: https://doi.org/10.1029/95JD01296

Purser, RJ, Wu, WS, Parrish, DF and **Roberts, NM.** 2003a. Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Weather Rev* **131**: 1524–1535. DOI: https://doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2

Purser, RJ, Wu, WS, Parrish, DF and **Roberts, NM.** 2003b. Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Weather Rev* **131**: 1536–1548. DOI: https://doi.org/10.1175//2543.1

Reid, SJ, Vaughan, G, Marsh, AR and **Smit, HGJ.** 1996. Accuracy of ozonesonde measurements in the troposphere. *J. Atm. Chem.* **25**: 215–226. DOI: https://doi.org/10.1007/BF00053792

Rienecker, MM, Suarez, MJ, Gelaro, R, Todling, R, Bacmeister, J, Liu,E, Bosilovich, MG, Schubert, SD, Takacs, L, Kim, G, Bloom, S, Chen, J, Collins, D, Conaty, A, da Silva, A, Gu, W, Joiner, J, Koster, RD, Lucchesi, R, Molod, A, Owens, T, Pawson, S, Pegion, P, Redder, CR, Reichle, R, Robertson, FR, Ruddick, AG, Sienkiewicz, M and **Woollen, J.** 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Clim.* **24**: 3624–3648. DOI: https://doi.org/10.1175/JCLI-D-11-00015.1

Rinsland, CP, Connor, BJ, Jones, NB, Boyd, I, -Matthews, WA, Goldman, A, Murcray, FJ, -Murcray, DG, David, SJ and **Pougatchev, NS.** 1996. Comparison of infrared and Dobson total ozone columns measured from Lauder, New Zealand. *Geophysical Research Letters*. **23**(9): 1025–1028. DOI: https://doi.org/10.1029/96GL00708

Saltzman, BE and **Gilbert, N.** 1959b. Microdetermination of Ozone in Smog Mixtures: Nitrogen Dioxide Equivalent Method. *American Industrial Hygiene Association Journal*. **20**: 379–386. DOI: https://doi.org/10.1080/00028895909343736

Stübi, R, Levrat, G, Hoegger, B, Viatte, P, Staehelin, J and **Schmidlin, FJ.** 2008. In-flight comparison of Brewer-Mast and electrochemical concentration cell ozonesondes. *J. Geophys. Res.* **113**(D13302). DOI: https://doi.org/10.1029/2007JD009091

Sullivan, JT, McGee, TJ, Sumnicht, GK, Twigg, LW and **Hoff, RM.** 2014. A mobile differential absorption lidar to measure sub-hourly fluctuation of tropospheric ozone profiles in the Baltimore–-Washington, D.C. region. *Atmos. Meas. Tech.* **7**: 3529–3548. DOI: https://doi.org/10.5194/amt-7-3529-2014

Vigouroux, C, De Mazière, M, Demoulin, P, Servais, C, Hase, F, Blumenstock, T, Kramer, I, Schneider, M, Mellqvist, J, Strandberg, A, Velazco, V, -Notholt, J, Sussmann, R, Stremme, W, -Rockmann, A, Gardiner, T, Coleman, M and **Woods, P.** 2008. Evaluation of tropospheric and stratospheric ozone trends over Western Europe from ground-based FTIR network observations. *Atmos. Chem. Phys*. **8**: 6865–6886. DOI: https://doi.org/10.5194/acp-8-6865-2008

Wu, WS, Purser, RJ and **Parrish, DF.** 2002. Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weather Rev.* **130**: 2905–2916. DOI: https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2

Zbinden, RM, Thouret, V, Ricaud, P, Carminati, F, Cammas, J-P and **Nédélec, P.** 2013. Climatology of pure tropospheric profiles and column contents of ozone and carbon monoxide using MOZAIC in the mid-northern latitudes (24° N to 50° N) from 1994 to 2009. *Atmos. Chem. Phys.* **13**: 12363–12388. DOI: https://doi.org/10.5194/acp-13-12363-2013

Ziemke, JR, Olsen, MA, Witte, JC, Douglass, AR, -Strahan, SE, Wargan, K, Liu, X, Schoeberl, MR, Yang, K, Kaplan, TB, Pawson, S, Duncan, BN, Newman, PA, Bhartia, PK and **Heney, MK.** 2014. Assessment and applications of NASA ozone data products derived from Aura OMI/MLS satellite measurements in context of the GMI chemical transport model. *J. Geophys. Res. Atmos.* **119**: 5671–5699. DOI: https://doi.org/10.1002/2013JD020914