REVIEW

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


From the earliest observations of ozone in the lower atmosphere in the 19th century, both measurement methods and the portion of the globe observed have evolved and changed. These methods have different uncertainties and biases, and the data records differ with respect to coverage (space and time), information content, and representativeness. In this study, various ozone measurement methods and ozone datasets are reviewed and selected for inclusion in the historical record of background ozone levels, based on relationship of the measurement technique to the modern UV absorption standard, absence of interfering pollutants, representativeness of the well-mixed boundary layer and expert judgement of their credibility. There are significant uncertainties with the 19th and early 20th-century measurements related to interference of other gases. Spectroscopic methods applied before 1960 have likely underestimated ozone by as much as 11% at the surface and by about 24% in the free troposphere, due to the use of differing ozone absorption coefficients.

There is no unambiguous evidence in the measurement record back to 1896 that typical mid-latitude background surface ozone values were below about 20 nmol mol⁻¹, but there is robust evidence for increases in the temperate and polar regions of the northern hemisphere of 30–70%, with large uncertainty, between the period of historic observations, 1896–1975, and the modern period (1990–2014). Independent historical observations from balloons and aircraft indicate similar changes in the free troposphere. Changes in the southern hemisphere are much less. Regional representativeness of the available observations remains a potential source of large errors, which are difficult to quantify.

The great majority of validation and intercomparison studies of free tropospheric ozone measurement methods use ECC ozonesondes as reference. Compared to UV-absorption measurements they show a modest (~1–5% ±5%) high bias in the troposphere, but no evidence of a change with time. Umkehr, lidar, and FTIR methods all show modest low biases relative to ECCs, and so, using ECC sondes as a transfer standard, all appear to agree to within one standard deviation with the modern UV-absorption standard. Other sonde types show an increase of 5–20% in sensitivity to tropospheric ozone from 1970–1995.

Biases and standard deviations of satellite retrieval comparisons are often 2–3 times larger than those of other free tropospheric measurements. The lack of information on temporal changes of bias for satellite measurements of tropospheric ozone is an area of concern for long-term trend studies.

Keywords: Ozone; Troposphere; Measurements; Trends; Historical; Climate
1. Introduction

Tropospheric ozone is a greenhouse gas and pollutant detrimental to human health and plant growth (Monks et al., 2015; WMO Reactive Gases Bulletin, 2018). Large changes after 1990 in the global distribution of the anthropogenic emissions that produce ozone have been reported, including reductions in North America and Europe and increases in Asia (Richter et al., 2005; Granier et al., 2011; Russell et al., 2012; Hilboll et al., 2013; Cooper et al., 2014; Zhang et al., 2016). This rapid shift, coupled with limited ozone monitoring in developing nations, has left scientists unable to answer the most basic questions: Which regions of the world have the greatest human and plant exposure to ozone pollution? How is ozone changing in nations with strong emission controls? To what extent is ozone increasing in the developing world? How can the atmospheric sciences community facilitate access to the ozone metrics necessary for quantifying ozone’s impact on climate, human health and crop/ecosystem productivity?

To answer these questions, the International Global Atmospheric Chemistry Project (IGAC) developed the Tropospheric Ozone Assessment Report (TOAR): Global metrics for climate change, human health and crop/ecosystem research (www.igacproject.org/activities/TOAR). Initiated in 2014, TOAR’s mission is to provide the research community with an up-to-date scientific assessment of tropospheric ozone’s global distribution and trends from the surface to the tropopause. TOAR’s primary goals are, 1) Produce the first tropospheric ozone assessment report based on the peer-reviewed literature and new analyses, and 2) Generate easily accessible, documented data on ozone exposure metrics at thousands of measurement sites around the world (Lefohn et al., 2018). Through the TOAR surface ozone database (Schultz et al., 2017; hereinafter TOAR-Surface Ozone Database) these ozone metrics are freely accessible for research on the global-scale impact of ozone on climate (Gaudel et al., 2018), human health (Fleming et al., 2018) and ecosystem productivity (Mills et al. 2018).

The assessment report is organized as a series of papers in a special feature of Elementa – Science of the Anthropocene (https://collections.elementascience.org/toar), with this paper comprising the Tropospheric Ozone Assessment Report: Tropospheric ozone from 1877 to 2016, observed levels, trends and uncertainties, subsequently abbreviated as TOAR-Observations. This paper describes the different tropospheric ozone measurement techniques used since the late 19th century to the present, and characterizes the uncertainty in the measurements and the spatial and temporal information obtained from each instrument type.

Knowledge of the uncertainties associated with tropospheric ozone measurements is important to reconciling measurements from different methods and platforms and for accurate and realistic model evaluation. It is also essential for the evaluation of trends. Historical ozone observations, those made before the widespread deployment of UV-based ozone instruments, are important to climate models. The global average radiative forcing of ozone (0.4 ± 0.2 W m⁻²; IPCC, 2013) is approximately 1/5 of the
radiative forcing due to CO₂, and slightly less than the radiative forcing due to methane (NOAA, 2018). This estimate has large uncertainty due to limited knowledge of pre-industrial concentrations of tropospheric ozone and its present-day spatial distribution (IPCC, 2013). Additional uncertainty arises from the detrimental impact of ozone on plant productivity, which due to feedbacks on CO₂ uptake, produces an indirect forcing (Sitch et al., 2007). Past efforts to evaluate 19th century ozone measurements have concluded that ozone in pre-industrial times was as low as 1/5 of its present concentration (e.g. Mareno et al., 1994; Volz and Kley 1988; Bojkov, 1986; Staehelin et al., 1994), based primarily on observations at Montsouris, Paris, France in the late 19th century. However, the validity of the early Montsouris measurements as representative of the regional atmosphere has been challenged (Calvert et al., 2015; Staehelin et al., 2017), and global atmospheric chemistry models have difficulty reproducing such a large historical increase from pre-industrial times (e.g. Wang and Jacob, 1998; Mickley et al., 2001; Lamarque et al., 2005; Young et al., 2013; Parrish et al., 2014; Young et al., 2018). It is therefore important to quantify uncertainties for these older measurement methods, to establish confidence limits for reproducibility and bias, and to answer the question: how well do we know historic levels of tropospheric ozone?

Section 2 of this paper describes the many methods that have been used to measure tropospheric ozone. Section 3 is an in-depth re-evaluation of the record of ozone in surface air away from cities and other interfer-ences. Section 4 addresses the measurement of ozone in the free troposphere, beginning with the relatively few historical measurements. Section 5 discusses several aspects of representativeness, and uncertainties associated with sampling of ozone in the troposphere. The paper concludes with a discussion of knowledge gaps and recommendations for future measurements.

2. Standards for the measurement of ozone in the atmosphere

Ozone is a highly reactive gas, with strong absorption bands in the IR and the UV. Three broad sets of techniques based on chemical reaction, UV absorption and IR absorption and emission have been used to measure ozone in the atmosphere. The methods derived from these techniques and their first use to measure ozone in the atmosphere are presented in Table 1. These methods have different measurement uncertainties and the results obtained from paired measurements using either the same or different techniques may differ from each other both systematically and randomly.

As a reactive gas ozone cannot currently be kept in containers nor does it persist in snow without ongoing loss. Hence no current measurements of past concentra-tions are possible (although they may be inferred from isotopic measurements of oxygen trapped in ice (Yeung et al., 2019)). It is also not possible to transport a sample of gas containing a known concentration of ozone from one location to another without ozone loss occurring within the container. Therefore, some other form of standard for ozone calibration to ensure world-wide trace-ability of measurement results is required. The current standard for tropospheric ozone measurement is based on its ultraviolet absorption cross-section at 253.65 nm of 1.148 x 10⁻²³ cm² molecule⁻¹. This standard originates from Hearn (1961), and has been adopted by the International Ozone Commission in 1984, the International Standards Organisation (ISO, 2017), the World Meteorological Organisation in its Guidelines for Continuous Measurement of Ozone in the Troposphere (Galbally et al. 2013) and is used by the International Bureau of Weights and Measures (BIPM) for ozone calibrations (BIPM, 2019).

To propagate this standard for surface ozone and aircraft ozone measurements, specially designed ozone photometers incorporating an ozone generator, and utilizing the measurement of the absorption of UV radiation of 253.65 nm wavelength within short cells (1 m or less in length) by sample air containing ozone, have been used as ozone transfer standards (ISO, 2017; Paur et al., 2003; Viallon et al., 2006a). By referencing ambient measurements to these standards, well-understood and traceable observa-tions of tropospheric ozone are made (Galbally et al., 2013; Tanimoto et al., 2007; Viallon, 2006a, b).

The numerical value of the ozone absorption cross-section is currently under review (Hodges et al., 2019; Orphal et al. 2016), with a recommendation that the value should be decreased by approximately 1.23% (Hodges et al., 2019). If accepted by the appropriate agencies (BIPM, WMO, ISO), this change will require all tropospheric ozone measurements on the current UV standard scale to be increased by 1.23%. This will not affect trends, but it will have a small effect on estimates that depend on the absolute ozone amount, such as calculations of ozone radiative forcing. This change will also improve agreement of the UV scale with gas phase titration (GPT) and the potassium iodide (KI) ECC ozonesondes.

A second ozone standard is gas phase titration of ozone against nitric oxide gas standards. Differences between GPT and standard UV photometry have been investigated by Tanimoto et al. (2006) and Viallon et al. (2006b, 2016) and found to be very small (~0.3%) when the newer values of the ozone absorption cross-sections (see Section 2.2.1) are used (Viallon et al. 2016). Thus GPT supports the proposed decrease in the ozone absorption cross-section at 253.65 nm. Because GPT is utilized as a standard and has not been used for ambient ozone measurements in either the historical record or the TOAR database, it is not listed in Table 1. Further information on GPT is included in the Supplemental Material (Text S-1).

These standards are propagated, via international co-ordination of the adoption of ozone absorption coefficients for UV and visible light (Orphal et al., 2016), to the communities using remote sensing methods for ozone measurement in the free troposphere. There is a recommendation to extend this co-ordination to infrared methods (Orphal et al., 2016).

The available record of surface ozone measurements divides into two periods: the modern period covers approximately 1975 to the present and is defined by the widespread availability of sensitive UV photometers for
surface ozone measurements; the historical period covers 1877–1975 and is defined by the use of other techniques and the absence of these UV photometers. There are a few years of overlap between the periods during the uptake of the modern technology.

A set of four criteria were developed and applied to select data for the historical reconstruction: (1) the measurement methods used should be related, through intercomparisons, 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) for surface ozone, 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 major artifacts or inconsistencies. The datasets that pass these four criteria are explicitly documented.

To commence the reconstruction of a historical tropospheric ozone record, the first of the four criteria must be addressed for each of the ozone data sets examined. As many historical measurement methods pre-date the current UV method, traceability may be derived through published side-by-side intercomparisons with an intermediate method.

The surface ozone method/instrument intercomparisons found in the literature are presented in Table 2. The ratio of each pair of methods corresponds to either the ratio of the mean values from each method or the slope of a regression whose intercept is assumed to be zero. Often only this number is recorded. In some cases an uncertainty is cited. Where not explicitly stated, we have assumed that this corresponds to one standard deviation. Table 2 is divided into 5 sections, the first four sections commencing with the comparison of a method with the UV method and then proceeding to other comparisons of that or closely related methods. The methods separated are Levy, Ehmert, Electrochemical Concentration Cell (ECC), and Colorimetric. The fifth section is for other relevant method comparisons not included in the first four sections. Two conclusions can be drawn from Table 2: (a) in the absence of other information, the relative bias of a historical set of ozone observations with the current UV standard lies in the range 0.7 to 1.2 at approximately 90% confidence limit; and (b) the uncertainty in the bias from one to another study of apparently identical instruments can be as large as 50%. Consequently, except in special

---

**Table 1:** The introduction of various techniques for measurement of ozone in the troposphere. (KI = Potassium Iodide).

<table>
<thead>
<tr>
<th>Date</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1845</td>
<td>KI-Starch papers</td>
<td>Schönbein (1845)</td>
</tr>
<tr>
<td>1876</td>
<td>KI manual volumetric</td>
<td>Albert-Lévy (1877)</td>
</tr>
<tr>
<td>1929</td>
<td>UV – Umkehr Inverse method</td>
<td>Götz et al. (1934)</td>
</tr>
<tr>
<td>1931</td>
<td>Long path UV</td>
<td>Götz and Ladenberg (1931), Fabry and Buisson (1931)</td>
</tr>
<tr>
<td>1934</td>
<td>Balloon borne UV</td>
<td>Regener and Regener (1934)</td>
</tr>
<tr>
<td>1938</td>
<td>Cryotrapping and subsequent analysis</td>
<td>Edgar and Paneth (1941a)</td>
</tr>
<tr>
<td>1941</td>
<td>Automatic KI</td>
<td>Paneth and Glückauf et al. (1941)</td>
</tr>
<tr>
<td>1943</td>
<td>Aircraft KI</td>
<td>Ehmert (1949)</td>
</tr>
<tr>
<td>1955</td>
<td>UV ozonesondes</td>
<td>Paetzold (1955)</td>
</tr>
<tr>
<td>1956</td>
<td>IR tropospheric ozone</td>
<td>Walshaw and Goody (1956)</td>
</tr>
<tr>
<td>1958</td>
<td>KI ozonesondes</td>
<td>Brewer and Milford (1960)</td>
</tr>
<tr>
<td>1960</td>
<td>Chemiluminescent ozonesondes</td>
<td>Regener (1960)</td>
</tr>
<tr>
<td>1972</td>
<td>UV surface ozone analysers</td>
<td>Bowman and Horak (1972)</td>
</tr>
<tr>
<td>1990</td>
<td>Satellite tropospheric ozone residual</td>
<td>Fishman et al. (1990)</td>
</tr>
<tr>
<td>1997</td>
<td>Satellite UV/VIS backscatter</td>
<td>Chance et al. (1997); Liu et al. (2005)</td>
</tr>
<tr>
<td>2001</td>
<td>Satellite IR atmospheric emission</td>
<td>Beer et al. (2001); Worden et al. (2007a)</td>
</tr>
<tr>
<td>2007</td>
<td>Satellite Multispectral</td>
<td>Worden et al. (2007b); Landgraf and Hasekamp (2007)</td>
</tr>
</tbody>
</table>
cases of traceability, each past set of observations should be seen as having a substantial unknown bias and the historical record at a particular location will not necessarily sensibly relate to current measurements at the same location. Due to such inconsistencies between historical and modern ozone observations, \textit{TOAR-Observations} estimates historical ozone levels on regional or zonal scales using all available data sets. The rationale for this new approach relies on the concept that the normalised biases of multiple sets of observations, if random, will tend to cancel out when averaged across multiple stations. A formal description of this approach is given in the Supplemental Material (Text S-2). The average of multiple sets of these past observations for multiple years and a particular geographic region is more appropriate for comparison with current observations. Therefore, we recommend that the

<table>
<thead>
<tr>
<th>Method Comparison</th>
<th>Ratio</th>
<th>Uncertainty</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>KI-arsenite/UV (A)</td>
<td>0.78</td>
<td>n/a</td>
<td>Dauvillier (1935)</td>
</tr>
<tr>
<td>KI-arsenite/UV (L)</td>
<td>1</td>
<td>±0.02</td>
<td>Volz and Kley (1988)</td>
</tr>
<tr>
<td>Ehmert/UV (A)</td>
<td>0.98</td>
<td>±0.09</td>
<td>Galbally (1979)</td>
</tr>
<tr>
<td>Ehmert/UV (A)</td>
<td>0.947</td>
<td>±0.009</td>
<td>Grasso (2011)</td>
</tr>
<tr>
<td>KI/UV (A)</td>
<td>0.93</td>
<td>±0.04</td>
<td>Vasy (1958)</td>
</tr>
<tr>
<td>ECC/Ehmert (A)</td>
<td>1.02</td>
<td>±0.12</td>
<td>WMO (1972), Galbally (1979)</td>
</tr>
<tr>
<td>MPI-Pruch/Ehmert (A)</td>
<td>1</td>
<td>±0.05</td>
<td>Pruchniewicz (1973)</td>
</tr>
<tr>
<td>NBKI colorimetric/Ehmert (L)</td>
<td>1.1</td>
<td>n/a</td>
<td>Renzetti (1959)</td>
</tr>
<tr>
<td>NBKI colorimetric/Ehmert (A)</td>
<td>1.22</td>
<td>±0.15</td>
<td>Galbally (1979)</td>
</tr>
<tr>
<td>Mast Brewer ozonesonde/Ehmert (A)</td>
<td>0.88</td>
<td>±0.10</td>
<td>Galbally (1979)</td>
</tr>
<tr>
<td>Cauer/Ehmert (A)</td>
<td>0.66</td>
<td>n/a</td>
<td>Warmbt (1964)</td>
</tr>
<tr>
<td>Cauer/Ehmert (corrected) (A)</td>
<td>0.9</td>
<td>n/a</td>
<td>Warmbt (1964)</td>
</tr>
<tr>
<td>Cryotrapping O$_3$/KI-thiosulfate</td>
<td>1</td>
<td>±0.05</td>
<td>Edgar and Paneth (1941a)</td>
</tr>
<tr>
<td>ECC/UV (A)</td>
<td>1.1</td>
<td>±0.14</td>
<td>Attmannspacher and Hartmannsgruber (1982)</td>
</tr>
<tr>
<td>ECC/UV (A)</td>
<td>1.01</td>
<td>±0.05</td>
<td>This study, Section 4.3</td>
</tr>
<tr>
<td>ECC/UV (A)</td>
<td>1.05</td>
<td>±0.04</td>
<td>IAGOS, this study, Section 4.8</td>
</tr>
<tr>
<td>HP-KI/UV (A)</td>
<td>0.94</td>
<td>±0.16</td>
<td>Attmannspacher and Hartmannsgruber (1982)</td>
</tr>
<tr>
<td>MPI-Pruch/HP-KI (A)</td>
<td>0.5</td>
<td>±0.04</td>
<td>This study, Section 2.1.3</td>
</tr>
<tr>
<td>NBKI colorimetric/UV (L)</td>
<td>0.97</td>
<td>n/a</td>
<td>Cherniack and Bryan (1965)</td>
</tr>
<tr>
<td>NBKI colorimetric/UV (A)</td>
<td>1.02</td>
<td>n/a</td>
<td>Cherniack and Bryan (1965)</td>
</tr>
<tr>
<td>1% NBKI/UV (L)</td>
<td>1.09</td>
<td>±0.02</td>
<td>Schnadt Poberaj et al (2007)</td>
</tr>
<tr>
<td>2% NBKI colorimetric/UV (L and A)</td>
<td>1.23</td>
<td>±0.06</td>
<td>Pitts et al. (1976a, b)</td>
</tr>
<tr>
<td>2% unbuffered KI titration/UV (L and A)</td>
<td>0.9</td>
<td>n/a</td>
<td>Pitts et al. (1976b)</td>
</tr>
<tr>
<td>Pressure/Volume/UV (L)</td>
<td>1.03</td>
<td>n/a</td>
<td>Watanabe and Stephens (1979)</td>
</tr>
<tr>
<td>Regener chemiluminescent/UV (L)</td>
<td>1</td>
<td>n/a</td>
<td>Regener (1964)</td>
</tr>
<tr>
<td>Ethylene-Chemiluminescent/UV (A)</td>
<td>0.94</td>
<td>±0.16</td>
<td>Attmannspacher and Hartmannsgruber (1982)</td>
</tr>
<tr>
<td>Ozonograph-KI/UV (A)</td>
<td>1.09</td>
<td>±0.20</td>
<td>Attmannspacher and Hartmannsgruber (1982)</td>
</tr>
<tr>
<td>Mast Ozone Meter/Pressure/Volume (L)</td>
<td>1.04</td>
<td>n/a</td>
<td>Watanabe and Stephens (1979)</td>
</tr>
<tr>
<td>Mast Ozone Meter/NBKI colorimetric (L and A)</td>
<td>0.86</td>
<td>n/a</td>
<td>Cherniack and Bryan (1965)</td>
</tr>
<tr>
<td>Mast Ozone Meter/NBKI colorimetric (L)</td>
<td>0.71</td>
<td>n/a</td>
<td>Gudiksen et al. (1966)</td>
</tr>
<tr>
<td>Regener chemiluminescent/Mast Ozone Meter (A)</td>
<td>1.2–1.8</td>
<td>n/a</td>
<td>Oltmans and Komhyr (1976)</td>
</tr>
</tbody>
</table>
evaluation of models be based on the regional or zonal means of multiple sets of observations rather than individual time series. For the discussion of individual historical time series uncertainties the reader is referred to the Supplemental Material (Text S-3).

Observations cited in this paper are quoted as reported, for ease of comparison with previous work, but expressed in their equivalent SI units. Concentrations are then converted to mole fractions (Text S-4) for purposes of comparison.

Each of the techniques used in ozone data sets either selected or rejected in this study will be discussed in turn.

2.1. Potassium Iodide measurement techniques
While ozone was originally identified and investigated by its odour (Schönbein, 1840; Rubin, 2001), the first quantitative measurements were based on the reaction of ozone with potassium iodide:

\[ \text{O}_3 + 2\text{KI} + \text{H}_2\text{O} \rightarrow \text{O}_2 + \text{I}_2 + 2\text{KOH} \]  

The basis of this measurement is the assumption that for each ozone molecule reacted, a molecule of iodine is produced; this ratio is the stoichiometry of the reaction. The amount of iodine produced is (in most methods) measured, and this in mole units, equals the amount of ozone in the air volume sampled. A number of techniques based on the KI reaction have been developed during the last two centuries, and the ozone-KI reaction is still in use in balloon borne ozonesondes (Table 1).

The stoichiometry of the reaction is crucial, and has been studied extensively. Many studies, however, were made at ozone concentrations much higher than those in the troposphere, because of the difficulty of working with low concentrations of ozone at the time (Saltzman and Gilbert, 1959; Boyd et al., 1970; Hodgeson et al., 1971; Kopczynski and Bufalini, 1971; Dietz et al., 1973). Byers and Saltzman (1959) found using GPT a reaction stoichiometry of 1.00 (with unquantified uncertainty) at pH 7, and that the reaction stoichiometry varied with pH, being lower by 50% at pH 14. This implies that the chemistry of the KI reaction with ozone is complex, involving reactions other than (1) that produce additional iodine, as well as reactions that cause loss of iodine (Byers and Saltzman, 1959; Staehelin and Hoigné, 1985).

Without buffering, the reaction will drive the solution alkaline, so the KI solution is in most methods buffered (NBKI, for neutral-buffered KI). Dietz et al. (1973) found a NBKI/UV ratio of 1.00 ± 0.03 at pH 7 for two measurements at 100 and 400 nmol mol⁻¹. Pitts et al. (1976a) found that the 2% NBKI method gave NBKI/IR = NBKI/UV = 1.23 ± 0.06 at 50% relative humidity for 0.1 to 1 ppm ozone and NBKI/IR = NBKI/UV = 1.14 ± 0.04 at 3% relative humidity. They had no explanation of the apparent water vapour dependence. However, it was reduced when potassium bromide was added (Lanting, 1979; Bergsheeoff et al., 1980), as is the case in ozonesondes. Slow side reactions involving the phosphate buffer may also change the stoichiometry from 1.0 (Saltzman and Gilbert, 1959; Flamm, 1977; Johnson et al., 2002).

Other compounds present in air can interfere with the KI-ozone reaction. NO, and H₂O₂ give positive interferences (Volz and Kley, 1988), NO, at a level of 5–10% (Pitts et al., 1976), although this appears to be quite variable (Cherniak and Bryan, 1965; Tarasick et al., 2000). SO₂ causes a negative interference of 1:1, i.e. a quantitative reduction in the ozone detected (Pitts et al., 1976; Schenkel and Broder, 1982; Volz and Kley, 1988). NH₃ is also a negative interferent (Anfossi et al., 1991), which increases the pH of the solution as well as reacting directly with iodine although the stoichiometry is not quantified (Downs and Adams, 1973).

Losses can occur in the inlet to the sampler, but even early experimenters appear to have been aware of this, and strove to avoid it. Inlet tubes (where described) were usually of glass (e.g. Dauvillier, 1934; Glückauf et al., 1944) and the type of glass was found to be important (Carbenay and Vassy, 1953). Other materials such as polyvinyl chloride became available later (e.g. Vassy, 1958), and may have caused negative biases in some cases (Alshuller et al., 1961; Potter and Duckworth, 1965) before Teflon was introduced (Gudiksen et al., 1966). In one case a cotton wool filter was used in the inlet (Edgar and Paneth, 1941b). No information is available with which to estimate inlet losses, but they could have negatively biased some of the KI measurements.

Loss due to evaporation of the iodine produced can also occur (Brewer and Milford, 1960; Kley et al., 1988). There are a number of methods based on the KI reaction (Table 1), and while all are similarly subject to interfering gases, they differ in terms of potential for iodine and/or ozone loss, and side reactions. In the Cauer method, the evaporation of the iodine is part of the analytical technique (Warmbt, 1964). The efficiency of the sampler needs to be considered for each case.

The contributions of each of these interfering or modifying factors cannot always be separately quantified. The best available summary of how a technique performs is obtained from comparisons in unpolluted ambient air with the UV method or with a method traceable to the UV method, as presented in Table 2.

2.1.1. Schönbein papers
The Schönbein paper method uses a KI and starch impregnated paper. Ozone diffusing to the paper surface reacts with the iodide, and the iodine produced forms a strongly blue-colored complex with the starch. Alternately, a pH indicator is present and the colour change is due to the alkalinity resulting from reaction (1), see Hartley (1881). Following exposure, the paper strips can be compared to a standard color scale to give a semi-quantitative ozone measurement (Fox, 1873). There are several variations on this technique, named after their inventors/developers: Schönbein, Sallo, de James, Teyr and Houzeau. These methods are described in limited detail (Houzeau, 1857; Fox 1873; Hartley 1881; Linville et al. 1980; Bokoj 1980; Kley et al. 1988; Anfossi et al. 1991).

Interest in ozone was very high in the late 19th Century, in part because of its role as an “air purifier” and the erroneous belief that it could eliminate pathogens, particularly...
cholera (Fox, 1873). Measurements were therefore made with Schönbein or related papers at hundreds of sites in Europe, the Americas, Australia, Asia, Africa, and Antarctica (Smyth, 1858; Fox, 1873; Royal Society, 1908; Bojkov, 1986; Galbally and Paltridge, 1989; Sandroni et al., 1992; Sandroni and Anfossi, 1994; Pavelin et al., 1999; Nolle et al., 2005).

There are two laboratory test chamber studies of the color development response to time and ozone concentration, that either directly or indirectly relate the filter paper method to the current UV-absorption standard. Linvill et al., (1980) found a filter paper response relationship to ozone exposure where color development was very strongly dependent on the relative humidity (RH) present in the chamber. A change from 3 to 4 (of 10) color units corresponds to a 10 nmol mol\(^{-1}\) ozone change at 80% RH and a 30 nmol mol\(^{-1}\) ozone change at 60% RH. Kley et al. (1988) found the papers gave, on exposure to a constant ozone level, an initial linear color increase for 3 hours or more, followed by a plateau and then a color decrease. Further exposure to ozone increased this color loss. Because of this complex behavior, there is a region between 6 and 10 hours exposure where ozone values between 10 and 50 nmol mol\(^{-1}\) correspond to less than 1 unit difference on the Schönbein scale. For longer exposures the response overlap and reverse order, i.e. longer exposures at high concentrations for selected conditions give lower color responses than some shorter exposures at lower concentrations. Consequently, it is impossible to uniquely relate a color development on these filter papers to an ozone concentration on the UV scale.

The filter paper method of measuring ozone concentrations is a passive measurement method that lacks a controlled diffusion barrier, the absence of which creates a wind speed dependence. There does not appear to be any information on the repeatability and reproducibility of these techniques in the field. It appears that the colour development (the signal) is dependent on ozone concentration, time of exposure, relative humidity, wind speed and light. The colour development has negative responses to ammonia and sulphur dioxide (Fox 1873; Linvill et al. 1980; Bokjov 1980; Kley et al. 1988). There may also be differences in response dependent on type of paper and method of preparation.

In 1876–1877, 289 parallel ozone measurements at the Montsouris Observatory in Paris were undertaken using KI papers and the more quantitative KI-arsenite method (Albert-Lévy, 1877). This study was repeated in the two following years (Marenco et al. 1994). The individual data are not available, but a frequency table is available from which a linear regression relationship can be determined. Several authors have used such a relationship in attempts to calibrate the KI paper measurements (Bojkov, 1986; Anfossi et al., 1991; Lisac and Grubišić, 1991; Sandroni et al., 1992; Marenco et al., 1994; Anfossi and Sandroni, 1997; Pavelin et al., 1999; Weidinger et al., 2011). These either use the Montsouris comparison directly, or use it to scale the chamber results of Linvill et al. (1980). In all cases they find that ozone was much lower in the 19th century than now. Evidently, this conclusion is entirely dependent on the calibration at Montsouris, a single site on the edge of an urban centre, not necessarily representative of the regional background atmosphere (the Montsouris measurements are discussed in section 3.1 below). Moreover, the results of the scaling are not consistent with the chamber measurements (compare Figure 1 of Pavelin et al., 1999 to Figure 1 of Linvill et al., 1980 or Figure 3 of Kley et al., 1988).

The KI papers appear to have been useful as a relative measure of ozone concentration, and showed many aspects of ozone variation and distribution that are now well known (Bojkov, 1986; Anfossi et al. 1991). However, given the high sensitivity of KI papers to relative humidity (greater than to ozone concentration), exposure time, wind speed, and other factors, and the radically different results from intercomparisons, the filter paper measurements cannot be related to modern ozone measurements with any degree of confidence, and are not recommended for quantitative use. The same recommendation was made by Fox (1873), Hartley (1881) and Kley et al. (1988).

### 2.1.2. The Albert-Lévy and Ehmert ozone measurement methods

In the Albert-Lévy and Ehmert KI methods, ozone is measured by bubbling a known quantity of air through an aqueous solution of iodide (I\(^{-}\)) and either arsenite (Albert-Lévy 1877; Dauvillier, 1934; Volz and Kley 1988) or thiosulfate (Ehmert 1951, 1952, 1959; Galbally 1969).

In the Albert-Lévy technique, the sampling solution contains iodide and arsenite. Ozone bubbling through the solution reacts with the iodide, producing iodine. The iodine produced reacts with the arsenite (AsO\(_3^2\)), converting it to arsenate (AsO\(_4^{3-}\)). Two titrations are performed, to determine the amount of arsenite in a vessel of the solution that has had air bubbled through it, and in an identical vessel that has not been exposed to bubbling. The titrations are conducted in an alkaline medium with a volumetric standard solution of iodine. The quantity of ozone in the air is calculated from the difference between the amounts of arsenite in the two vessels. Measurements were continuous, 24-hour sampling averages. Volz and Kley (1988) replicated the apparatus and method of Albert-Lévy and found agreement in the laboratory with the UV method to ±2%.

Dauvillier (1935) undertook an intercomparison of the UV and the KI-arsenite methods using atmospheric samples over a snow surface in Abisko, Sweden. Between 22 December 1934 and 7 March 1935, there were 50 simultaneous measurements and the ratio of the derived ozone amounts was 0.78. This suggests that in practice the KI-arsenite method may underestimate atmospheric ozone levels.

In the Ehmert technique (Ehmert 1951, 1952, 1959; Galbally 1969), the sampling solution is neutral buffered and contains iodide and thiosulfate (S\(_2\)O\(_3^{2-}\)). Ozone bubbling through a vessel of the solution reacts with iodide producing iodine which converts the thiosulfate to tetrathiionate (S\(_4\)O\(_6^{2-}\)). With the Ehmert technique the quantity of thiosulfate in the vessel is calculated electrochemically using a coulometric (where chemical transformation is
equated to electron flow) analysis (Ehmert, 1959; Galbally, 1969). The thiosulfate loss equals the ozone amount sampled. Dividing this amount by the air volume sampled gives the ozone concentration. Measurements can be made with air sampling as short as 30 minutes.

As presented in Table 2, the Ehmert, ECC and UV methods had ratios indistinguishable from 1.0 with an uncertainty of approximately ±10%. Other variations of the Ehmert method involve injection of known amounts of thiosulfate into the reacting solution which allows continuous or semi-continuous measurements (Paneth and Glückauf 1941; Glückauf et al., 1944; Bowen and Regener 1951; Carbenay and Vassy, 1953; Regener 1959). An ozone sonde based on this method was developed (Kobayashi and Toyama, 1966a).

The Ehmert and ECC methods are considered suitable methods for the analyses presented here and are utilized here as intermediate methods when relating other methods to the UV standard.

2.1.3. Ozone measurement methods where the iodine accumulates in solution
A number of the KI methods accumulate iodine in solution while the air is sampled. The first approach to this is the colorimetric method of measuring ozone, where the iodine produced in the neutral buffered 1% KI solution is subsequently measured spectrosopically at 352 nm and quantified by comparison against iodine standards (Byers and Saltzman 1959; Saltzman and Gilbert 1959a). Saltzman and Gilbert (1959a) demonstrated that the iodide concentration, the pH of the solution and the time delay before measuring the iodine absorbance were all critical parameters. They also demonstrated the effectiveness of SO₂ as a negative interferent in the KI method. There have been multiple comparisons of the colorimetric NBKI method against other methods, as listed in Table 2. It shows a high bias with respect to UV methods, with some exceptions (e.g. Cherniak and Bryan, 1965). Before the mid-1970s, benchtop UV photometers and chemiluminescent ozone analysers were calibrated to an equal calibration source, typically NBKI (e.g. Clements, 1975; Pitts et al., 1976; Torres and Bandy, 1978).

The second approach involves removal of iodine produced in the solution. In the simplest case the anode and cathode are chemically separated by the steady one-way flow of the sensing solution. The iodine produced from ozone is measured via the current from a platinum cathode downstream of the air mixing zone. The reaction at the cathode is:

\[ I_2 + 2e \rightarrow 2I^- \]  (3)

The charge flow in the external circuit is proportional to the ozone reacted. This “coulometric” method is used in the transmogrifier (Brewer and Milford 1960) and its commercial adaption as the Mast Ozone Meter (Mast and Saunders, 1962), as well as the instrument developed at the Max Planck Institute for Aeronomy (Pruchniewicz 1970, 1973) which will be subsequently described as MPI-Pruch. This “coulometric” method is also used in the Brewer-Mast and Electrochemical Concentration Cell (ECC) ozonesondes and their adaptions for use in surface air. In the case of the ECC sonde an ion bridge is used instead of relying on the one-way flow of the sensing solution.

The Mast ozone meter has undergone other testing to that in Table 2 (e.g. Gudiksen et al., 1966; Potter and Duckworth, 1965), indicating that under field conditions it gave responses of 50–70% of the NBKI result. A correction factor of 0.8 is used here, corresponding to the Mast Ozone Meter ratio to the UV method based on information in Table 2 and references above.

There was widespread use of the surface ECC methods in the 1970s (Oltmans, 1981) and there are comparisons against the UV and Ehmert methods, see Table 2. The HP-KI method, used for surface ozone observations at Hohenpeissenberg Observatory between 1971 and 1986, is a variation of the ECC method as developed by Kobayashi and Toyama (1966b). The HP-KI method was compared with the Ehmert method via an ozone generator supplied by V. H. Regener and gives a ratio of 1.0 ± 0.02 (Attmannspacher and Hartmannsgruber, 1982).

The Pruch-Mast analyser (Pruchniewicz, 1973) is another KI method relying on one way flow of the sensing solution. Comparisons give 1.0 ± 0.05 for ambient ozone measurements with the MPI-Pruch analyser and the Ehmert method at 4 sites and also laboratory comparisons against an ozone generator supplied by V. H. Regener (Pruchniewicz, 1973). However, new evaluations conducted for TOAR-Observations cast doubt on the reliability of this method in the field. A comparison of 15 months of overlapping observations of ambient surface ozone measurements by the MPI Pruch and HP-KI methods at the Hohenpeissenberg Observatory from January 1971 to May 1972, shows that the MPI Pruch method indicates 0.50 ± 0.04 of the ozone level given by the HP-KI method. Similar results were obtained comparing a MPI-Pruch analyser at Zugspitze with nearby ozonesonde results from the same height in the atmosphere. These comparisons are in sharp contrast with previously described comparisons of the HP-KI method and the MPI Pruch method (Pruchniewicz, 1973; Attmannspacher and Hartmannsgruber, 1982). One possible explanation is that the instruments, due to some unknown factor, were far more variable in their efficiency at measuring ozone than the initial tests indicated. Given the information available, no conclusion about the cause can be drawn. This is discussed further in the Supplemental Material (Text S-3).

2.1.4. The Cauer ozone measurement method
The Cauer technique (Cauer 1935, 1951) involves bubbling a large volume of air (100 to 300 litres) through a 10 ml solution buffered by sodium acetate and containing 50 μg iodine as KI. The ozone in the air converts the iodide to iodine and the iodine evaporates into the air stream. At the end of the sampling the remaining iodide is determined and the loss of iodide is equated to the quantity of ozone in the air sampled. When divided by the air volume sampled this gives the ozone concentration. Cauer (1951) writes "For experienced chemists, this analysis requires 20 minutes, but necessitates great care, and is difficult
for non-chemists. The Cauer method was used in multiple studies between 1935 and 1955, and in the national network of the then East Germany from 1952 until 1982 (Cauer 1951; Teichert 1955; Warmbt 1964; Feister and Warmbt 1987). It was compared with the Ehmert method by Warmbt (1964). An initial study, with the two systems presumably in their normal operating conditions, gives a Cauer/Ehmert ratio of 0.66. A more intense study, standardising various components and including blank corrections, gives a Cauer/Ehmert ratio of 0.90. However, for low ozone concentrations, when the Cauer method measured less than 5 nmol mol⁻¹, the Ehmert method often measured 10–20 nmol mol⁻¹ (Warmbt, 1964). These results cause considerable uncertainty concerning what corrections should be applied to Cauer data from different measurement periods. Here a correction factor is utilized that corresponds to the 0.9 ratio for Cauer/Ehmert in Table 2 and the Ehmert agreement with the UV method.

2.1.5. Cryotrapping
An unusual technique that was used by only one group was the cryotrapping of ozone on silica gel, with subsequent distillation to remove NO₂, and analysis for ozone by KI and UV methods (Edgar and Paneth, 1941a). This provided a sound measurement method and atmospheric ozone concentrations were obtained over the UK for multiple days (Edgar and Paneth, 1941b).

2.2. Other methods of measuring tropospheric ozone
2.2.1. Early ultraviolet absorption methods (1929–1960)
The ultraviolet absorption method for measuring ozone is based on the strong optical absorption of ozone in the Hartley (200–300 nm) and Huggins (320–360 nm) bands. Measurements involve an artificial light source (a mercury or hydrogen lamp) and a prism-based spectrograph with a detector, typically a photographic plate, that records the intensity of light at multiple wavelengths. In early measurements, the source and detector were separated by a long atmospheric path, of hundreds of metres to several km. The long distances were required to obtain adequate attenuation of the UV radiation, owing to sensitivity issues with the detectors. The ozone concentration was calculated from the ratio of intensities measured, at night, at long and short distances from the light source (Fabry, 1950). This measurement of ozone was difficult and the uncertainties involved (e.g. Kay 1953) do not appear to be quantified.

Some 15 sets of measurements of ozone at the Earth’s surface and in the free troposphere were made with the UV absorption technique between 1929 and 1960 (Fabry and Buisson, 1931; Götz and Ladenberg, 1931; Götz and Maier-Leibnitz, 1933; Chalonge et al., 1934; Chalonge and Vassy, 1934; Regener and Regener, 1934; Dauvillier, 1935; Götz et al., 1935; O’Brien et al., 1936; reported in Craig, 1950; Regener, 1938a; Coblenz and Stair, 1939, 1941; Götz and Penndorf, 1941; Vassy, 1941, data reported in Fabry, 1950; Rasool et al. 1956; Paetzold, 1959). The wavelengths used span the Hartley and Huggins bands.

Because the early UV methods are in principle the same as current methods, the main issue for comparisons of ozone data is identifying the wavelengths and absorption cross-sections used. The values used for ozone absorption cross-sections in the Hartley and Huggins bands have changed with time, particularly up to the time of the determination of Hearn (1961).

To illustrate these variations, Table 3 extends in time and expands in wavelengths the information presented in the final table of Hearn (1961). Ozone absorption cross-sections are presented in units of cm² molecule⁻¹, with earlier units converted. Between 1913 and 2015 there is a 20% range in absorption cross-sections at 253.7 nm and a range that varies from <1% to nearly 30% at wavelengths up to 334.2 nm. These wavelengths cover the range used to measure tropospheric ozone by the UV method between 1929 and 1960.

Early UV ozone measurements discussed in the following sections are corrected for UV cross-sections to the current standard (Hearn, 1961), following Table 3.

2.2.2. Recent ultraviolet absorption method (1970–present)
In the early 1970s there was a revolution in tropospheric ozone measurements. A newly developed UV absorption photometer (Bowman and Horak 1972) provided a stable, low-maintenance continuous ozone analyser based on an absolute method of measuring ozone with an incremental sensitivity of 1 nmol mol⁻¹. UV photometers measure the UV light absorption in the Hartley band (220–310 nm) where ozone is a strong absorber. Usually, a mercury lamp emitting light at 253.7 nm is used as the light source.

UV absorption photometers are currently the most commonly used method for in-situ ozone observations. The method fulfils well the requirements of analyser performance such as signal-to-noise ratio, detection limit, stability of sensitivity, and negligible interferences when measuring in clean air (Galbally et al., 2013). Moreover, it requires little maintenance during operation. Consequently, it is also the recommended technique for continuous ozone observations in ambient air, e.g. in the WMO Global Atmosphere Watch Programme (Galbally et al., 2013), in the United States (US EPA, 2013), Europe (European Union, 2012), and India (Central Pollution Control Board, 2009). Data quality and consistency of records within and across networks have improved over time, as seen by comparison of time series (Logan et al., 2012; Parrish et al., 2012), and station audits by the World Calibration Centre for Surface Ozone, Carbon Monoxide, Methane and Carbon Dioxide (WCC-Empa) (Buchmann et al., 2009). The great majority of in-situ records from more than 9000 sites in the TOAR database that have operated for 3 years or more between approximately 1975 and 2015 were recorded with UV absorption photometers.

A comprehensive uncertainty evaluation was undertaken by the WMO in the Guidelines for Continuous Measurement of Ozone in the Troposphere (Galbally et al., 2013). There, the total expanded (95% confidence) measurement uncertainty for UV absorption photometers was estimated to be

$$
\sigma(O_3) = 2 \cdot \sqrt{(0.81)^2 + (0.0089 \cdot O_3)^2} \text{ nmol mol}^{-1}
$$

(4)
Table 3: (a) Absorption cross-sections of ozone \([\times 10^{20} \text{ cm}^2 \text{ molecule}^{-1}]\) at selected Hg line wavelengths (in air). (b) Ratios relative to Gorshelev et al. (2014). DOI: https://doi.org/10.1525/elementa.376.t3

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>253.65</th>
<th>263.30</th>
<th>265.20</th>
<th>269.90</th>
<th>275.20</th>
<th>280.40</th>
<th>289.36</th>
<th>292.50</th>
<th>296.73</th>
<th>302.15</th>
<th>302.50</th>
<th>312.50</th>
<th>334.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabry and Buisson (1931)</td>
<td>1046</td>
<td>917.0</td>
<td>–</td>
<td>779.9</td>
<td>582.8</td>
<td>36.17</td>
<td>175.7</td>
<td>116.6</td>
<td>66.85</td>
<td>–</td>
<td>30.85</td>
<td>80.56</td>
<td>0.514</td>
</tr>
<tr>
<td>Läuchli (1928)</td>
<td>1277</td>
<td>–</td>
<td>1054</td>
<td>–</td>
<td>–</td>
<td>390.8</td>
<td>–</td>
<td>–</td>
<td>59.13</td>
<td>–</td>
<td>–</td>
<td>8.227</td>
<td>0.600</td>
</tr>
<tr>
<td>Ny and Choong (1933)</td>
<td>1222</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>185.1</td>
<td>–</td>
<td>70.19</td>
<td>31.88</td>
<td>–</td>
<td>–</td>
<td>0.574</td>
</tr>
<tr>
<td>Vigroux (1953)</td>
<td>1065</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>156.0</td>
<td>–</td>
<td>58.53</td>
<td>28.20</td>
<td>–</td>
<td>–</td>
<td>0.497</td>
</tr>
<tr>
<td>Inn &amp; Tanaka (1953)</td>
<td>1141</td>
<td>990.7</td>
<td>955.6</td>
<td>784.2</td>
<td>570.8</td>
<td>366.8</td>
<td>146.5</td>
<td>101.1</td>
<td>57.59</td>
<td>28.45</td>
<td>27.00</td>
<td>7.584</td>
<td>0.523</td>
</tr>
<tr>
<td>Hearn (1961)</td>
<td>1148</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>147.4</td>
<td>–</td>
<td>59.72</td>
<td>28.62</td>
<td>–</td>
<td>–</td>
<td>0.476</td>
</tr>
<tr>
<td>Mauersberger et al. (1986)</td>
<td>1137</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Malicet et al. (1995)</td>
<td>1131</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>151.3</td>
<td>–</td>
<td>61.46</td>
<td>29.85</td>
<td>–</td>
<td>–</td>
<td>0.466</td>
</tr>
<tr>
<td>Gorshelev et al. (2014)</td>
<td>1122</td>
<td>988.9</td>
<td>939.1</td>
<td>806.2</td>
<td>582.5</td>
<td>386.8</td>
<td>153.8</td>
<td>106.0</td>
<td>61.59</td>
<td>30.36</td>
<td>28.58</td>
<td>7.279</td>
<td>0.493</td>
</tr>
<tr>
<td>Viallon et al. (2015)</td>
<td>1127</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hodges et al. (2019)</td>
<td>1133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>253.65</th>
<th>263.30</th>
<th>265.20</th>
<th>269.90</th>
<th>275.20</th>
<th>280.40</th>
<th>289.36</th>
<th>292.50</th>
<th>296.73</th>
<th>302.15</th>
<th>302.50</th>
<th>312.50</th>
<th>334.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabry and Buisson (1931)</td>
<td>0.93</td>
<td>0.93</td>
<td>0.97</td>
<td>1.00</td>
<td>0.94</td>
<td>1.14</td>
<td>1.10</td>
<td>1.09</td>
<td>–</td>
<td>1.08</td>
<td>1.11</td>
<td>1.04</td>
<td>0.97</td>
</tr>
<tr>
<td>Läuchli (1928)</td>
<td>1.14</td>
<td>–</td>
<td>1.12</td>
<td>–</td>
<td>–</td>
<td>1.01</td>
<td>–</td>
<td>–</td>
<td>0.96</td>
<td>–</td>
<td>–</td>
<td>1.13</td>
<td>1.22</td>
</tr>
<tr>
<td>Ny and Choong (1933)</td>
<td>1.09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.20</td>
<td>–</td>
<td>1.14</td>
<td>1.05</td>
<td>–</td>
<td>–</td>
<td>1.16</td>
</tr>
<tr>
<td>Vigroux (1953)</td>
<td>0.95</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.01</td>
<td>–</td>
<td>0.95</td>
<td>0.93</td>
<td>–</td>
<td>–</td>
<td>1.01</td>
</tr>
<tr>
<td>Inn &amp; Tanaka (1953)</td>
<td>1.02</td>
<td>1.00</td>
<td>1.02</td>
<td>0.97</td>
<td>0.98</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>1.04</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Hearn (1961)</td>
<td>1.02</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.96</td>
<td>–</td>
<td>0.97</td>
<td>0.94</td>
<td>–</td>
<td>–</td>
<td>0.97</td>
</tr>
<tr>
<td>Mauersberger et al. (1986)</td>
<td>1.01</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Malicet et al. (1995)</td>
<td>1.01</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.98</td>
<td>–</td>
<td>1.00</td>
<td>0.98</td>
<td>–</td>
<td>–</td>
<td>0.95</td>
</tr>
<tr>
<td>Gorshelev et al. (2014)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Viallon et al. (2015)</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hodges et al. (2019)</td>
<td>1.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
with the ozone reading of the analyser given in nmol mol$^{-1}$. Thus, at a mean level of 30 nmol mol$^{-1}$ the total expanded uncertainty (95% confidence) is ± 1.7 nmol mol$^{-1}$.

A comparison of this bottom-up uncertainty analysis with field results is shown in Figure 1 with the results of 559 calibrations of UV absorption photometers in the Swiss National Air Pollution Monitoring Network. If the assumptions for the uncertainty estimation are correct, 95% of the calibration results shown in Figure 1 should be found within the central part surrounded by the grey shaded area. In fact, in Figure 1 less than 1.5% of the observed calibration results lie outside the estimated uncertainty range, so for this extensive dataset of UV absorption photometer field calibrations, the results fall well within the uncertainty estimate (4). This reflects the robust nature of properly undertaken ozone measurements with the UV photometric method.

The foregoing analysis does not cover all sources of uncertainty, omitting those associated with such causes as inlet losses and potential interferences. The uncertainty of the absorption cross-section is not included, as the calibrations are performed with primary standard analysers based on the same UV absorption cross-section. For the purposes of the subsequent analyses, at a mean level of 30 nmol mol$^{-1}$, the total expanded uncertainty (95% confidence) in modern ozone measurements is <2 nmol mol$^{-1}$.

### Figure 1: Intercept vs. slope plot for 559 calibrations of various ozone analysers with transfer standards within the Swiss National Air Pollution Monitoring Network between November 2005 and April 2017 for two different types of UV absorption photometers. Green and purple lines illustrate the slopes for 0 and 100 nmol mol$^{-1}$; the grey lines denote maximum uncertainties for ozone mole fractions from 0 to 100 nmol mol$^{-1}$, i.e. slope-intercept combinations within the grey lines correspond to deviations smaller than the maximum uncertainty. The maximum uncertainties for 0 and 100 nmol mol$^{-1}$ are highlighted in green and purple. DOI: https://doi.org/10.1525/elementa.376.f1

2.2.3. Chemiluminescent methods of measuring tropospheric ozone

There are multiple chemiluminescent methods for measuring tropospheric ozone, of which the methods involving ozone reaction with either ethylene, rhodamine B or nitric oxide are relevant to this paper.

The ethylene chemiluminescent ozone analyser (Warren and Babcock, 1970) is based on the reaction of ozone and ethylene, which produces electronically excited formaldehyde. As this formaldehyde returns to the ground state, light is emitted in a band centred at 430 nm, which is detected by a photomultiplier. The count rate varies linearly with ozone concentration, provided the cell pressure and temperature, the gain of the detector (the photomultiplier tube), the ethylene and sample flows and the composition of other components of the air sample are unchanged. The ethylene chemiluminescent ozone analyser is a sensitive (±1 nmol mol$^{-1}$) and stable instrument with a fast response (~1 s) and importantly, is not subject to SO$_2$ or NO$_2$ interference. The analyser requires regular calibration with a standard ozone analyser, which in the 1970s was either a KI or UV instrument. When calibrated, the ethylene chemiluminescent analyser gives results that closely match those of the ultraviolet method as seen in the intercomparisons at Hohenpeissenberg, Germany (Attmannspacher and Hartmannsgruber 1982) and Cape Grim, Australia (Elsworth and Galbally 1984). These analysers were used widely in North America in the 1970s (Heidorn and Yap, 1986), following the discovery that tropospheric ozone was damaging to tobacco crops (Macdowall et al., 1964; Mukammal, 1965). Although not in common use currently, in the TOAR surface ozone database chemiluminescence is listed as the method of ozone measurement at 627 of more than 9000 sites.

The rhodamine-B solid chemiluminescent ozone analyser was developed as an ozonesonde (Regener 1960) which was calibrated both with a surface based analyser, traceable to a UV measurement, prior to launch and against the total column ozone measured by a UV (Dobson) instrument at the same location and near the time of the ozonesonde flight. This system was also developed into an aircraft and surface based analyser (Regener, 1964). The analyser was regularly calibrated with an internal ozone generator that in turn had been calibrated against UV based ozone measurements (Regener, 1964). This chemiluminescent ozone analyser was used in the 1960s in the USA and Antarctica. The Regener ozonesonde was subject to large changes in sensitivity due to the effects of ozone, humidity and temperature (Regener 1964; Chatfield and Harrison, 1977; Hering and Dütsch, 1965) and the surface-based analyser may have had similar deficiencies.

In the nitric oxide chemiluminescent ozone analyser (Fontijn et al. 1970; Stedman et al. 1972) the reaction of ozone and nitric oxide produces electronically excited nitrogen dioxide. As the nitrogen dioxide returns to the ground state, it emits light with an intensity maximum at 1200 nm and a range between 600 and 3000 nm. This light is detected by a photomultiplier. The count rate varies linearly with ozone concentration, provided the cell pressure and temperature, the gain of the detector (the
photomultiplier tube), the nitric oxide and sample flows and the composition of other components of the air sample are unchanged. The nitric oxide chemiluminescent ozone analyser is a stable instrument with a sensitive (∼0.1 nmol mol⁻¹) and fast response (∼1 s) and not subject to SO₂ or NO₃ interference. A current description of the instrument is available (Campos et al., 2006). The analyser requires regular calibration with a standard ozone analyser. There are no records in the TOAR surface ozone database that have employed nitric oxide-based chemiluminescence as the method of measurement, but these analysers have been used extensively in aircraft measurements of tropospheric ozone (Lenschow et al. 1980) for more than 20 years (Section 4.7).

2.2.4 Differential Optical Absorption Spectrometry (DOAS)
The DOAS technique measures the spectrally-resolved absorption features in a beam of light that is returned by a retro-reflector located at some distance from the instrument (Platt et al. 1979). A telescope is used to send and receive a beam of white light, typically from a Xenon arc lamp. A photodiode array detector is used for simultaneous detection of the UV spectrum. The absorption features measured in the returning light beam are a convolution of all the absorption bands of molecules present in the beam path. The concentrations of ozone and other absorbing species are extracted based on well-characterized absorption cross-section data. The precision of a DOAS system for O₃ is estimated at 3% due to uncertainties of the absorption cross section (∼1%), and stray light in the spectrometer (∼2–3%), while noise and unexplained spectral structures determine biases and detection limits (2–4 nmol mol⁻¹), which scale inversely with the path length (Stutz and Platt, 1996). A field study comparison of DOAS with UV photometric analysers found differences of ±7%, attributed to spatial and temporal atmospheric inhomogeneities, including the fact that the DOAS beam scanned over a path higher in altitude than the sampling point of the UV instrument (Williams et al., 2006). In the TOAR database DOAS is listed as the method of ozone measurement at 39 sites.

3. Ozone measurements in surface air 1870s–1970s
For the purpose of reconstructing the historical surface ozone record an extensive literature search was conducted, which unearthed several data sets that have not before been coherently analysed in a single study. As discussed in Section 2, the approach used in TOAR-Observations is to reconstruct regional or zonal average ozone values based on all available historical data sets rather than relying on individual data sets as in previous studies. Following is a description of the application of the four data selection criteria for the historical reconstruction.

Criterion 1: all measurements via the Schönbein and related filter paper methods are rejected; the early UV measurements (1929–1960) are corrected to the currently accepted values of absorption cross-sections (Hearn, 1961); the observations made by the Mast Ozone Meter and related transmogrifier, and the Cauer method have been adjusted following Table 2.

Criterion 2: It is known that SO₂ causes quantitative negative interference with the KI measurement methods (Albert-Lévy, 1907; Glückauf, 1941, 1944; Saltzman and Gilbert 1959a), with a stoichiometry reported as 1.0 (Schenkel and Broder, 1982; Volz and Kley, 1988). SO₂ is known to be present in very low concentrations in the background atmosphere of <1 nmol mol⁻¹ (Seinfeld and Pandis 2000), thus not a significant interferent there. However in urban areas in Europe following the industrial revolution coal burning was widespread (Mylona, 1996; Smith et al., 2011) and the resulting sulfur-based acid pollution is widely documented (Smith, 1872; Ladureau, 1883, Witz, 1885). This criterion can lead to some ambiguity, as “clean” sites such as Arkona, on the Baltic coast, and the hilltop site of Hohenpeissenberg in southern Germany are subject to modest levels of SO₂ interference (Feister and Warmbt, 1987; Low et al., 1990, 1991). Thus KI-based ozone measurements need to be scrutinised for possible SO₂ interference before being accepted. The interference should be quantified and small compared with the ozone signal, and the uncertainty introduced into the ozone reading due to correction should be ideally ≤5%.

Criterion 3: The ozone sampled in the air in the surface layer should be representative of the unpolluted planetary boundary layer. At times of good turbulent mixing in the lower atmosphere, due to either convection driven by solar radiation or mechanical turbulence from wind shear, vertical gradients of ozone mole fraction diminish and ozone levels in near surface air are representative of the planetary boundary layer (Auer 1939; Glückauf 1944; Teichert 1955; Galbally and Derwent 1979; Fabian and Pruchniewicz, 1977; Galbally et al. 1986). At rural sites with plant and soil surfaces in flat plains away from fresh anthropogenic sources of NO, when turbulence diminishes at night, so that the rate of ozone supply from above is less than that lost due to ozone destruction, i.e. dry deposition, at the underlying surface, a night-time decrease of ozone occurs (Auer 1939; Glückauf 1944; Teichert 1955; Galbally 1968, 1972; Garland and Derwent 1979; Fabian and Pruchniewicz, 1977; Galbally et al. 1986). Consequently, nighttime measurements or 24-h average measurements of surface ozone at such continental sites are not representative of the well mixed boundary layer. Over ice, snow and water surfaces, this ozone decrease with weak mixing does not occur because the rate of ozone loss at the underlying surface is much smaller. The diurnal cycle is different at mountain top sites in steep terrain. At night the slopes cool and typically a downslope wind develops that entrains air from above, as described for Mauna Loa (Price and Pales, 1963). During the day the slopes warm and, away from urban pollution, an upslope wind brings air from below that has been depleted in ozone. Data selected in this study take into account topography and surface characteristics to ensure that only data from well mixed conditions, or conditions reasonably assumed to be well mixed, are selected.

Criterion 4: 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) said, the results are absurd. Serious artifacts and inconsistencies associated with individual datasets are discussed in the Supplemental Material (Text S-3).

For every site, the method used, the possible pollution sources, the topography, surface cover and prevailing meteorology and credibility of the results have been examined to assess the suitability of the observations for inclusion in the historical record. One excluded site is explicitly included in the following discussion. This is Montsouris, Paris, which is discussed here because it has been used as a central set of data for a number of previous historical reconstructions of tropospheric ozone. For the other historical data discussed below, in all cases the method is valid and with one exception, the sites are presumed free of interfering pollutants and representative of the well-mixed boundary layer. In a few cases some of the data are excluded on credibility grounds, and some are accepted with caution; the reasons for these decisions are discussed. Before discussing the available historical data which passed the selection criteria, it is worth repeating the cautionary note from Section 2: for comparison with current observations these data should be treated as a group, because there are substantial unknown uncertainties associated with the bias corrections of each individual set of observations.

There are some other features of the data that are important to note. Some studies involve measurements at multiple sites; other studies have long records at one site that have been broken into distinct periods. Thus the number of published studies and the number of data sets do not correspond. The 60 data sets accepted into the record, their site names, locations, data references etc. are presented in Tables 4, 5 and 6. During the period 1870s–1950, most of the available measurements were made using KI techniques with a small number of spectroscopic measurements by UV absorption. The studies were occasional scientific experiments, usually of limited duration and initially in central and northern Europe. From 1950 to the 1970s the coverage became global and measurements at a given location moved to yearly or greater duration.

The standard deviations presented here 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. 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. For more information see the Supplemental Material (Text S-5).

The observations are grouped into the tropical (0°–30°), temperate (30°–60°) and polar (60°–90°) regions of each hemisphere. They are presented in Tables 4–6 and discussed in the following text for different periods. Ozone is summarized for each region using a weighted mean, with the weighting proportional to the number of days with observations at each site.

3.1. The Montsouris Observatory ozone measurements 1876–1910

The ozone measurements at the Municipal Observatory at the Parc de Montsouris at the southern edge of Paris (Albert-Lévy, 1877) from 1876–1910, are the oldest active sampling ozone measurements known. As noted in Section 2.1.1, they are pivotal to conclusions in previous work that ozone in pre-industrial times was much less than its present concentration (e.g. Mareno et al., 1994). Hence the Montsouris observations are examined in detail here.

The technique used was the KI-arsenite system and is related to the UV method as described in Section 2.1.2, and is a valid method.

Volz and Kley (1988) examined daily measurements of ozone at Montsouris in 1876–86 and 1905 and correlated the results with the wind direction. The area to the southwest of Montsouris had no significant urban centers in 1876–1905 and Volz and Kley concluded that air from the southwest would be free of urban air pollutants which might interfere with the analysis. It was found that with wind from the southwest the average ozone levels measured in 1876–86 and 1905 were approximately 8 and 10 nmol mol⁻¹, and with wind from directions other than southwest the ozone levels were lower by 2 and 5 nmol mol⁻¹, respectively. The lower ozone levels observed when the wind direction was from Paris were attributed to the presence of SO₂ in the Paris urban plume. The corrected ozone data gave an average ozone level of 11 nmol mol⁻¹ over the period 1876–1910 (Volz and Kley 1988). This is much lower than that found at clean mid-latitude sites today.

Observers in the 19th century were aware of the potential for interference of reducing gases with the KI reaction, and, as noted by Volz and Kley (1988), an attempt was made to quantify these interferences. Average values for 1905–07 of “gaz réducteurs” correspond to ~3 nmol mol⁻¹, with weekly values as high as 16 nmol mol⁻¹ (Albert-Lévy, 1907, 1908). However, exactly how to interpret these measurements remains uncertain, as SO₂ does not react strongly with KI in the absence of ozone.

The period 1870s to 1900s was one of rapid change in Paris. During the period 1870–1880, Paris was a city of 2 million people and coal supplied 58% and wood 42% of the city’s total energy needs (Kim and Barles, 2012). Coal burning releases SO₂, and two approaches are taken to estimate SO₂ in Paris at this time. Firstly, measurements of ambient SO₂ in Paris began in the 1950s, and records of coal use in the city are available from 1875. These concentration and emission data are combined to estimate SO₂ levels of 55 nmol mol⁻¹ in Paris in the early 20th century (Ionescu et al., 2012). However, this approach neglects activities outside the city boundary where both building construction and the presence of gasworks that provided gas to Paris were located (Kim and Barles, 2012; Kesztembaum and Rosenthal 2014). Secondly, Montsouris measurements were made of sulphate in rainwater with an average of 13.9 mg l⁻¹ (range 3.5–37.0 mg l⁻¹) as SO₄²⁻; (Albert-Lévy, 1907, 1908). Sulphate in rainwater is closely associated with atmospheric sulphur dioxide, and such sulphate values are typical of highly polluted
areas and correspond to a $\text{SO}_2$ level of $\sim 25$–$75 \text{ nmol mol}^{-1}$ (e.g. Sequeira, 1975; Davies, 1979; Aas et al., 2007; Gonçalves et al., 2007). Both approaches indicate that $\text{SO}_2$ interference would have biased the Montsouris measurements low. The inferred $\text{SO}_2$ levels suggest a large degree of interference (Section 2.1), but in the absence of actual measurements there is inadequate knowledge to quantify the problem.

Other measurements at the Observatory (Albert-Lévy, 1877; Hartley, 1881) include on average $12 \text{ nmol mol}^{-1}$ of oxides of nitrogen (measured as nitric acid) during this period. This measured value is lower than the $28 \text{ nmol mol}^{-1}$ estimate based on coal and other fuel use in the early 20th century (Ionescu et al., 2012). However, it compares well with values of $\sim 8 \text{ nmol mol}^{-1}$ found in London by Reynolds (1930) and Edgar and Paneth (1941b). As nitrogen oxides are emitted as both NO and $\text{NO}_2$, they can interact with $\text{O}_3$ in the gas phase through NO titration removing ozone and $\text{NO}_2$ photolysis producing $\text{O}_3$. particularly if reactive organic compounds are present. Also the NO, can act as a positive interferent in the KI ozone method. Therefore, it is not immediately apparent whether the influence of these nitrogen oxides increased or decreased the measured ozone values.

Paris at the end of the 19th century was home to large numbers of horses and dairy cattle (Barles, 2012). Measurements of $\text{NH}_3$ were also made at the Observatory (Hartley 1881; Albert-Lévy, 1903); an average of $28 \text{ nmol mol}^{-1}$ is reported for the period 1883–1901. As already discussed, $\text{NH}_3$ is a negative interferent in the KI ozone measurement method, although the stoichiometry of the interference with the arsenite method is not known.

Overall the Montsouris observations fail Criterion 2 that the likelihood for significant contamination of the ozone measurements due to interfering pollutants in the sampled atmosphere should be low. Correction for $\text{SO}_2$ and other interferences is not feasible, due to their estimated magnitude, lack of a full understanding of the interference, and the absence of reliable atmospheric observations of the interfering compounds at Montsouris at that time.

Criterion 3 asks: are the measurements representative of the well-mixed boundary layer? The low sensitivity of the KI-arsenite method required 24-hour averages. The sampling was from a balcony of the Observatory building 5 m above the ground. The 24-hour averages at a clean air site would be biased low, being $0.8$ of daytime average and $0.7$ of daytime maximum measurements, given a typical diurnal cycle (Galbally and Roy, 1980) due to ozone dry deposition as discussed earlier. Furthermore wind speed and direction observations in Paris show an average diurnal pattern with northeasterly winds at night between 2000 and 0400 h, and southwesterly winds during the afternoon during 1100 to 1600 h between March and October 1890–1896 (Angot, late 1890s). These overnight winds would bring $\text{SO}_2$ enriched air to the Observatory causing chemical interference with the measurements, while cleaner air in the afternoon would have reduced the cumulative $\text{SO}_2$ effect.

At that time W.M. Hartley, user of the KI-arsenite method in his laboratory studies and a scientist familiar with the Montsouris work expressed concern over the possibility of ozone loss in the sampling system and wrote: “It is impossible, therefore, to accept the figures given in the Annuaire de L’Observatoire de Montsouris as indicating anything like the true proportion of ozone usually present in country air…” (Hartley, 1881). Indeed, the Montsouris measurements are consistent with other historical measurements in urban areas (Text S-6, Table S-1 and Figure S-1).

In summary, the measurements of ozone concentration at Montsouris were made with a valid measurement technique, however it is very likely that there is a large negative bias in the measurements, of comparable magnitude to the observed ozone concentration, due to the presence of $\text{SO}_2$ and ammonia and a further uncertain bias due to nitrogen oxides. Also the 24-hour sampling includes low night-time ozone values due to the occurrence of nocturnal inversions and ozone dry deposition as well as chemical interference due to the recirculation of air over Paris. Consequently, the measured ozone concentrations for Montsouris for 1876–1910 are biased low, are not representative of the regional atmosphere and are not used in this assessment of historical ozone concentrations.

3.2. Other surface measurements: 1896–1901
During 1896–1901 other surface ozone measurements, in clean air, were made with the KI-arsenite method (Albert-Lévy, 1877) and are listed in Table 4. De Thierry (1896) made measurements at Chamonix (~1 km asl) and Grands-Mulets (~3km asl) on Mont Blanc in southeastern France on 3 days in August and September 1896. Lespieau, (1906) made 13 measurements over 4 days in 1900 and 6 days in 1901 on glaciers at three levels (~1.25 km, ~3 km and ~4.8 km) on Mont Blanc. The mean of these observations, weighted by the number of measurement days, (hereinafter referred to as the ’weighted mean’), in the temperate zone over Europe is 25 nmol mol$^{-1}$ with a range of 20–63 nmol mol$^{-1}$ for the period 1896–1901.

3.3. 1929–1934
The first measurements of surface ozone with the UV method were made in 1929 (Fabry and Buisson 1931; Götz and Ladenberg 1931). In Europe there were multiple measurements with long path UV (Götz and Maier-Leibnitz, 1933; Chalonge and Vassy, 1934; Götz and Penndorf 1941). The study of Chalonge et al. (1934) involved two groups making simultaneous UV measurements at two alpine sites, swapping sites mid-way during the experiment. No significant differences were observed in this early instrument intercomparison. The ozone mixing ratio, corrected for the Fabry and Buisson (1931) UV absorption coefficients, was 31 nmol mol$^{-1}$ at Jungfraujoch, Switzerland (3450 m) and 18 nmol mol$^{-1}$ at nearby Lauterbrunnen (800 m), showing an increase with altitude.

The weighted mean of these observations in the temperate zone over Europe is 25 nmol mol$^{-1}$ with a range from 18–32 nmol mol$^{-1}$ for the period 1929–1934.
<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Value(s)</th>
<th>Days</th>
<th>Mean (nmol mol⁻¹)</th>
<th>SD (nmol mol⁻¹)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1896–1901 Northern Temperate – Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Chamonix 45.9N, 6.9E, 1050 m</td>
<td>August 1896</td>
<td>KI-arsenite</td>
<td>De Thierry (1897)</td>
<td>35–39 µg m⁻³</td>
<td>2</td>
<td>20.4</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>2 Grands-Mulets 45.9N, 6.9E, 3020 m</td>
<td>September 1896</td>
<td>KI-arsenite</td>
<td>De Thierry (1897)</td>
<td>94 µg m⁻³</td>
<td>1</td>
<td>63.0</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>3 Glacier des Bossons 45.9N, 6.9E, 1250 m</td>
<td>August 1900, 1901</td>
<td>KI-arsenite</td>
<td>Lespieau, 1906</td>
<td>41 µg m⁻³</td>
<td>5</td>
<td>22.9</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>4 Grands-Mulets 45.9N, 6.9E, 2640 m</td>
<td>August 1900, 1901</td>
<td>KI-arsenite</td>
<td>Lespieau, 1906</td>
<td>33 µg m⁻³</td>
<td>2</td>
<td>22.3</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>5 Sommet, Mont Blanc 45.9N, 6.9E, 4810 m</td>
<td>Sept 1900, Aug 1901</td>
<td>KI-arsenite</td>
<td>Lespieau, 1906</td>
<td>26 µg m⁻³</td>
<td>3</td>
<td>21.5</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

Weighted mean 25.2

1929–1934 Northern Temperate – Europe

<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Value(s)</th>
<th>Days</th>
<th>Mean (nmol mol⁻¹)</th>
<th>SD (nmol mol⁻¹)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Provence 44.2N, 5.2E, 300 m</td>
<td>Oct. 1929–Mar. 1930</td>
<td>UV</td>
<td>Fabry and Buisson, 1931</td>
<td>22 nmol mol⁻¹</td>
<td>1</td>
<td>22.4</td>
<td>n/a</td>
<td>1</td>
</tr>
<tr>
<td>7 Arosa 46.8N, 9.7E, 1860 m</td>
<td>April 30, 1930</td>
<td>UV</td>
<td>Götz and Ladenburg (1931)</td>
<td>29 nmol mol⁻¹</td>
<td>1</td>
<td>32.2</td>
<td>n/a</td>
<td>2</td>
</tr>
<tr>
<td>8 Arosa</td>
<td>Sept. 1932–Apr. 1933</td>
<td>UV</td>
<td>Götz and Maier-Leibnitz (1933)</td>
<td>8–31 nmol mol⁻¹ (L1929); 9–36 nmol mol⁻¹ (FB1931)</td>
<td>5</td>
<td>24.9</td>
<td>10.5</td>
<td>3</td>
</tr>
<tr>
<td>9 Arosa</td>
<td>March–May 1934</td>
<td>UV</td>
<td>Götz &amp; Penndorf (1941)</td>
<td>22–32 nmol mol⁻¹</td>
<td>11</td>
<td>28.2</td>
<td>2.8</td>
<td>4</td>
</tr>
<tr>
<td>10 Chur 46.9N, 9.5E, 600 m</td>
<td>Sept. 1932–April 1933</td>
<td>UV</td>
<td>Götz and Maier-Leibnitz (1933)</td>
<td>18–19 nmol mol⁻¹ (L1929); 21 nmol mol⁻¹ (FB1931)</td>
<td>2</td>
<td>20.4</td>
<td>n/a</td>
<td>3</td>
</tr>
<tr>
<td>11 Lauterbrunnen 46.6N, 7.9E, 800 m</td>
<td>18–31 August 1933</td>
<td>UV</td>
<td>Chalonge, Gotz &amp; Vassy (1934)</td>
<td>11–26 nmol mol⁻¹</td>
<td>10</td>
<td>17.5</td>
<td>4.5</td>
<td>5,1</td>
</tr>
<tr>
<td>12 Jungfraujoch 46.6N, 8.0E, 3450 m</td>
<td>18–31 August 1933</td>
<td>UV</td>
<td>Chalonge, Gotz &amp; Vassy (1934)</td>
<td>24–35 nmol mol⁻¹</td>
<td>7</td>
<td>30.9</td>
<td>5.6</td>
<td>5,1</td>
</tr>
</tbody>
</table>

Weighted mean 24.9

(Contd.)
<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Value(s)</th>
<th>Days</th>
<th>Mean (nmol mol⁻¹)</th>
<th>SD  (nmol mol⁻¹)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 Mt Ventoux, 44.2N, 5.3E, 1912 m</td>
<td>October, 1938</td>
<td>UV</td>
<td>Vassy, 1941</td>
<td>25 nmol mol⁻¹</td>
<td>1</td>
<td>25.5</td>
<td>n/a</td>
<td>6,1</td>
</tr>
<tr>
<td>14 Friedrichshafen, 47.7N, 9.5E, 400 m</td>
<td>8–12 Sept. 1938</td>
<td>KI</td>
<td>Regener (1938b)</td>
<td>15–24 nmol mol⁻¹</td>
<td>4</td>
<td>21.0</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>15 Jungfraujoch, 47.5N, 9.7E, 1060 m</td>
<td>August 1938</td>
<td>KI</td>
<td>Regener (1938b)</td>
<td>24–43 nmol mol⁻¹</td>
<td>5</td>
<td>30.0</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>16 Pfänder Mountain, 54.8N, 16W, 100 m</td>
<td>September, 1940</td>
<td>Ehmert</td>
<td>Ehmert and Ehmert (1949); Volz et al. (1988)</td>
<td>22 nmol mol⁻¹</td>
<td>7</td>
<td>22.0</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>17 Durham, UK, 53.6N, 30W, 6 m</td>
<td>1940–1941</td>
<td>KI coulometric</td>
<td>Paneth and Glückauf 1941; Glückauf 1944</td>
<td>27 nmol mol⁻¹</td>
<td>20</td>
<td>27.0</td>
<td>n/a</td>
<td>8</td>
</tr>
<tr>
<td>18 Kew, 51.5N, 0.3W, 6 m</td>
<td>May–July 1938 &amp; 1939</td>
<td>Cryogenic, KI and UV analysis</td>
<td>Paneth and Glückauf 1941; Glückauf 1944</td>
<td>22–28 nmol mol⁻¹</td>
<td>3</td>
<td>24.3</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>19 Southport, 53.6N, 30W, 6 m</td>
<td>May–July 1938 &amp; 1939</td>
<td>Cryogenic, KI and UV analysis</td>
<td>Paneth and Glückauf 1941; Glückauf 1944</td>
<td>27 nmol mol⁻¹</td>
<td>20</td>
<td>27.0</td>
<td>n/a</td>
<td>8</td>
</tr>
</tbody>
</table>

Weighted mean 25.5

1950s to 1970s Northern Temperate – Europe, North Africa, North America, Japan

<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Value(s)</th>
<th>Days</th>
<th>Mean (µg m⁻³)</th>
<th>SD  (µg m⁻³)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Mt. Capillo, NM, 34.7N, 106.4W, 3100 m</td>
<td>Nov–Dec 1950</td>
<td>KI coulometric</td>
<td>Bow and Regener (1951)</td>
<td>44.6 ± 10.4 µg m⁻³</td>
<td>30</td>
<td>26.0</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>21 Arosa</td>
<td>Apr.1950–Jul.1951</td>
<td>Ehmert</td>
<td>Götz and Volz (1951)</td>
<td>44.6 ± 10.4 µg m⁻³</td>
<td>400</td>
<td>26.4</td>
<td>6.2</td>
<td>10</td>
</tr>
<tr>
<td>22 Arosa</td>
<td>Apr.1954–Oct.1958</td>
<td>Ehmert</td>
<td>Perl 1965</td>
<td>39.3 ± 10.8 µg m⁻³</td>
<td>1500</td>
<td>23.2</td>
<td>6.4</td>
<td>10</td>
</tr>
<tr>
<td>23 Lindenberg, 52.2N, 14.1E, 98 m</td>
<td>Aug–53; June–July 1954</td>
<td>Cauer</td>
<td>Tiechert (1955)</td>
<td>31 µg m⁻³</td>
<td>31</td>
<td>20.2</td>
<td>n/a</td>
<td>11</td>
</tr>
<tr>
<td>24 Arkona, 54.7N, 13.4E, 42 m</td>
<td>1956–1971</td>
<td>Cauer</td>
<td>Feister &amp; Warmbt (1987)</td>
<td>36.1 µg m⁻³</td>
<td>5500</td>
<td>22.7</td>
<td>4.7</td>
<td>12</td>
</tr>
<tr>
<td>25 Arkona, 54.7N, 13.4E, 42 m</td>
<td>1972–1984</td>
<td>Cauer</td>
<td>Feister &amp; Warmbt (1987)</td>
<td>44.5 µg m⁻³</td>
<td>4300</td>
<td>25.8</td>
<td>5.3</td>
<td>13</td>
</tr>
<tr>
<td>26 Hohenpeissenberg, 48N, 11E, 975 m</td>
<td>1971–1975</td>
<td>Wet chemical idometric</td>
<td>Atmannspacher et al. (1977)</td>
<td>1800</td>
<td>32.3</td>
<td>11.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Contd.)
<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Value(s)</th>
<th>Days</th>
<th>Mean (nmol mol⁻¹)</th>
<th>SD (nmol mol⁻¹)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>27 Hohenpeissenberg</td>
<td>1969–1972</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>64.8 µg m⁻³</td>
<td>960</td>
<td>30.3</td>
<td>14.3</td>
<td>14</td>
</tr>
<tr>
<td>28 Norderney</td>
<td>1969–1975</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>41.9 µg m⁻³</td>
<td>2000</td>
<td>19.6</td>
<td>9.0</td>
<td>14</td>
</tr>
<tr>
<td>29 Zugspitze</td>
<td>1969–1975</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>45.1 µg m⁻³</td>
<td>1920</td>
<td>21.1</td>
<td>5.9</td>
<td>14</td>
</tr>
<tr>
<td>30 Cagliari</td>
<td>1970–1975</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>41 µg m⁻³</td>
<td>1480</td>
<td>19.2</td>
<td>7.2</td>
<td>14</td>
</tr>
<tr>
<td>31 Westerland</td>
<td>1971–1975</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>47.1 µg m⁻³</td>
<td>1240</td>
<td>22.0</td>
<td>5.8</td>
<td>14</td>
</tr>
<tr>
<td>32 Kairouan</td>
<td>1970–1975</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>56.1 µg m⁻³</td>
<td>1336</td>
<td>26.2</td>
<td>9.0</td>
<td>14</td>
</tr>
<tr>
<td>33 Mt Norikura</td>
<td>Sept 1959; Aug, Sept 1960</td>
<td>Ehmert</td>
<td>Miyake et al. (1962)</td>
<td>48.7 µg m⁻³</td>
<td>18</td>
<td>22.8</td>
<td>9.7</td>
<td>22</td>
</tr>
<tr>
<td>34 Mt Norikura</td>
<td>Sept 1959; Aug, Sept 1960</td>
<td>Ehmert</td>
<td>Miyake et al. (1962)</td>
<td>35.7 µg m⁻³</td>
<td>18</td>
<td>16.7</td>
<td>9.1</td>
<td>22</td>
</tr>
<tr>
<td>35 Tomisaki</td>
<td>Aug, Sept 1963, Feb 1965</td>
<td>Ehmert</td>
<td>Kawamura and Sakurai (1966)</td>
<td>44.4 µg m⁻³</td>
<td>23</td>
<td>20.7</td>
<td>6.7</td>
<td>22</td>
</tr>
<tr>
<td>Weighted mean</td>
<td></td>
<td></td>
<td></td>
<td>24.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dauvillier (1934) used the KI-arsenite method at Scorsby Sund on the east coast of Greenland, from November 1932 to August 1933 (Table 5). His daily data (reported as 24-hour means) can be notionally divided into two sets: 225 days of data show a background value of about 50 µg m⁻³, with less than 5% over 100 µg m⁻³ (approximately 47 nmol mol⁻¹), and 45 days of data that show major events where ozone went as high as 570 µg m⁻³. The latter maximum is approximately 270 nmol mol⁻¹. Ozone during December 1932 was particularly high, averaging approximately 100 nmol mol⁻¹. This monthly mean is more than twice the highest December monthly mean at Arctic sites in the TOAR Surface Ozone Database. These high events were recorded by Dauvillier (1934) as interesting and remarkable. He associated them with aurora and suggested enhanced transport from the stratosphere or ionosphere to the surface air. Dauvillier (1934) also states that during the winter the inlet was close to or at the snow surface. There are two possible explanations of these high ozone levels, either (a) the KI-arsenite method is subject to large positive interference in the recorded calm conditions in the polar night or (b) there was enhanced stratospheric-tropospheric exchange or tropospheric ozone production that led to high ozone levels in surface air, that has not recurred in the Arctic during the period of UV surface ozone measurements, approximately 1980 to the present. Two points are worth noting: Dauvillier (1935) subsequently validated his method in the Arctic in winter against UV surface ozone observations; also, the ozonesonde record from Resolute Bay, in the Canadian Arctic, shows occasional very high values at the surface, only in winter, including a value of 164 nmol mol⁻¹ in 1966, and Chung and Dann (1985) record surface ozone levels measured with an ethylene chemiluminescent analyser in December in Saskatchewan, Canada of up to 228 nmol mol⁻¹. With the evidence available it is impossible to resolve which explanation is correct. The considered judgement is that the 45 days of observations during the major events are not credible and omitted from the subsequent analysis, while the 225 days of background values are retained as credible. Dauvillier (1935) presents measurements from Abisko, Sweden between December 1934 and March 1935, giving average ozone concentrations of 41 µg m⁻³ from 68 UV measurements and 33 µg m⁻³ from 56 measurements with the KI-arsenite method.

Considering the qualifications above, the methods used are valid, the sites are presumed free of interfering pollutants and representative of the well-mixed boundary layer. The weighted mean of these observations in the Northern Polar region is 22 nmol mol⁻¹ with a range of 15–24 nmol mol⁻¹ for the period 1932–1935.

3.4. 1938–1941
The only UV measurements of surface ozone found during this period were at Mt. Ventoux (1912 m, southern France) by A. Vassy in October 1938, as reported by Fabry (1950). After correction for the Fabry and Buissou (1931) UV absorption coefficients these observations average 26 nmol mol⁻¹.

The KI method became more widely used during this period. At Jungfraujoch (3450 m) on 5 days in August 1938, Regener (1938b) observed 30 (range 24–43) nmol mol⁻¹ ozone, and in September at Friedrichshafen (at 400 m near the shore of Lake Constance, southern Germany), 21 nmol mol⁻¹ (range 15–24). Ehmert and Ehmert (1949) made measurements using the Ehmert method on Pfänder Mountain (1060 m in western Austria) in September, 1940. The data have been reinterpreted by Volz et al. (1988), where the original ozone measurements gave 15 nmol mol⁻¹ and the revised value is 22 nmol mol⁻¹.

Edgar and Paneth (1941b) present measurements of ozone in the air in a street of South Kensington, London, UK, from the rooftop of the Royal College of Science building in South Kensington, at the Kew Observatory outside London and at Southport on the northwest coast of England, made from February 1938 to July 1939. They used a cryogenic trapping and purification method followed by both UV and KI analysis (Edgar and Paneth, 1941a; 1941b). Considering only the non-urban observations at Kew and Southport, the mean and range are 24 (17–29) nmol mol⁻¹.

Glückauf (1944), using an automated KI-thiosulfate method, presents occasional ozone data in a meteorological analysis of observations near Durham, UK. The following results are extracted from his paper. On 21 March 1941 in wind from a clean sector, the ozone mixing ratio was 31 ± 2 nmol mol⁻¹. The level in November 1940 is reported as half this. Daily maximum values varied from 24 nmol mol⁻¹ in November to 68 nmol mol⁻¹ in May. On 28 and 29 August 1941, in prolonged strong steady winds, ozone reached daily maxima of 25 and 27 nmol mol⁻¹ with a 37-hour average of 21.5 nmol mol⁻¹. Analyses of five warm fronts indicated ozone mole fractions of 25 to 34 nmol mol⁻¹. Analyses of four cold fronts indicated ozone mole fractions of 27 to 68 nmol mol⁻¹. The high values, likely from subsidence behind the front, may have been partly of stratospheric origin. If the maximum values are treated as 95th percentiles and an annual cycle is fit to the monthly observations then the mean mole fraction of ozone in clean, well-mixed conditions is 27 nmol mol⁻¹.

Accepting the specific analysis above of Glückauf (1944), for this period the methods used are valid, the sites, or air masses sampled, are presumed free of interfering pollutants and representative of the well mixed boundary layer. The weighted mean of these observations in the temperate zone over Europe is 26 nmol mol⁻¹ with a range from 21–30 nmol mol⁻¹ for the period 1938–1941. These results are indistinguishable from those both 5 years and 4 decades earlier.

3.5. 1951–1970s
The period 1950s to 1970s is the first period in which there were globally distributed measurements of ozone in surface air. Series of measurements lasting a year or more became common and short-term campaigns fewer. The surface ozone records during this period include: the upsurge in observations during the International Geophysical Year of 1957–1958; measurements made initially at Dresden-Wahnsdorf in 1954 and then expanded in
Table 5: Surface measurements in polar regions. Mean values and standard deviations in nmol mol\(^{-1}\). See Text S-7 for Notes. DOI: https://doi.org/10.1525/elementa.376.t5

<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Range</th>
<th>Days</th>
<th>Mean (nmol mol(^{-1}))</th>
<th>SD (nmol mol(^{-1}))</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1932–1935 Northern High Latitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 Scoresby Sund</td>
<td>Nov. 1932 to Aug. 1933</td>
<td>KI-arsenite</td>
<td>Dauvillier (1934)</td>
<td>50 µg m(^{-3})</td>
<td>251</td>
<td>23.5</td>
<td>8</td>
<td>15,9</td>
</tr>
<tr>
<td>37 Abisko, Sweden</td>
<td>22 Dec 1934–7 Mar 1935</td>
<td>KI-arsenite</td>
<td>Dauvillier (1935)</td>
<td>33 µg m(^{-3})</td>
<td>56</td>
<td>15.2</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>38 Abisko, Sweden</td>
<td>Dec.1934–Mar.1935</td>
<td>UV</td>
<td>Dauvillier (1935)</td>
<td>41 µg m(^{-3})</td>
<td>68</td>
<td>19.5</td>
<td>6</td>
<td>16,1</td>
</tr>
<tr>
<td>1950s to 1970s Northern High Latitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 Tromso</td>
<td>1970–1975</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>40.4 µg m(^{-3})</td>
<td>1480</td>
<td>18.9</td>
<td>2.6</td>
<td>14</td>
</tr>
<tr>
<td>40 Bredkalen</td>
<td>December 1968–January 1970</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>45.7 µg m(^{-3})</td>
<td>292</td>
<td>21.4</td>
<td>7.1</td>
<td>14</td>
</tr>
<tr>
<td>41 Kise</td>
<td>August 1970–November 1972</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>44.8 µg m(^{-3})</td>
<td>680</td>
<td>20.9</td>
<td>4.4</td>
<td>14</td>
</tr>
<tr>
<td>42 College, Alaska</td>
<td>August–December 1950</td>
<td>KI-arsenite</td>
<td>Wilson et al. (1952)</td>
<td>68 ± 27 µg m(^{-3})</td>
<td>149</td>
<td>31.4</td>
<td>12.7</td>
<td>17</td>
</tr>
<tr>
<td>43 Barrow, Alaska</td>
<td>January 1965–September 1967</td>
<td>Mast ozone meter</td>
<td>Kelley (1970)</td>
<td>34.2 ± 15.4 nmol mol(^{-1})</td>
<td>1003</td>
<td>34.2</td>
<td>15.4</td>
<td>17a</td>
</tr>
<tr>
<td>1950s to 1970s Southern High Latitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44 Little America</td>
<td>March 1957–October 1958</td>
<td>Bowen and Regener (1951)</td>
<td>Wexler et al. (1960)</td>
<td>43.3 µg m(^{-3})</td>
<td>443</td>
<td>19.2</td>
<td>7.1</td>
<td>18,19</td>
</tr>
<tr>
<td>45 Halley Bay</td>
<td>January–December 1958</td>
<td>transmogrifier</td>
<td>MacDowall (1962); also Roscoe and Roscoe (2006)</td>
<td>13.1 nmol mol(^{-1})</td>
<td>275</td>
<td>15.8</td>
<td>3.7</td>
<td>20</td>
</tr>
<tr>
<td>46 Amundsen-Scott</td>
<td>March 1961–July 1963</td>
<td>Bowen and Regener (1951)</td>
<td>Oltmans &amp; Komhyr (1976)</td>
<td>30.3 ± 4.8 nmol mol(^{-1})</td>
<td>780</td>
<td>30.3</td>
<td>4.8</td>
<td>19</td>
</tr>
<tr>
<td>48 Hallett, Antarctica</td>
<td>January 1962–February 1963</td>
<td>Regener chemiluminescent</td>
<td>Aldaz (1965)</td>
<td>19.1 ± 5.1 nmol mol(^{-1})</td>
<td>365</td>
<td>19.1</td>
<td>5.1</td>
<td>19</td>
</tr>
<tr>
<td>49 Base Roi Baudouin</td>
<td>April 1965–December 1966</td>
<td>Mast ozone meter</td>
<td>Wisse and Meerb (1969)</td>
<td>21.9 ± 5.6 nmol mol(^{-1})</td>
<td>600</td>
<td>26.3</td>
<td>6.7</td>
<td>19,20</td>
</tr>
</tbody>
</table>

Weighted mean

| Mean (nmol mol\(^{-1}\)) | 21.5 |
| Weighted mean            | 23.5 |
1956 to a network of 7 surface stations across Germany (Warmbt, 1964; Feister and Warmbt, 1987); and from 1969–1975 the Troposphärisches Ozon (TROZ) network of 16 surface ozone stations on a meridional zone from Norway to South Africa, run by the Max Planck Institute for Aeronomy (Fabian and Pruchniewicz, 1977). Several other sites utilized here were run on an individual or national network basis.

For the TROZ network, the largest network for this period, a coulometric KI method (Section 2.1.3) was used (Pruchniewicz 1970, 1973). Detailed consideration of the TROZ data reveals issues with (a) its traceability to the modern UV standard, (b) interferences or pollution at the sites, and (c) special features in selection of representative surface ozone data (reported ozone values are representative of daytime maximum values rather than daytime means). These issues are discussed in the Supplemental Material (Text S-3). Because this data set has data from 15 stations (one station was omitted for reasons described below), located between 90°’N and 30°’S, it potentially dominates the 1950s–1970s historical record. To understand the sensitivity of the historic analysis to the TROZ data and to provide robust conclusions, the quantification of the change in ozone from the historical to the modern period is performed both with and without the TROZ data (see Section 3.6).

Five surface ozone observing sites operating in Northern Polar region in the 1950s to 1970s are presented in Table 5. Three sites are from the TROZ network and two from Alaska: College (137 m, located in a suburb of Fairbanks in the center of the state) and Barrow (15 m, near the shore of the Beaufort Sea and now called Utqiagvik) (Wilson et al. 1952; Kelley 1973). The College data, taken with the KI-arsenite method, show spikes of very high values like the Dauvillier (1934) record; these are truncated on credibility grounds (see Text S-3 for detailed discussion). The Barrow data are well-documented, but they show odd seasonal behaviour and some surprisingly high values, and so have questionable credibility, but are accepted (Text S-3). Considering the qualifications above, the methods used are valid, the sites are presumed free of interfering pollutants and representative of the well mixed boundary layer. The weighted mean of the accepted observations in the Northern Polar region (1950s–1970s) is 24 nmol mol⁻¹ with a range of 19–34 nmol mol⁻¹. If the College and Barrow data are entirely omitted this average becomes 19 nmol mol⁻¹, but this value relies entirely upon the TROZ network.

In the Northern Temperate region there are multiple data sets. Bowen and Regener (1951) present ozone data from Capillo Peak Observatory in New Mexico (approximately 2800 m asl). Four years of ozone data were obtained at the Arosa Light Observatory (located on the northern edge of the town of Arosa in a Swiss valley, 1810 m (Staehelin et al., 2018)) using the Ehmert technique (Perl, 1965). In August 1953 and June/July and September 1954 ozone observations were made utilizing an 80 m tower using the Cauer method (Section 2.1.4) at Lindenberg (98 m) in northeastern Germany (Teichert 1955). Measurements were made at Hohenpeissenberg (975 m) from 1971 to 1975 using the HP-KI method. Six other sites, all from the TROZ network, are listed in Table 4. The TROZ station Lindau is omitted due to the influence of high levels of local pollution (Fabian and Pruchniewicz, 1976). The long-term site at Arkona on the north coast of Germany using the Cauer method commenced in 1956 (Warmbt 1964; Feister and Warmbt 1987). With the exception of Arkona, the methods used are valid, the sites are presumed free of interfering pollutants and representative of the well mixed boundary layer. With a correction for SO₂ interference (Warmbt 1964; Feister and Warmbt 1987), the Arkona data are accepted (see note #12 in Text S-7). Measurements from three sites in Japan made with the Ehmert method are included (Miyake et al. 1962; Kawamura and Sakurai 1966). The methods used are valid, and after the correction of the Arkona data, all sites are presumed free of the influence of interfering pollutants and representative of the well mixed boundary layer. The weighted mean of these observations in the Northern Temperate region is 22 nmol mol⁻¹ with a range of 19–32 nmol mol⁻¹.

In the Northern Tropics in the 1950’s to 1970’s surface ozone observing sites are Fort Lamy in Chad, Africa (12°’N), from the TROZ network (Fabian and Pruchniewicz, 1977), Pune (Poona) 19°’N and Ahmedabad, 24°’N in India (Tiwari and Sreedharan 1973; Naja and Lal 1996) and Mauna Loa, Hawaii 20°’N (Price and Pales, 1963) (Table 6). The weighted mean of these observations in the Northern Tropics is 23 nmol mol⁻¹ with a range from 16–31 nmol mol⁻¹ for the period 1950–1975.

In the Southern Tropics in the 1950’s to 1970’s surface ozone sites of Luanda, 9°’S, Sa da Bandeira 15°’S, Alexander Bay 28°’S and Windhoek 23°’S are from the TROZ network and all are located in Africa (Fabian and Pruchniewicz, 1977). The weighted mean of these observations in the Southern Tropics is 18 nmol mol⁻¹ with a range from 14–24 nmol mol⁻¹ for the period 1970–1975.

In the Southern Temperate region in the 1950’s to 1970’s surface ozone sites are Hermanus, South Africa 34°’S (Fabian and Pruchniewicz, 1977), 20 days measurements during campaigns at locations in south east Australia, −34°’S (Galbally, 1968, 1970, 1971, 1972) and two years of measurements at Macquarie Island 54°’S, south of New Zealand (Galbally and Roy, 1981). The weighted mean of these observations in the Southern Temperate region is 22 nmol mol⁻¹ with a range from 21–25 nmol mol⁻¹ for the period 1967–1975.

Six surface ozone observing sites in the Southern Polar region operated in this period, all in Antarctica. These are Little America (Wexler et al., 1960), Halley Bay (MacDowall, 1962), Hallett and Amundsen-Scott South Pole Station (Aldaz, 1965; Oltmans & Komhyr, 1976), Base Rio Baudouin (Wisse and Meerburg, 1969), and Mirny (Kolbig and Warmbt, 1978) (Table 5). The 1958 data from Halley Bay have questionable credibility (Roscoe and Roscoe 2006; see Text S-3). The Amundsen-Scott South Pole Station is at 2835 m altitude, and therefore it is expected, as observed, that the ozone levels there are higher than those at the lower altitude stations. There are several other sites at southern high latitudes that
<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Range</th>
<th>Days</th>
<th>Mean (nmol mol(^{-1}))</th>
<th>SD</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 Ahmedaabad, India</td>
<td>23N, 72.6E, 55 m</td>
<td>Ehmert</td>
<td>Naja and Lal (1996)</td>
<td>20 ± 6 nmol mol(^{-1})</td>
<td>360</td>
<td>20.0</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>51 Mauna Loa, Hawaii</td>
<td>19.5N, 150W, 3400 m</td>
<td>Bowen and Regener (1951)</td>
<td>Price and Pales (1963)</td>
<td>~30 nmol mol(^{-1}); winter low 20, spring high 50</td>
<td>670</td>
<td>29.9</td>
<td>7.4</td>
<td>10</td>
</tr>
<tr>
<td>52 Pune (Poona), India</td>
<td>18.5N, 73.5E, 560 m</td>
<td>September 1969–August 1970</td>
<td>Brewer bubbler</td>
<td>Tiwari and Sreedharan (1973)</td>
<td>1264</td>
<td>30.9</td>
<td>4.6</td>
<td>9,22,20</td>
</tr>
<tr>
<td>53 Fort Lamy</td>
<td>12.1N, 15.0E, 300 m</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>33.4 µg m(^{-3})</td>
<td>1264</td>
<td>15.6</td>
<td>3.7</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weighted mean</td>
<td></td>
<td>21.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 Alexander Bay</td>
<td>28.6S, 16.5E, 21 m</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>36.7 µg m(^{-3})</td>
<td>1360</td>
<td>17.1</td>
<td>6.5</td>
<td>14</td>
</tr>
<tr>
<td>55 Luanda</td>
<td>8.9S, 13.2E, 74 m</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>28.8 µg m(^{-3})</td>
<td>1360</td>
<td>13.5</td>
<td>4.5</td>
<td>14</td>
</tr>
<tr>
<td>56 Sa da Bandeira</td>
<td>14.9S, 13.6E, 1760 m</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>34.1 µg m(^{-3})</td>
<td>1144</td>
<td>15.9</td>
<td>8.5</td>
<td>14</td>
</tr>
<tr>
<td>57 Windhoek</td>
<td>22.6S, 17.1E, 1725 m</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>50.6 µg m(^{-3})</td>
<td>1408</td>
<td>23.6</td>
<td>7.9</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weighted mean</td>
<td></td>
<td>17.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58 Hermanus</td>
<td>34.4S, 19.3E, 13 m</td>
<td>MPI-Pruch</td>
<td>Fabian &amp; Pruchniewicz (1977)</td>
<td>44.3 µg m(^{-3})</td>
<td>1464</td>
<td>20.7</td>
<td>5.8</td>
<td>14</td>
</tr>
<tr>
<td>59 SE Australia</td>
<td>34.45S, 19.3E, 13 m</td>
<td>Ehmer plus Mast Brewer</td>
<td>Galbally (1968, 1971, 1972)</td>
<td>51.2 ± 14.4 µg m(^{-3})</td>
<td>730</td>
<td>24.8</td>
<td>6.9</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weighted mean</td>
<td></td>
<td>22.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Surface measurements in the tropics and in southern temperate latitudes. Mean values and standard deviations in nmol mol\(^{-1}\). See Text S-7 for Notes. DOI: https://doi.org/10.1525/elementa.376.t6
operated during this period (for example Oltmans and Komhyr, 1976), but data are currently not available. With the qualifications for Halley Bay, above, the methods used are valid, the sites are presumed free of interfering pollutants and representative of the well mixed boundary layer. The weighted mean of these observations, with the Halley Bay data, is 24 nmol mol\(^{-1}\), with a range from 16–30 nmol mol\(^{-1}\) for the period 1957–1966. Without the Halley Bay data, the range is 19–30 nmol mol\(^{-1}\), but the weighted mean remains at 24 nmol mol\(^{-1}\).

3.6. Changes in surface ozone

The sets of observations that pass the 4 data selection criteria imposed here for historical surface ozone observations are presented in Tables 4–6 and Figures 2–6, grouped by region. To quantify the changes of ozone levels from the historical (pre 1975) to the modern period (1990–2014), the historical data are compared to all available modern ozone observations in the same regions, according to the criteria described below. The modern data are extracted from the TOAR Surface Ozone Database and averaged across 5 × 5 degree grid cells at 5-year intervals (Schultz et al., 2017).

The historical observations have been selected to be representative of the well-mixed boundary layer, and the comparable metrics from the TOAR Database for rural stations are daytime average values and daily 8 hour maxima (DMA8). As the “rural” designation in the gridded average product from the TOAR database (Schultz et al., 2017) excludes sites at elevations above 2000 m (because of the typical increase of ozone with altitude) a proper comparison should exclude the measurements at higher mountain sites. In Europe (Figure 2) these are Grands-Mulets, Mont Blanc, Jungfraujoch, Zugspitze, Mt. Norikura in Japan and in North America, Capillo Peak. Except for the value of 63 nmol mol\(^{-1}\) in 1896 at Grands-Mulets, disregarding these eight points does not substantially change the picture in Figure 2 (see Figure S-2).

In Figure 2 there is no evidence of a change in rural background ozone during the historical record (1896–1975), as noted in Section 3.4. The averages for the four historical time periods in Table 4 differ by less than 3 nmol mol\(^{-1}\). This is in sharp contrast with previous analyses for the historical period and arises mainly due to the application of the 4 criteria to select valid historical data. The difference is due to: (a) the rejection of the Montsouris data, (b) the corrections detailed in Table 4, which have raised values for the early UV measurements by as much as 11%; (c) the inclusion of some early data that are not frequently cited and (d) other corrections as noted for Mast Ozone Meter, transmogrifier, and Cauer method data. Interestingly, the overall historical average found here is similar to older well-informed estimates: for example, Fabry (1950) remarks that surface ozone is typically about 20–25 nmol mol\(^{-1}\), with only modest daily variation.

Figure 2 indicates an increase of about 12 or 16 nmol mol\(^{-1}\) between the historical record and the modern period, depending on which metric (12-hr daytime average or DMA8) is applied to the modern data. The diurnal variation of surface ozone depends on both the underlying surface and the topography. At remote sites, as are considered here, measurements over flat continental surfaces show a distinct daytime maxima, while measurements over water, snow and ice show little diurnal variation and

![Figure 2](https://doi.org/10.1525/elementa.376.f2)

**Figure 2:** Historical measurements of surface ozone in the Northern Temperate region, 30°N–60°N (30 European, three Asian, one North American and one North African data sets; see Table 4 for details). Error bars represent standard deviations of the measurement averages (atmospheric variability), not uncertainty of the measurement. 5-year averages of modern UV measurements at sites below 2000 m, classified as ‘rural’, in the 5 × 5 degree gridded product from the TOAR database are also shown, both daytime averages (day) and daily 8-hour maxima (DMA8) (Schultz et al., 2017). DOI: https://doi.org/10.1525/elementa.376.f2
measurements from hilltops or from elevated sites in valleys often show night-time maxima. Afternoon averages were chosen for the historical data where available (see Tables 4–6 and notes above), but in a number of cases only daytime or 24-hour averages were available, and the UV measurements were all made at night, so it is difficult to ascertain whether DMA8 or the 12-hour daytime average is the best metric for comparison. However both show the same behaviour.

In the Northern Polar region (Figure 3), and in the Tropics (Figure 5), there is also some evidence of an increase, although there are fewer historical datasets with

![Northern High Latitudes: Historical surface ozone measurements](image1)

**Figure 3:** Historical measurements of surface ozone in the Northern Polar region (60°N–90°N; see Table 5 for details). Error bars represent standard deviations of the measurement averages (atmospheric variability), not uncertainty of the measurement. Five-year averages (daytime mean) of modern UV measurements at sites below 2000 m, classified as “rural”, in the TOAR 5 x 5 degree gridded average product are also shown, for both daytime averages (day) and daily 8-hour maxima (DMA8) (Schultz et al., 2017). DOI: [https://doi.org/10.1525/elementa.376.f3](https://doi.org/10.1525/elementa.376.f3)

![Southern High Latitudes: Historical surface ozone measurements](image2)

**Figure 4:** Historical measurements of surface ozone in the Southern Polar region (60°S–90°S; see Table 5 for details). Error bars represent standard deviations of the measurement averages (atmospheric variability), not uncertainty of the measurement. Five-year averages (daytime mean) of modern UV measurements at sites below 2000 m, classified as “rural”, from the TOAR 5 x 5 degree gridded average product are also shown, for both daytime averages (day) and daily 8-hour maxima (DMA8) (Schultz et al., 2017). DOI: [https://doi.org/10.1525/elementa.376.f4](https://doi.org/10.1525/elementa.376.f4)
which to make the comparison. In the southern hemisphere (Figures 4 and 6), there is no clear evidence of a change.

A quantification of the change in ozone mole fraction from the historic to the modern measurements has been performed for each latitude band in Tables 4–6 and is presented in Table 7. Table 7a presents changes of surface ozone from the historical period to the modern period using all available historical observations below 2000 m elevation. The modern period is based on rural ozone observations below 2000 m for the years 1990–2014, and 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.

Figure 5: Historical measurements of surface ozone in the tropics (30°S–30°N; see Table 6 for details). Error bars represent standard deviations of the measurement averages (atmospheric variability), not measurement uncertainty. Five-year averages (daytime mean) of modern UV measurements at sites below 2000 m, classified as “rural”, in the TOAR 5 × 5 degree gridded average product are also shown, both daytime averages (day) and daily 8-hour maxima (DMA8) (Schultz et al., 2017). DOI: https://doi.org/10.1525/elementa.376.f5

Figure 6: Historical measurements of surface ozone in the Southern Temperate region (30°S–60°S; see Table 6 for details). Error bars represent standard deviations of the measurement averages (atmospheric variability), not uncertainty of the measurement. Five-year averages (daytime mean) of modern UV measurements at sites below 2000 m, classified as “rural”, in the TOAR 5 × 5 degree gridded average product are also shown, for both daytime averages (day) and daily 8-hour maxima (DMA8) (Schultz et al., 2017). DOI: https://doi.org/10.1525/elementa.376.f6
Table 7: (a) Change of ozone from the historical period to the modern period based on all available historical observations below 2000 m elevation. Results in bold font are different from zero at the 95% confidence level. For the Northern Temperate zone, historical data were compared with modern data for a large region covering western Europe (16°×15° grid cell averages), and also for a single 5°×5° grid cell which encompasses most of the historical observations from western Europe. For the latter comparison monthly means from the individual rural sites were used, rather than a single average across the grid cell. (b) As for Table 7a, but with all questionable historical data sets omitted; in addition, results are shown for the Northern Temperate and Northern High Latitude zones with historical data sets limited to those that measured ozone via long-path UV methods. For details of the analysis and data sets used, see Tables S2 and S3. DOI: https://doi.org/10.1525/elementa.376.t7

<table>
<thead>
<tr>
<th>(a) Region</th>
<th>Historical ozone¹</th>
<th>Modern¹</th>
<th>Ozone change²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern High Latitude</td>
<td>23.1 (8)</td>
<td>30.5 (21)</td>
<td>31.8% (2.9%, 60.7%)</td>
<td>0.05</td>
</tr>
<tr>
<td>Northern Temperate (35–55°N and 5°W–15°E)</td>
<td>23.6 (27)</td>
<td>36.1 (16)</td>
<td>53.1% (29.8%, 76.4%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Northern Temperate (45–50°N, 5–10°E, individual sites)</td>
<td>23.6 (27)</td>
<td>32.5 (23)</td>
<td>37.8% (22.8%, 52.8%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Northern Low Latitude</td>
<td>22.2 (3)</td>
<td>31.8 (12)</td>
<td>43.4% (9.9%, 76.9%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Southern Low Latitude</td>
<td>17.5 (4)</td>
<td>16.3 (5)</td>
<td>−7.0% (−54.8%, 40.8%)</td>
<td>0.78</td>
</tr>
<tr>
<td>Southern Temperate</td>
<td>23.3 (3)</td>
<td>23.5 (10)</td>
<td>0.8% (−34.3%, 35.8%)</td>
<td>0.97</td>
</tr>
<tr>
<td>Southern High Latitude</td>
<td>20.5 (5)</td>
<td>24.1 (6)</td>
<td>17.9% (−17.2%, 52.9%)</td>
<td>0.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Region</th>
<th>Historical ozone¹</th>
<th>Modern¹</th>
<th>Ozone change²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern High Latitude</td>
<td>19.4 (1) UV data only</td>
<td>30.5 (21)</td>
<td>57.1% (15.2%, 99.0%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Northern Temperate (35–55°N and 5°W–15°E)</td>
<td>23.6 (22)</td>
<td>36.1 (16)</td>
<td>52.9% (29.4%, 76.4%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Northern Temperate (45–50°N, 5–10°E, individual sites)</td>
<td>24.4 (7) UV data only</td>
<td>36.1 (16)</td>
<td>47.7% (16.9%, 78.5%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Northern Low Latitude</td>
<td>N/A</td>
<td>31.8 (12)</td>
<td>24.8% (−10.2%, 60.0%)</td>
<td>0.30</td>
</tr>
<tr>
<td>Southern Low Latitude</td>
<td>24.6 (2)</td>
<td>23.5 (10)</td>
<td>−4.6% (−46.5%, 37.4%)</td>
<td>0.86</td>
</tr>
<tr>
<td>Southern High Latitude</td>
<td>21.6 (4)</td>
<td>24.1 (6)</td>
<td>11.6% (−24.9%, 48.0%)</td>
<td>0.56</td>
</tr>
</tbody>
</table>

¹Values indicate mean ozone in nmol mol⁻¹; sample size (for historical data = number of sites) in parentheses.
²Values indicate the percent change in mean ozone from the historical to the modern period with the confidence interval in parentheses, for the 12-hr daytime mean metric, using Welch’s t-test.

Across 5°×5° grid cells. Within all zones the changes in mean ozone mole fraction are either positive, or there is no clear indication of a change. Together they, along with the comparable data for DMA8 (see Text S-8 and Tables S-2 and S-3), indicate that surface ozone has increased by 30–70% from historical levels to the present in rural air in the temperate and polar regions of the Northern Hemisphere, and negligibly changed in the remote locations in the Southern Hemisphere. Statistical tests were performed using Welch’s generalization of Student’s t-test for unequal variances (Welch, 1947).

As described above and in the Supplemental Material (Text S-3), some of the historical data sets have questionable reliability due to a variety of documented issues. To understand the sensitivity of the results in Table 7a to these less reliable data sets, analyses were also performed with all questionable historical data sets omitted (see Text S-8 and Tables S-2 and S-3 for details), and also with all KI-based data omitted (this limits the historical measurements to seven long-path UV data sets in the Northern Temperate region and one in the Northern High Latitude zone). These results (Table 7b) are very similar, indicating that the results in Table 7a are robust with respect to the choice of historical data sets included in the analysis. Other tests (omitting only the observations from the Fabian and Pruchniewicz (1977) TROZ network, and using the non-parametric Wilcoxon rank sum test (Mann and Whitney, 1947) also yielded similar results (Table S-2), with smaller p-values for the non-parametric test. Comparisons were also made to the modern data using the daily 8-hour maximum (DMA8) metric; as expected these increases are about 20% larger (Table S-3).

The modern set of measurements utilized for this analysis is comprised of all rural sites with surface ozone measurements within the geographical boundaries of the region. There is considerable variability in the behaviour...
and levels of ozone in both rural surface air and the lower atmosphere within Europe (Chevallier et al. 2007, Lyapina et al., 2016). To test the sensitivity of these calculated changes to the choice of modern data sets, two domains were selected to represent the modern ozone observations for the Northern Temperate region. The first domain covers most of Western Europe and utilizes monthly mean observations from 16 grid cells, each 5° × 5°. The second domain focuses on a single 5° × 5° grid cell that encompasses most of the historical ozone observations from Western Europe, especially those that were based on UV methods. Within this single grid cell monthly means from the individual rural sites were used.

The differences in the estimated changes for the Northern Temperate region, based on these two modern domains are notable, both in Table 7a and 7b. However, for both modern domains the differences between Table 7a and 7b are quite small. This suggests that the uncertainty in the estimated increases in Table 7 depends more on the modern region chosen for comparison than on the historical data. Data representativeness thus seems to be the more important source of uncertainty.

It is therefore not surprising that the increases determined here are different from some past analyses (e.g. Parrish et al., 2012, 2014). Past analyses have used data from a few selected stations with long-term records, while this analysis has used all available historical and modern measurements. A more detailed matching of co-located historical and modern sites will provide additional insight and is being undertaken.

In the southern hemisphere, there is little evidence for an increase of ozone from the historical to the modern period.

It is worth emphasizing that these results do not depend on particular individual records, as has been noted with regard to several data sets judged to be of questionable credibility; indeed, if only the long-path UV measurements are retained, the increase in Europe is in the range 33–48%, with p-values between 0.01 and 0.03.

In summary, our best estimate for the increase of ozone at northern temperate and high latitudes is a range of 32–53%, based on all historical measurements, using the 12-hour daytime average as the modern metric, and 43–71% using the daily 8-hour maximum metric. These increases are different from zero at the 95% confidence level. Similar results are found using non-parametric statistical tests, and for calculations using (the most reliable) subsets of the historical data.

4. Free tropospheric measurements

Many of the measurement techniques described in Section 2 have been applied to measurements in the free troposphere through balloon and aircraft profiling, and more recently through ground and satellite-based remote sensing. Recent studies have examined free tropospheric ozone data quality issues by comparing time series of surface observations with commercial aircraft and ozonesonde profiles from nearby locations (Logan et al., 2012; Tanimoto et al., 2015). In some cases laboratory and field intercomparisons can provide information about instrument response changes with time (e.g. Attmannspacher and Dütsch, 1970, 1978; Hilsenrath, 1986; Kerr et al., 1994; Smit et al., 2007). The different methods and their biases and uncertainties as established through intercomparisons are each reviewed, and related to the UV standard.

4.1. Early measurements

Historical observations of free tropospheric ozone that pass the relevant data selection criteria are presented in Table 8 and Figure 7. The first direct measurement of ozone from a balloon ascent was made with three flights near Stuttgart, Germany in 1934 (Regener and Regener, 1934). A quartz UV spectrophotograph was used to observe the change with altitude of the total amount of ozone above the balloon. The derived profile increases very smoothly from a value equivalent to 40 nmol mol⁻¹ near the ground to ~180 nmol mol⁻¹ at 10 km; this integrates to about 40 Dobson units (DU) below 10 km.

A similar differential method was used on the manned flight of the high-altitude balloon Explorer II on November 11, 1935 in the western USA. This showed very little ozone below the tropopause; less than 10 DU (O’Brien et al., 1936; reported in Craig, 1950 and Fabry, 1950). This first flight also showed an unusually sharp tropopause.

Additionally, ozone measurements using wavelengths 310–330 nm were made on two unmanned balloon flights in Germany on 30 October 1937 and 11 December 1937 by Regener (1938a). These appear to have been more sensitive (reduced stray light) and found a good deal of variation in the troposphere, with maximum values that correspond to ~110 nmol mol⁻¹ at 3 km. These high values are questioned by Fabry (1950), who also points out that the differencing method is subject to large errors below the tropopause, where changes are small. However, 0–10 km column amounts of 23 and 37 DU on the ascent and descent of 30 October 1937 and 15 DU on both the ascent and descent of 11 December 1937 (Regener 1938b) compare well, on average, to November averages of 1–10 km columns measured over Europe by ozonesondes since 1990 (27–35 DU). These estimates of historical ozone levels include a 10% upward correction for the ultraviolet absorption coefficients used (Lauchli, 1928, 1929). The large variation between the ascent and descent of 30 October 1937 is probably an indication of the uncertainty of the differencing method, since it is unlikely that ozone changed by that much in less than a day.

Coblentz and Stair made a series of 19 balloon launches in 1938–40 from the eastern USA with a filter-based optical ozonesonde using wavelengths 290–330 nm (Coblentz and Stair, 1939, 1941); average values correspond to ~24 nmol mol⁻¹ in the lower troposphere and ~96 nmol mol⁻¹ near 10 km. These estimates include a 13% upward correction for the ultraviolet absorption coefficients used (Ny and Choong, 1933 and Fabry and Buisson, 1931).

In the early 1950s Paetzold (1955a, b) conducted 32 balloon flights with a UV spectrophotograph using wavelengths 295–318 nm, over Weissenau, Germany, of which 25 gave results in the 0–10 km region. From the cross-sections he
Table 8: Historical measurements of free tropospheric ozone (0–10 km column) in Europe and North America. Mean values and standard deviations in DU. See Text S-7 for Notes. DOI: https://doi.org/10.1525/elementa.376.t8

<table>
<thead>
<tr>
<th>Location</th>
<th>Period</th>
<th>Method</th>
<th>Reference</th>
<th>Range</th>
<th>Profiles</th>
<th>Mean (DU)</th>
<th>SD (DU)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stuttgart</td>
<td>26 June, 7, 31 July 1934</td>
<td>Differential UV</td>
<td>Regener and Regener (1934)</td>
<td>40 DU</td>
<td>3</td>
<td>40</td>
<td>n/a</td>
<td>24</td>
</tr>
<tr>
<td>48.8N, 9.2E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapid City</td>
<td>28 July 1934; 11 November 1935</td>
<td>Differential UV</td>
<td>O’Brien et al. (1936)</td>
<td>10 DU</td>
<td>2</td>
<td>10</td>
<td>n/a</td>
<td>25</td>
</tr>
<tr>
<td>44.1N, 103.2W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedrichshafen</td>
<td>30 October 1937; 11 December 1937</td>
<td>Differential UV</td>
<td>Regener (1938a)</td>
<td>15–37 DU</td>
<td>4</td>
<td>22.5</td>
<td>10.4</td>
<td>26</td>
</tr>
<tr>
<td>47.7N, 9.5E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beltsville</td>
<td>June 1938, 1939, 1940</td>
<td>Differential UV</td>
<td>Coblentz and Stair (1939, 1941)</td>
<td>22–40 DU</td>
<td>11</td>
<td>28</td>
<td>10.4</td>
<td>27</td>
</tr>
<tr>
<td>39.0N, 76.9W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darmstadt</td>
<td>August 15–25, 1942</td>
<td>KI</td>
<td>Ehmert (1949)</td>
<td>11–19 DU</td>
<td>5</td>
<td>14.4</td>
<td>3.4</td>
<td>28</td>
</tr>
<tr>
<td>49.9N, 8.7E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weissenau</td>
<td>February 1953–May 1954</td>
<td>Differential UV</td>
<td>Paetzold (1955a, b)</td>
<td>0–50 DU</td>
<td>31</td>
<td>22.3</td>
<td>12.8</td>
<td>29</td>
</tr>
<tr>
<td>47.8N, 9.6E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tromso</td>
<td>May–July 1955</td>
<td>Ehmert</td>
<td>Brewer (1955)</td>
<td>24–44 DU</td>
<td>4</td>
<td>33.2</td>
<td>8.3</td>
<td>30</td>
</tr>
<tr>
<td>69.6N, 19.0E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
quotes for 306 and 318 nm an estimated 24% upward correction is required, so that these flights show a seasonal maximum in April of 27 DU and a minimum in October of 15 DU, and an extreme range of 0–50 DU for the 0–10 km ozone amounts.

In 1942, ozone sampling with a wet chemical technique was conducted on 6 aircraft flights over Germany giving ozone profiles up to 9 km (Ehmert 1949). Subsequently Kay (1953) and Brewer (1955) presented data using the Ehmert method from aircraft flights over the UK and Norway. The former are quite low, and failed to detect the tropopause, possibly because of losses in the intake (Brewer, 1955). These data are excluded based on criterion 4. The latter show an average profile increasing from 30 nmol mol$^{-1}$ at launch to ~95 nmol mol$^{-1}$ just below the tropopause, very similar to a typical contemporary profile.

While some data sets show a large variability (e.g. Paetzold 1955a, b), the approach here is to include all of the early measurements that meet the relevant selection criteria (1, 2 and 4). The weighted mean of the observations in Table 8 is 23 DU, with a range of 0–50 DU for the period 1934–1955.

The corresponding modern average 0–10 km tropospheric column amount of ozone is 36 DU, for both the northern Europe and the eastern US regions in Figure 7. The modern values shown are annual averages for the 15° × 15° regions around the locations of most of the historical ascents (northern Europe and the eastern US), calculated from ozonesondes using the TOST product (Section 4.4). The error bars indicate standard deviations of the monthly averages (atmospheric variability), not uncertainty of the measurement. DOI: https://doi.org/10.1525/elementa.376.f7

### Figure 7: Historical measurements of free tropospheric ozone. Values shown are integrated 0–10 km ozone amounts.

The difference between the ascent and descent of 30 October 1937 (dashed line) is probably an indication of the uncertainty of the differencing method, since it is unlikely that ozone changed by that much in less than a day. The modern values shown are annual averages for the 15° × 15° regions around the locations of most of the historical ascents (northern Europe and the eastern US), calculated from ozonesondes using the TOST product (Section 4.4). The error bars indicate standard deviations of the monthly averages (atmospheric variability), not uncertainty of the measurement. DOI: https://doi.org/10.1525/elementa.376.f7

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean 0-10 km ozone column (DU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>0</td>
</tr>
<tr>
<td>1910</td>
<td>10</td>
</tr>
<tr>
<td>1920</td>
<td>20</td>
</tr>
<tr>
<td>1930</td>
<td>30</td>
</tr>
<tr>
<td>1940</td>
<td>40</td>
</tr>
<tr>
<td>1950</td>
<td>50</td>
</tr>
<tr>
<td>1960</td>
<td>60</td>
</tr>
<tr>
<td>1970</td>
<td>70</td>
</tr>
<tr>
<td>1980</td>
<td>80</td>
</tr>
<tr>
<td>1990</td>
<td>90</td>
</tr>
<tr>
<td>2000</td>
<td>100</td>
</tr>
<tr>
<td>2010</td>
<td>110</td>
</tr>
</tbody>
</table>

Regener and Regener (1934)  
O'Brien et al. (1936)  
Coblentz and Stair (1939; 1941)  
Ehmert (1949)  
Paetzold (1955a,b)  
Brewer (1955)  
Northern Europe 40–55°N, 0–15°E  
Eastern US 30–45°N, 70–85°W

### 4.2. Umkehr

Umkehr measurements rely on the fact that the scattering intensity of solar UV (effective scattering height) changes with solar zenith angle (SZA). Since most of the ozone in the atmospheric column is in the stratosphere, the information content of the tropospheric part of the retrieval is limited, and only ~50% or less of the information in layer 1 (1000–250 hPa in the standard retrieval) comes from the troposphere (Stone et al., 2015; Text S-9), while the rest of the information comes from the adjacent stratospheric layers.

In 1932–33 a series of 46 Umkehr profiles were made at Arosa, Switzerland (Götz et al., 1934) for determining the atmospheric profile of ozone. These yielded estimates of ~30–70 nmol mol$^{-1}$ at 2 km, ~80–100 nmol mol$^{-1}$ at 5 km and ~180–220 nmol mol$^{-1}$ at 10 km. These values are high compared to modern measurements, but carry a substantial uncertainty due to stratospheric influence, and are therefore excluded from the analysis of Section 4.1.

The accuracy of Umkehr profile retrievals has improved with time as a result of modifications to the retrieval algorithm and better a priori information from ozonesonde and satellite-derived climatologies. Early comparisons
(Kulkarni and Pittock, 1970) indicate a high bias of ~45% in the troposphere (likely in part due to the bias to lower response of the Brewer-Mast sondes; see Section 4.3). A 1989 field study (Komhyr et al., 1995) comprised of 6 morning and 6 afternoon co-incident Umkehr and ozonesonde measurements found tropospheric ozone in Umkehr profiles to be on average 16% low against ECC sondes launched in the morning, and 29% high in the afternoons. In a 2004–2005 comparison, 60 co-incident Umkehr and ozonesonde profiles taken at Belsk, Poland showed on average a slight (~2% ±25%) low bias of the Umkehr partial ozone column below 250 hPa, relative to the sondes column (Krzyścin and Rajewska-Wiecz, 2007). Umkehr profile measurements are currently made twice daily, at ~16 sites worldwide (Figure 8).

4.3. Ozonesondes

Ozone soundings were undertaken at a global network of 11 sites from 1962 to 1966 by the US Environmental Science Services Administration. This network operated in parallel with a North American network of 13 sites, coordinated by the US Air Force Cambridge Research Laboratories from 1963–1965. Together these networks released over 2000 Regener, Brewer-Mast and carbon-iodine sondes (Komhyr and Sticksel, 1967, 1968; Hering, 1964; Hering and Borden, 1964, 1965, 1967). Regener chemiluminescent sondes were used regularly for only a brief period in the 1960s, as they showed somewhat erratic response, with an average bias of about ~40% in the troposphere (Chatfield and Harrison, 1977; Wilcox, 1978; Hering and Dütsch, 1965). Regular chemical ozone soundings at a number of sites in Europe, North America, Australia and Antarctica began in the latter half of the 1960s. Balloon-borne ozonesondes therefore provide the longest time series of the vertical ozone distribution throughout the troposphere. However, ozone soundings are limited in spatial and temporal coverage. Routine ozonesonde launches have been made at less than 100 stations worldwide; these are unevenly distributed, although this is much improved since the introduction of the Southern Hemisphere ADDitional Ozonesondes (SHADOZ) network in the 1990s. Launch frequency is typically weekly, and at most 2–3 times per week. Vertical resolution is high: the ozone sensor response time ($e^{-1}$) of about 25–40 seconds (Smit and Kley, 1998) gives the sonde a vertical resolution of about 100–200 metres for a typical balloon ascent rate of 4–5 m s$^{-1}$ in the troposphere.

All “modern” sonde types – ECC, Brewer-Mast (BM), Brewer-GDR (GDR), Indian and the Japanese KC – use the reaction of ozone with aqueous potassium iodide (KI), which assumes that two electrons are produced for each molecule of ozone, as the method of ozone detection. As discussed in Section 2.1, there can be variations in the stoichiometry of the ozone-iodide reaction. Also, there may be losses of ozone in the pump and of ozone or iodine to the walls of the sensor chamber, as well as iodine evaporation and possibly adsorption to the platinum cathode (Tarasick et al., 2002). Both the GDR and Indian sondes are similar in design to the BM sonde (Brewer and Milford, 1960), while the KC sonde is similar to the early Komhyr carbon-iodine (CI) sonde (Komhyr, 1964; Komhyr et al., 1968). The ECC (Komhyr, 1969) is an electrochemical concentration cell with two chambers connected by an ion bridge. These different instrumental layouts cause differences in response (Smit, 2002) and ozone losses depend both on these differences and on sonde preparation. Losses can be as large as 40% in poorly prepared sondes, although such issues are much less common after about 1980. Slow side

![Figure 8](https://doi.org/10.1525/elementa.376.f8)
reactions in the sensing solution can cause excess iodine to be produced (Saltzman and Gilbert, 1959a; Flamm, 1977; Johnson et al., 2002). This effect is modest except when there are sharp ozone gradients, as it causes a slow (~20 min) second-order time response. As for surface KI monitors, interference from other gases can be a problem (generally restricted to the boundary layer) in polluted areas (Schenkel and Broder, 1982; Tarasick et al., 2000). Pump rate or temperature errors, as well as radiosonde pressure biases, will also produce ozone measurement errors (positive or negative), but these are generally small (<1%) in the troposphere. Background currents cause ozone offsets that are typically as large as 5% in the upper troposphere. The use of sensing solutions other than those recommended for each type of ECC sonde can introduce additional biases of 2–8% (Smit and ASOPOS panel, 2011), although these can be corrected. Ozonesondes can, therefore, under certain conditions significantly underestimate ozone concentrations, but it is difficult to explain positive errors larger than about 10–20%.

Ozonesondes record a surface measurement at release time, but this is subject to negative errors if the sonde is not allowed to run for a few minutes after removal of the ozone filter. Also, the sonde is released at about 1 metre above the ground, and strong gradients often exist in the first few metres above the ground (Galbally, 1968). Aliasing from diurnal cycles is also an issue in the planetary boundary layer if release times are variable (Tarasick et al., 2005; Thompson et al., 2014). Ozonesonde profiles are very useful for detecting the transition from the boundary layer to the free troposphere, however.

Numerous field and laboratory intercomparisons of ozonesondes have attempted to characterize ozone-sonde biases and uncertainties. These show considerable variability (Text S-10, Figure S-4, Tables S-4 to S-8 and Figures S-5 to S-8) in part due to differences in preparation, and also as in a number of studies a UV reference photometer was not available, and so the results are relative to an average profile, different for each study. A clearer picture emerges if only the comparisons with a UV reference (the modern standard) are retained. This is done for ECC sondes in Figures 9 and 10. Several adjustments for consistency are described in the figure caption. Despite the fact that there have been several models of

![Figure 9: Published intercomparison results of Electrochemical Concentration Cell (ECC) sondes with UV photometer measurements. The symbol size is inversely proportional to the standard error of the mean difference, while the error bars show standard deviations. The curves are running means (weighted by the standard error) of the mean difference (black) and the one standard deviation uncertainty (grey). Included are unpublished data (1980) from Barnes et al. (1985). The 1985 point has been adjusted to account for a positive bias due to the use of 1.5% KI solution (6.2 ± 1%, from Barnes et al. (1985) and Peterson (1978)). The 1980 point and that for Torres & Bandy (1978) were not adjusted for this bias, as it apparently was not observed before the early 1980s (as discussed by Barnes et al. (1985)). Instead, following JCGM 100:2008, an additional uncertainty of 6.2/√3 = 3.6% has been added to these points. The data from Hilsenrath (1986) are for the NOAA (1% KI) sondes. Data from Smit and Kley (1998) are averaged over all ECC results, and from Smit and Sträter (2004) are an average of results for the two types using recommended solutions (1% for SciPump and 0.5% for EnSci). Sondes were operated according to standard operating procedures for the agency participating. Data after about 1995, with some exceptions, have not been normalized to a total ozone measurement. DOI: https://doi.org/10.1525/elementa.376.f9](https://doi.org/10.1525/elementa.376.f9)
ECC sondes, the running averages show no significant trend in bias over the 50-year period, despite the model changes. This is independent of normalization to coincident total ozone measurements, as normalization factors at most long-term sites show no trend. A mean offset, weighted by the standard error of each point is therefore found (shown as a red dashed line). Using the ECC sondes as a transfer standard, this value is then used to transfer the UV reference to the bias results for other sondes, for intercomparisons without a UV reference photometer (Figures 11–14). The calculated uncertainty of this weighted mean is also added, and we note that this may be an underestimate, as sample sizes in intercomparisons are small, and experimental conditions may not reflect the full range of conditions found in long-term field operations.

The result (Figures 11 and 12) indicates that the BM sonde shows an increase in tropospheric response of about 20% between the 1970s and the 1990s. Improved preparation procedures for BM sondes (Attmannspacher and Dütsch, 1978; Claude et al., 1987; DeBacker, 1999; Favaro et al., 2002; Tarasick et al., 2002) may have contributed to this, and there may have been minor changes in sonde manufacture over the long period of record (World Climate Research Programme, 1998). This is consistent with the conclusion of Schnadt Poberaj et al. (2009), who compared European BM sondes with UV-measurements from the GASP and MOZAIC flights (Sections 4.7 and 4.8, below).

Over the same period, the KC sondes also show a modest increase in tropospheric response (Figures 13 and 14). In addition, although almost all current ozonesonde data are from ECC sondes, the transition to the ECC has been gradual (Figures 15 and 16). Since the other sonde types historically show negative biases in the troposphere relative to the ECC sonde (see also Tables S7 and S8), this transition may itself introduce an apparent trend in free tropospheric ozone derived from ozone sondes, if data are combined without adjustments. The BM sondes were used extensively in the 1970s and in Europe in the 1980s, and are currently in use at only one site (Hohenpeissenberg). Large amounts of data prior to the early 1990s exist from the Brewer-GDR, Indian, and the Japanese KC sondes.

Currently, station records are being re-evaluated for artifacts introduced by changes of sonde type, manufacturer, strength of sensing solution, or preparation procedure, under the Ozonesonde Data Quality Assessment activity (Smit et al., 2017). This has resulted in changes to the Canadian record of 2–5% in the 1980–2015 period, and as much as 20% to the pre-1980 BM data (Tarasick et al., 2016). Tropical stations in the SHADOZ network show changes of up to 8% (Witte et al., 2017). A large part of this is due to the fact that after about 1995, stations using the new EnSci sonde with 1% KI solution show a 4–8% positive bias in the lower troposphere and 2–6% in the upper troposphere (Table S-5). Sites in Europe and the US show changes of 1–2% (van Malderen et al., 2016; Witte et al., 2019).

Figure 10: As Figure 9, for the upper troposphere. Beekmann et al. (1995) shows a very high bias; the normalized data, corrected for background current (3.5 ± 3.5%) have been used, as suggested. DOI: https://doi.org/10.1525/elementa.376.f10
Figure 11: Intercomparison results of Brewer-Mast (BM) sondes. Those without a UV photometer reference have been adjusted to the UV standard using the ECC average bias. Adjusted uncertainties include the quoted ECC uncertainty from the intercomparison, and the bias uncertainty from Figure 9. Lehmann (2005) uses SAGE and Dobson data as a UV reference. Sondes were operated according to standard operating procedures for the agency participating. Profile data were normalized to a total ozone measurement. DOI: https://doi.org/10.1525/elementa.376.f11

Figure 12: As Figure 11, for the upper troposphere. DOI: https://doi.org/10.1525/elementa.376.f12
**Figure 13:** As Figure 11, for Japanese KC sondes. Sondes were operated according to standard operating procedures for the agency participating. Profile data were normalized to a total ozone measurement. DOI: https://doi.org/10.1525/elementa.376.f13

**Figure 14:** As Figure 13, for the upper troposphere. DOI: https://doi.org/10.1525/elementa.376.f14
**Figure 15:** Ozonesonde profiles 1960–1989. The color bar indicates the number of profiles of each type available from each station. Data from the World Ozone and UV Data Centre (WOUDC). Most stations launch weekly; at three sites in Europe launches are three times weekly. DOI: https://doi.org/10.1525/elementa.376.f15

**Figure 16:** Ozonesonde profiles 1990–2015. The color bar indicates the number of profiles of each type available from each station. Data from the WOUDC and the Network for Detection of Atmospheric Composition Change (NDACC). DOI: https://doi.org/10.1525/elementa.376.f16
4.4. Ozonesonde derived data products

Ozonesonde data are global and long-term, but sparse in space and time. Several products attempt to remedy this deficiency by combining ozonesonde data with satellite ozone data or meteorological information. The ML climatology (McPeters and Labow, 2012; McPeters et al., 2007) uses Microwave Limb Sounder (MLS) data to produce a zonally-averaged climatology in 10° latitude bands from 0 to 65 km. BSVertOzone (Hassler et al., 2018; Bodeker et al., 2013) uses several satellite data sources and a sophisticated regression-interpolation scheme to produce a zonally-averaged climatology in 5° latitude bands from 0 to 70 km.

The Trajectory-mapped Ozonesonde dataset for the Stratosphere and Troposphere (TOST), used elsewhere in this paper and in TOAR-Climate (Gaudel et al., 2018), uses the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess, 1998) and meteorological fields from National Centers for Environmental Prediction (NCEP) reanalyses to fill the gaps between ozonesonde stations, by extending each ozone record along its trajectory path forward and backward for 4 days. Over this 4-day period ozone production and loss is assumed to be negligible. Ozone values along these trajectory paths are binned into a 3-dimensional grid of 5° × 5° × 1 km (latitude, longitude, and altitude), from sea level or ground level up to 26 km (Figure 17). Tropospheric column ozone (TCO) is calculated from ozone mole fractions below the tropopause, found using the WMO 2 K/km lapse-rate definition applied to the NCEP reanalysis data.

TOST has been evaluated using individual ozonesondes, excluded from the mapping, by backward and forward trajectory comparisons, and by comparisons with aircraft profiles and surface monitoring data (Tarasick et al., 2010; Liu, G. et al., 2013; Liu, J. et al., 2013). Differences are typically about 10% or less, but there are larger biases in the UTLS, the boundary layer, and in areas where ozonesonde measurements are sparse. The accuracy of the TOST product depends largely on the accuracy of HYSPLIT and the meteorological data on which it is based. Data products are available at http://woudc.org/data/products/#related-ozone-products.

4.5. Tropospheric ozone lidar

UV DIAL (Differential Absorption Lidar) technique is a well-established technique for tropospheric ozone monitoring, from as low as 100 m to the tropopause, from ground-based sites (Browell et al. 1983, Ancellet and Ravetta, 2003) or aircraft (Browell et al. 1983, Ancellet and Ravetta, 2003). Differences between existing lidar instruments are in the wavelength choice and number of wavelength pairs used for the ozone retrieval, the size of the telescope or the laser power. Vertical resolution is variable, depending on wavelength choice, signal strength and integration time, but can be as fine as 50–100 m in the lower troposphere. Temporal resolution can be very high (1 min). Temporal coverage is generally limited by the need for human operators. The recent deployment of an autonomous system (Strawbridge et al., 2018), however, removes this handicap.

---

**Figure 17:** Number of data values resulting from the trajectory mapping averaged in each 5° x 5° x 1 km bin at 5–6 km in the TOST decadal average ozone field for October 2000–2009. The standard errors are generally of the order of a few nmol mol⁻¹, although where data density is low they can be higher. DOI: https://doi.org/10.1525/elementa.376.f17
DIAL errors may arise from electronic interference in the detected signals, unaccounted effects of particulate backscatter and extinction, errors in ozone absorption across sections, insufficient knowledge of the near-range geometrical overlap function, and errors in the cross-sections or a priori concentrations of interfering atmospheric molecules (Eisele and Trickl, 2005; Leblanc et al., 2016b). Beam misalignment can be avoided by comparing the ozone profiles obtained for different wavelength pairs. Relative uncertainties can be as low as 2–5% in the lower and middle troposphere, for averaging times of ~1 min, if the “on” (i.e. more strongly-absorbed) wavelength selected is below 280 nm (Trickl, 2019). For longer “on” wavelengths, used for higher altitudes, the uncertainties grow, and significantly longer signal averaging or a reduction of the vertical resolution must be applied.

The accuracy of tropospheric ozone measurements using lidar systems has been analysed using three different approaches: 1) a system uncertainty analysis based on estimated uncertainties from component sources, 2) simultaneous differences with other techniques (ozonesonde, UV instruments on aircraft and mountain stations) during intercomparison campaigns, 3) differences between seasonal averages from two instruments over a long time period.

The system uncertainty analysis considers random uncertainty (altitude, resolution and ozone-dependent signal to noise ratio), as well as the sources of systematic uncertainty noted above. Details may be found in Leblanc et al. (2016a, b).

Table 9 summarizes results of the small number of published intercomparison campaigns. The weighted mean of the quoted lidar bias in Table 9 is approximately 0% ±2% in the lower troposphere, and −3% ±3% in the upper troposphere. The bias is also small below 3 km, but uncertain as lidar systematic errors increase at low altitude (high aerosol load, lidar misalignment).

Several sites (Boulder, Observatoire de Haute Provence, La Reunion (France), Garmisch-Partenkirchen, Huntsville) currently have DIAL running in parallel with other techniques (Figures 8 and 16), and so further comparisons are possible.

One example of lidar-ECC comparison results is shown in Figure 18 for a set of 13 co-located and simultaneous measurements with the Table Mountain Facility (TMF) lidar during the SCOOP campaign (Leblanc et al. 2018). Average agreement is excellent; throughout the profile it is within the theoretical total uncertainty. For the lidar, this includes detection noise, and systematic errors as described above, while for the ozonesonde profiles it is assumed to be 5%.

Gaudel et al. (2015) examined differences between 5-year ozone seasonal averages, using regular, not necessarily simultaneous, lidar and ECC profiles. Over a 20 year period, the ECC sondes averaged about 1 nmol mol⁻¹ higher in the free troposphere above 4 km. Seasonal differences fluctuated generally between ±5 nmol mol⁻¹, with a maximum of 11 nmol mol⁻¹, at 6–8 km. This was shown to be due to significant transport differences, correlated with the requirement for clear sky conditions for lidar measurements, inducing a meteorological bias in lidar sampling.

The three different analyses yield results consistent with a precision better than 10% and a slight negative bias of 0–3% with ECC sondes. When adjusted for the mean ECC biases found in Section 4.3, this corresponds to a mean bias of +1% ±8% in both lower and upper troposphere. These results depend on good aerosol and cloud screening, based on the backscatter lidar signal analysis; under these conditions the lidar accuracy (as from these published intercomparison campaigns) is better than that calculated from the system uncertainty analysis.

4.6. Ground-based Fourier Transform Infra-Red (FTIR)

Global FTIR observations are coordinated by the Infrared Working Group of NDACC (Network for the Detection of Atmospheric Composition Change, https://www2.acom.ucar.edu/irwg www.ndacc.org). Calibration and retrievals for ozone are standardized across the network (Vigouroux et al., 2015). The retrieval follows Optimal Estimation (OE) theory (Rodgers, 2000) and requires a priori data for the atmospheric state and other forward model parameters. A priori atmospheric composition profiles are constant for all retrievals in the time series for a given site. They are derived from climatological runs of the WACCM V4 (Whole Atmosphere Community Climate Model) model. Daily a priori temperature and pressure profiles are from NCEP. The instruments are solar viewing and so observations are biased to clear sky daytime, with seasonal limitations at high latitudes. Typically, several observations are taken per day if there is a clear line of sight to the sun.

Typical degrees of freedom for signal (DOFS) are 4–5 (see Text S-11). Averaging kernels are shown in Figure S-11a (and see Figure 1 of Vigouroux et al. (2015)). They represent the contribution to the retrieval from the measurement. Table S-10 lists stations that obtain at least 0.8 DOFS for the ground – 8 km layer. The remainder of the information is from the a priori best estimate. Those stations with sufficiently long and dense time series are used in TOAR-Climate (Gaudel et al., 2018) where most have DOFS (to 8km) >0.9. Only a small portion of this first summed partial kernel is sensitive to the lower stratosphere (Figure S-11b). Further details about the information content of ozone retrievals can be found in Schneider et al. (2008), Vigouroux et al. (2015), and in Wespes et al. (2012) focusing on tropospheric ozone.

Figure 19 shows time series of the partial columns of ozone measured by the FTIR at the Izaña Atmospheric Observatory (IZO), together with columns derived from the coincident ozonesonde profiles. The sondes were launched weekly at Santa Cruz de Tenerife (35 km northeast of IZO) from 1999 to 2006 and at Guimar station (15 km east of IZO) since October 2006. The FTIR data are averaged in a temporal window of 6 hours around the ECC launch time (12 UTC).

Table 10 gives the ozone uncertainty budgets estimated by the OE approach and from ECC sonde intercomparisons at the Izaña Atmospheric Observatory. Uncertainties in the tropospheric columns (from station altitude = 2.37 km for IZO) are shown. Total theoretical random uncertainty
Table 9: Lidar bias and precision derived from intercomparison campaigns between lidar, ozonesonde, aircraft and mountain site simultaneous measurements. Bias and Precision represent means and standard deviations of differences from the ECC sonde profiles. Theoretical accuracy is the expected lidar accuracy from the system uncertainty analysis. (OHP: Observatoire de Haute Provence; GAP: Garmisch-Partenkirchen). DOI: https://doi.org/10.1525/elementa.376.t9

<table>
<thead>
<tr>
<th>Campaign</th>
<th>Reference</th>
<th>Altitude (km)</th>
<th>Bias</th>
<th>Precision</th>
<th>Theoretical accuracy</th>
<th>Altitude (km)</th>
<th>Bias</th>
<th>Precision</th>
<th>Theoretical accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHP, 1989: 8 ECC+BM sondes</td>
<td>Beekmann et al., 1994</td>
<td>3–8</td>
<td>4% ± 2%</td>
<td>5%</td>
<td>10%</td>
<td>5–9</td>
<td>6% ± 3.5%</td>
<td>9%</td>
<td>10%</td>
</tr>
<tr>
<td>OHP, 1991: 8 ECC+BM, 8 aircraft profiles</td>
<td>Beekmann et al., 1995</td>
<td>1–6</td>
<td>2.8% ± 0.8%</td>
<td>3%</td>
<td>&lt;10%</td>
<td>6–8</td>
<td>10% ± 3%</td>
<td>10%</td>
<td>&lt;20%</td>
</tr>
<tr>
<td>Huntsville, 2008: 12 ECC sondes</td>
<td>Kuang et al., 2011</td>
<td>2–7</td>
<td>3% ± 2.8%</td>
<td>4%</td>
<td>10%</td>
<td>3–10</td>
<td>1.5% ± 10%</td>
<td>4%</td>
<td>10%</td>
</tr>
<tr>
<td>Saga 2013, 2 ECC sondes</td>
<td>Uchino et al., 2014</td>
<td>4–6</td>
<td>1% ± 1.5%</td>
<td>9%</td>
<td>10%</td>
<td>5–12</td>
<td>0% ± 1.5%</td>
<td>5–20%</td>
<td>5–20%</td>
</tr>
<tr>
<td>OHP, 5-year seasonal mean, ~240 sondes</td>
<td>Gaudel et al., 2015</td>
<td>2–5</td>
<td>0% ± 1.5%</td>
<td>2–5%</td>
<td>2–5%</td>
<td>5–12</td>
<td>0% ± 1.5%</td>
<td>5–20%</td>
<td>5–20%</td>
</tr>
<tr>
<td>TMF, 2016: 15 ECC</td>
<td>Leblanc et al., 2018</td>
<td>2–12</td>
<td>1% ± 1.5%</td>
<td>2–5%</td>
<td>2–5%</td>
<td>5–12</td>
<td>0% ± 1.5%</td>
<td>5–20%</td>
<td>5–20%</td>
</tr>
<tr>
<td>GAP, &gt;20 BM sondes; mountain sites</td>
<td>Trickl, 2019</td>
<td>2–5</td>
<td>0% ± 1.5%</td>
<td>2–5%</td>
<td>2–5%</td>
<td>5–12</td>
<td>0% ± 1.5%</td>
<td>5–20%</td>
<td>5–20%</td>
</tr>
</tbody>
</table>
is estimated as the sum of the random parameter uncertainty and the smoothing uncertainty (associated with the limited vertical sensitivity of the FTIR technique). Significant contributors to the random parameter uncertainty are uncertainties in the temperature vertical profiles, the instrumental line shape, and the measurement noise, while the systematic uncertainty is dominated by spectroscopic uncertainties (Schneider and Hase, 2008; Schneider et al., 2008; García et al., 2012). The overall theoretical uncertainty of ~11% in tropospheric ozone partial columns is dominated by the smoothing uncertainty.

Smoothing uncertainty is also important in sonde-FTIR comparisons: the scatter (one standard deviation) of the relative differences is reduced from ~9% to ~7% when comparing ECC sonde profiles smoothed with the FTIR averaging kernels.

The mean bias between ozone partial columns from FTIR and ECC sondes of 4%, which can be up to ~6% with alternate retrieval strategies (García et al., 2012), is in excellent agreement with the positive bias of 1–5% (+5%) for ECC sondes found from UV-referenced sonde intercomparison studies (Section 4.3), and with the positive bias of 5–8% found from MOZAIC/IAGOS comparisons (Section 4.8).

Recent laboratory work recommends simultaneous measurements of ozone absorption coefficients in the IR and the UV (Orphal et al., 2016), which will lend further confidence in FTIR methods by tying them directly to the UV standard.

**Figure 18:** Blind comparison between the Table Mountain Facility (TMF) tropospheric ozone lidar profiles and ozonesonde profiles, from co-located and simultaneous measurements during the SCOOP campaign in August 2016. DOI: https://doi.org/10.1525/elementa.376.f18

**Figure 19:** Partial columns (2.4–13 km) measured at Izaña Atmospheric Observatory, with integrated ozone partial columns from ECC sondes at a nearby site. DOI: https://doi.org/10.1525/elementa.376.f19
Table 10: Estimated random and systematic uncertainties relative to the FTIR retrieved tropospheric ozone partial column (2.37–13 km) for the Izaña Bruker 120/5HR, obtained by comparing to 515 coincident ECC sondes (García et al., 2012). DOI: https://doi.org/10.1525/elementa.376.t10

<table>
<thead>
<tr>
<th>Uncertainty (%)</th>
<th>Theoretical random parameter uncertainty</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoothing uncertainty</td>
<td>0.5 nmol mol⁻¹</td>
<td>±3.5%</td>
</tr>
<tr>
<td>Total theoretical random uncertainty</td>
<td>~11</td>
<td></td>
</tr>
<tr>
<td>Theoretical systematic uncertainty</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Experimental random uncertainty – ECC sondes</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Experimental systematic uncertainty: FTIR–ECC sondes</td>
<td>~4</td>
<td></td>
</tr>
</tbody>
</table>

4.7. Ozone measurements from aircraft

In addition to the early measurements described in Section 4.1, observations were made using a KI method by Fabian and Pruchniewicz (1977) on 34 regular airline flights. The first major program of ozone measurements from regular passenger aircraft began after the observation of high ozone inside the cabin of planes flying over the US (V. Mohnen, personal communication). Ozone levels over 350 nmol mol⁻¹ were observed on some flights over the US in 1973, and as high as 600 nmol mol⁻¹ on polar flights (Bischof, 1973). The problem was exacerbated in 1975 when the long-range Boeing 747 SP was introduced, as this flew higher and further north, and so frequently well into the lower stratosphere. Ozone levels over 600 nmol mol⁻¹ were observed frequently and passengers and crew complained of severe headaches and nosebleeds. This became a crisis for airlines, which was initially dealt with by pilot advisories (FAA, 1977) and flight planning to avoid areas of expected high ozone (based on the measurements available in the 1970s). New FAA regulations, AC_120–38 (FAA, 1980) which restrict maximum cabin ozone levels to 250 nmol mol⁻¹ (peak) and 100 nmol mol⁻¹ (3-hour average) were developed and are still in effect. Most passenger jet aircraft now have ozone destruction filters on the cabin air intakes, but not all, as avoidance is still an option. This is not always successful, however, as even these high limits are sometimes exceeded (Bekö et al., 2015).

This urgent issue, combined with concern about adverse effects of aircraft exhaust emissions on the atmosphere, led to a collaboration between NASA and several US airlines to operate the Global Atmospheric Sampling Program (GASP). From March 1975 to June 1979, GASP provided the first representative ozone measurements from regular aircraft (Falconer and Holdeman, 1976; Nastrom, 1977). A commercially available 253.7 nm UV photometer was modified for automated operation. The air sample from the inlet for gas phase measurements was pressurized by means of a PTFE (Teflon)-coated diaphragm pump to well above cabin pressure. Losses in the inlet and pump were as large as 16% in 1975–76, reduced to <6% in 1977. Overall uncertainty is estimated at ±8.4% in 1975–76, and ±3.3% from 1977, with a known bias of +9% from the calibration via NBKI (Schnadt Poberaj et al., 2007). Four passenger B-747 aircraft made a total of 6149 flights, with measurements at altitudes between 6 and 13.7 km. The program mostly covered the North Atlantic and Pacific Oceans, as well as North America, but also to a lesser extent Europe (Schnadt Poberaj et al., 2009), with a few flights to India, Singapore, Australia, New Zealand, and Brazil.

NOXAR (Nitrogen Oxides and Ozone along Air Routes) provided measurements onboard a B-747 from Zurich (Switzerland) to Atlanta, Boston, New York and Chicago and also Beijing, Bombay and Hong Kong. The system was operated from May 5, 1995, until May 13, 1996 (540 flights), and from August 12 until November 23, 1997 (104 flights). The analyzer was a modified Environics S-300 253.7 nm UV absorption instrument. The air sampling contained an aerofoil-sectioned aluminum boom with a PTFE core, just forward of the aircraft’s rearmost starboard door. The boom extended 23 cm into the slip-stream from the aircraft skin. Ambient air at ~200 hPa was compressed to aircraft cabin pressure (~800 hPa) using a PTFE-coated diaphragm pump. Losses in the inlet and pump were as large as 7%, adding ±3.5% to the overall uncertainty. Precision is estimated at ±0.5 nmol mol⁻¹ and overall uncertainty at ±5 nmol mol⁻¹ ±6% (Dias-Lalcaca et al., 1998; Brunner et al., 2001).

More than three decades of dedicated research aircraft measurements, mostly with UV absorption instruments and also with some O₃-NO chemiluminescence measurements, are also archived (e.g. LARC, 2019; BADC, 2019; ESRL, 2019; NCAR, 2019). Typically these flights were directed at observing particular atmospheric phenomena (e.g. biomass burning plumes), and so sampling may be biased accordingly.

4.8. MOZAIC/IAGOS

In-service Aircraft for a Global Observing System (IAGOS) and its predecessor Measurement of Ozone and water vapor by Airbus in-service airCraft (MOZAIC) have been making automatic and regular measurements of O₃ vapor and standard meteorological parameters onboard long-range commercial Airbus A330/A340 aircraft since August 1994 (Marenco et al., 1998, Petzold et al., 2015).

Ozone measurements are made by dual-beam UV absorption monitors with a response time of 4 s, a detection limit of 2 nmol mol⁻¹, and an uncertainty of ±2 nmol mol⁻¹ ±2%, including a 1% uncertainty in the reference instrument (Thouret et al., 1998; Nédélec et al., 2015). As in GASP and NOXAR, the sampled air is compressed by a Teflon-coated diaphragm pump before entering the UV-photometer, but losses in the inlet and pump are estimated at less than 1%, based on laboratory and ground tests (Thouret et al., 1998). Quality assurance procedures have not changed since the beginning of the record in 1994. No calibration drift has been observed, nor inconsistency between MOZAIC and IAGOS instruments (Blot et al., in preparation). Ozone monitors are calibrated annually to a reference analyser at the Bureau Internationale
des Poids et Mesures (BIPM), and also compared every 2 hours to an in-flight ozone calibration source. MOZAIC can be considered a reference dataset (e.g. Thouret et al., 1998, 2006; Schnadt Poberaj et al., 2009; Logan et al., 2012), due to its known calibration history.

Previous comparisons of MOZAIC/IAGOS data with ozonesondes show negative biases of a few per cent (sonde values higher), with larger differences in the earlier part of the MOZAIC record (Thouret et al., 1998; Staufer et al., 2013, 2014). Recent results also show small (6% or less) negative biases against ECC sondes (Zbinden et al., 2013; Tanimoto et al., 2015). Despite the large number of profiles in either case, coincidences between aircraft and sonde launches are few. However, a comparison (Figure 20) of trajectory-mapped averages of ozonesonde and MOZAIC/IAGOS profile data (see description of TOST, above) indicates that over 1994–2012 sonde measurements are about 5 ± 1% higher in the lower troposphere, and 8 ± 1% higher in the upper troposphere, consistent with the 1 ± 5% and 5 ± 5% average biases found for ECC sondes in Section 4.3, from UV-referenced sonde intercomparison studies. In addition, some of the routine soundings during this period will be from EnSci sondes used with 1% KI solution, which may account for the additional bias (Section 4.3).

Unlike ozonesonde sites, airports are typically urban, but MOZAIC/IAGOS ozone data do not appear largely affected by local boundary layer chemistry (Petetin et al., 2016, 2018). This is also indicated by Figure 20. The IAGOS database (http://www.iagos.org) currently contains data from more than 100,000 vertical profiles of tropospheric ozone, measured during takeoff and landing from 148 airports around the world since August 1994. The data sampled from the ascents and descents at these airports are unevenly distributed both spatially and temporally because the frequency of visits to airports by aircraft that take part in MOZAIC/IAGOS varies, depending on commercial airlines’ operating constraints. In particular, data are sparser in the southern hemisphere (Figure 21).

A year-by-year comparison (Figure S-10) 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 larger differences in the first two years of the MOZAIC/IAGOS record have been noted previously (Logan et al., 2012; Staufer et al., 2014) and may also be due to sampling differences.

4.9. Tropospheric ozone satellite and residual measurements

The measurement of tropospheric ozone from space is a challenge, because of the large stratospheric ozone burden that satellite instruments must look through. Typical variations in the stratosphere at mid-latitudes (more than 10%) are larger than the entire amount of ozone in the troposphere. A number of techniques have been developed to derive information about tropospheric ozone from nadir-pointing spectrometer data. The tropospheric ozone residual (TOR) technique (Fishman et al. 1990, 1996, 2003; Ziemke et al., 1998; Chandra et al., 2003) uses height-resolved ozone information from the Solar Backscatter Ultraviolet (SBUV/2) or the Stratosphere Aerosol and Gas Experiment (SAGE), Halogen Occultation Experiment (HALOE) or Microwave Limb Sounder (MLS) instruments to subtract stratospheric ozone from the total column ozone measured by the Total Ozone Mapping Spectrometer (TOMS), or more recently, the Ozone Monitoring Instrument (OMI) (Ziemke et al., 2006; Jing et al., 2006). An extension of this technique uses forward tra-

![Figure 20: Average (1994–2012) relative differences (%) of trajectory-mapped MOZAIC/IAGOS profile data minus trajectory-mapped ozonesonde data (Osman et al., paper in preparation). Variations with latitude in the left-hand plot are likely due to differences in sonde type and preparation, which may cause biases of several percent. When averaged over latitude, sonde measurements are about 5 ± 1% higher, in the lower troposphere, and 8 ± 1% higher in the upper troposphere, consistent with the average biases found from UV-referenced sonde intercomparison studies (Section 4.3). DOI: https://doi.org/10.1525/elementa.376.f20](https://doi.org/10.1525/elementa.376.f20)
jectory model calculations or potential vorticity mapping with MLS data (Schoeberl et al., 2007; Yang et al., 2007), to produce stratospheric column estimates with higher horizontal resolution, suitable for producing a daily TOR product. Winds are from NASA’s Modern Era Retrospective Reanalysis (MERRA). Similarly, the Global Modeling and Assimilation Office (GMAO) assimilated ozone product is produced by using OMI and MLS retrievals as input to the global data assimilation system used to produce MERRA (Wargan et al., 2015). Tropospheric ozone has also been derived from the TOMS data alone by assuming the longitudinal distribution of stratospheric ozone (Kim et al., 1996; Hudson and Thompson, 1998).

Cloud differential methods employ the fact that, particularly in the tropics, the tops of the highest clouds are essentially at the tropopause, and so the tropospheric ozone column can be found from the difference in total ozone measured in adjacent cloudy and cloud-free pixels. The Convective Cloud Differential (CCD) method (Ziemke et al., 1998) or the Cloud-Clear Pair (CCP) method (Newchurch et al., 2003) use this approach with TOMS. This method has been applied to OMI, GOME, GOME-2, and SCIAMACHY data (Ziemke et al., 2017; Valks et al., 2014; Leventidou et al., 2016). A second method, called “cloud slicing” (Ziemke et al., 2001, 2003, 2009, 2017), uses measurements of above-cloud column ozone together with cloud-top pressure data to derive ozone column amounts in the upper troposphere. Used in combination, these methods can estimate 400 to 1000 hPa lower tropospheric column ozone. Similarly, lower tropospheric ozone amounts near mountainous regions have been derived from TOMS data using a topographic contrast method (Jiang and Yung, 1996; Kim et al., 1996; Newchurch et al., 2001), and tropical tropospheric ozone has been derived from TOMS data alone, based on differences in ozone-column retrieval sensitivity as a function of scan angle (Kim et al., 2001).

More recent satellite instruments with higher spectral resolution and broader spectral coverage retrieve tropospheric ozone directly from the backscattered radiances in the Hartley (200–320 nm) and Huggins (320–350 nm) bands. Information on the ozone vertical distribution is derived from the effective scattering depth at different wavelengths, and also from the temperature dependence of the ozone absorption cross-sections in the Huggins bands, which separates ozone in the warmer troposphere from colder stratospheric ozone (Chance et al., 1997). GOME profiles have been retrieved at 20 or more layers from the surface to ~60 km using the OE technique (Munro et al., 1998; Hoogen et al., 1999; van der A et al., 2002; Liu et al., 2005), Tikhonov-Philips (TP) regularization (Hasekamp and Landgraf, 2001), and neural networks (Del Frate et al., 2002; Müller et al., 2003). Layer values are not independent, however; total degrees of freedom for signal (DFS) are about 5–6.5, with most in the stratosphere; only about 1 independent point is retrieved in the troposphere. These algorithms are reviewed and compared in Meijer et al. (2006). Similar methods have been applied to GOME-2 (Cai et al., 2012; van Peet et al., 2014; Miles et al., 2015; van Oss et al., 2015), SCIAMACHY (Sellitto et al., 2012a, b), OMI (Kroon et al., 2011; Liu et al., 2010a, b; Mielonen et al., 2015; Sellitto et al., 2011; Di Noia et al., 2013), and OMPS (Bak et al., 2017). Table S-11 compares GOME and OMI retrievals.
The Tropospheric Emission Spectrometer (TES) 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.

The Infrared Atmospheric Sounding Interferometer (IASI) instrument is also an FTIR, similarly operating in the 3.7 to 15.5 µm spectral range, but with 0.5 cm$^{-1}$ spectral resolution (see Table S-11 for details). Ozone profiles are retrieved with similar vertical resolution in the troposphere to that of TES. Table 11 summarizes the characteristics of these satellite data products. More detailed descriptions are found in the Supplemental Material (Text S-12).

Table 11: Summary of characteristics of tropospheric ozone satellite and residual measurement products. Resolution is given as nadir footprint and either tropospheric column ozone (TCO) or degrees of freedom for signal (DFS) in the troposphere. DOI: https://doi.org/10.1525/elementa.376.t11

<table>
<thead>
<tr>
<th>Product</th>
<th>Dates</th>
<th>Type</th>
<th>Coverage</th>
<th>Resolution</th>
<th>Sampling</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOR</td>
<td>1979–2005</td>
<td>Residual TOMS +SAGE or SBUV</td>
<td>Global w/o polar night</td>
<td>1° × 1.25° TCO</td>
<td>Monthly</td>
<td>Fishman et al. (2003)</td>
</tr>
<tr>
<td>OMI/MLS</td>
<td>2004–2015</td>
<td>Residual OMI-MLS</td>
<td>Global w/o polar night</td>
<td>1° × 1.25° TCO</td>
<td>Monthly</td>
<td>Ziemke et al. (2006)</td>
</tr>
<tr>
<td>OMI/MLS (GMAO DA)</td>
<td>2005–2014</td>
<td>Assimilated product</td>
<td>Global w/o polar night</td>
<td>2° × 2.5° TCO</td>
<td>Daily</td>
<td>Wargan et al. (2015)</td>
</tr>
<tr>
<td>GOME-1,2, SCIA CCD</td>
<td>1996–2012</td>
<td>Cloud differential</td>
<td>Tropics</td>
<td>2.5° × 5° TCO</td>
<td>Monthly</td>
<td>Leventidou et al. (2016)</td>
</tr>
<tr>
<td>GOME</td>
<td>1995–2003</td>
<td>UV spectral fitting, neural network</td>
<td>Global w/o polar night</td>
<td>960 × 80 km ≤1.2 DFS</td>
<td>3-day</td>
<td>Munro et al. (1998); Liu et al. (2005); Müller et al. (2003)</td>
</tr>
<tr>
<td>GOME-2</td>
<td>2007–present</td>
<td>UV spectral fitting</td>
<td>Global w/o polar night</td>
<td>40 × 80/640 km ≈1 DFS</td>
<td>Daily</td>
<td>van Oss et al. (2015)</td>
</tr>
<tr>
<td>GOME-2</td>
<td>2007–present</td>
<td>UV spectral fitting</td>
<td>Global w/o polar night</td>
<td>160 × 160 km ≈1 DFS</td>
<td>Daily</td>
<td>Miles et al. (2015)</td>
</tr>
<tr>
<td>OMI profile</td>
<td>2004–present</td>
<td>UV spectral fitting</td>
<td>Global w/o polar night</td>
<td>13 × 48 km ≤1.2 DFS</td>
<td>Daily</td>
<td>Kroon et al. (2011)</td>
</tr>
<tr>
<td>OMI profile</td>
<td>2004–present</td>
<td>UV spectral fitting</td>
<td>Global w/o polar night</td>
<td>52 × 48 km ≤1.2 DFS</td>
<td>Daily</td>
<td>Liu et al. (2010a, b), Huang et al. (2017, 2018)</td>
</tr>
<tr>
<td>TES</td>
<td>2004–present</td>
<td>IR spectral fitting</td>
<td>50S to 70N, 16 tracks</td>
<td>5 × 8 km ≤1.6 DFS</td>
<td>2-day</td>
<td>Nassar et al. (2008); Boxe et al. (2010)</td>
</tr>
<tr>
<td>IASI-LISA</td>
<td>2007–present</td>
<td>IR spectral fitting</td>
<td>Global</td>
<td>12 × 25 km ≤1.6 DFS</td>
<td>Twice daily</td>
<td>Dufour et al. (2012)</td>
</tr>
<tr>
<td>IASI-SOFRID</td>
<td>2007–present</td>
<td>IR spectral fitting</td>
<td>Global</td>
<td>12 × 25 km²</td>
<td>Twice daily</td>
<td>Barret et al. (2011)</td>
</tr>
<tr>
<td>OMI+TES</td>
<td>2004–2008</td>
<td>IR + UV spectral fitting</td>
<td>82S to 82N, 16 tracks</td>
<td>13 × 48 km 2.0 DFS</td>
<td>2-day</td>
<td>Fu et al. (2013)</td>
</tr>
<tr>
<td>IASI+GOME-2</td>
<td>2009–present</td>
<td>IR + UV spectral fitting</td>
<td>Global w/o polar night</td>
<td>12 × 25 km 1.7 DFS</td>
<td>Daily</td>
<td>Cuesta et al. (2013)</td>
</tr>
</tbody>
</table>
optimizes the retrieval constraints to maximize the DOFS in the lower troposphere (Dufour et al., 2010, 2012, 2015). These IASI (TP) retrievals are able to depict the horizontal distribution of ozone plumes within the lower troposphere (Figure 22) with a relative maximum of sensitivity typically between 3 to 4 km, in case of positive thermal contrasts (i.e. over land during summer) but with limited sensitivity to near surface ozone.

Multispectral measurements can also enhance retrieval sensitivity to lower tropospheric ozone. Examples include: UV radiance + polarization (Hasekamp and Landgraf, 2002), UV + IR (Worden et al., 2007b; Landgraf and Hasekamp, 2007), UV + IR + VIS (Natraj et al., 2011), and VIS + IR (Hache et al., 2014) measurements using OE or TP regularization techniques. The polarization measurements in the UV show higher sensitivity to ozone in the troposphere than to ozone in the stratosphere.

IASI+GOME2: A multispectral satellite approach that simultaneously fits IASI measurements in the thermal IR and co-located UV spectra from GOME-2, at the IASI spatial resolution (12 × 25 km) has allowed the spaceborne observation of ozone plumes below 3 km, both over land and ocean (Cuesta et al., 2013). Sensitivity in the surface-3 km layer peaks at 2 to 2.5 km asl over land (Figure 22), while the DOFS for this layer are 0.35, 40% more than IASI (TP). Validation with ozonesondes in 2009–10 show that ozone is retrieved in this layer with a mean bias of 4% and a precision of 17%, when smoothing by the retrieval vertical sensitivity (9% mean bias and 27% precision for direct comparisons).

TES+OMI: A similar multispectral approach combines radiances from TES and OMI (Fu et al., 2013). The joint TES/OMI retrieval provides 2 DOFs in the troposphere with approximately 0.4 DOFs for near surface ozone (surface to 700 hPa).

Figures 23 and 24 display bias and uncertainty information for satellite retrievals and data products, from published validation studies. Systematic biases, however, can vary by region (see TOAR-Climate, Gaudel et al., 2018). In all cases evaluations were with respect to ECC sondes; there are very few comparisons with other tropospheric data sources (e.g. Safieddine et al., 2016). The published

![Figure 22: Satellite retrieval of ozone on 19 August 2009 over Europe (a) in the lower troposphere (up to 6 km asl) derived from the IASI Tikhonov-Philips approach (Eremenko et al., 2008) and (b) in the lowermost troposphere (up to 3 km asl) from the IASI+GOME2 multispectral synergism (Cuesta et al., 2013). Examples of averaging kernels in the troposphere for a satellite pixel over land for (c) IASI (TP) and (d) IASI+GOME2. DOI: https://doi.org/10.1525/elementa.376.f22](https://doi.org/10.1525/elementa.376.f22)
Biases and uncertainties have therefore been adjusted with the biases derived for ECC sondes in Section 4.3. Evaluations are in general single averages over a short period of time; there are very few published evaluations of instrument drift with respect to time (Huang et al., 2017; Boynard et al., 2018; Gaudel et al., 2018).

It is apparent from Figures 23 and 24 that while biases are fairly modest, ranging between −10% and +20%, but often much smaller, standard deviations are large, compared to those of the other measurement systems discussed above: about 10–30%, versus 5–10% for sondes, aircraft, lidar and ground-based FTIR. Nevertheless, as...
satellite data products offer global or near-global coverage with few gaps, their value is correspondingly large. Other measurement systems suffer from errors of representativeness, when point measurements are interpolated or extrapolated to infer information at points other than the place and time of the original measurement.

5. Representativeness

Ozone is a highly reactive secondary pollutant with many processes such as photochemical formation, deposition and titration playing a role in determining atmospheric mole fractions. Although several thousand ground-based stations measure ozone concentrations at high temporal frequencies world-wide (TOAR-Surface Ozone Database), the globe is nevertheless undersampled, since surface ozone over land surfaces may vary locally on scales of a few kilometres or less. (Here the spatial representativeness of observations down to an urban/regional scale is considered. The issue of measurements being representative of the well-mixed boundary layer addressed earlier with Criterion 3, is on a smaller scale and not addressed in this section.) Sites are unevenly distributed, with relatively few in the tropics and southern mid-latitudes (Sofen et al., 2016, TOAR-Surface ozone database). Spatial representativeness is therefore often the largest source of uncertainty in the use of ground-based data. For a well-calibrated analyzer, measurement uncertainty is in the 1-nmol mol⁻¹ range (Section 2.2.2), while two locations with different site characteristics (e.g. urban vs rural) may show average differences an order of magnitude larger. Land use may also change with time, complicating trend analysis by inducing changes that do not reflect background ozone trends. Remote baseline sites with minimal influence of local processes are usually representative of a larger area than sites located for catching the plumes of particular sources (e.g. monitoring stations at curbsides) which reflect local conditions. The application of the data may also determine the area of representativeness, as correlation lengths are typically longer for monthly averages than for daily data (e.g. Sofen et al., 2016).

Standard attempts to assess the representativeness of in-situ air quality monitoring stations classify the sites into categories like urban, suburban, rural, or remote, in terms of their exposure to sources and sinks (e.g. European Union, 2008). These classifications are somewhat qualitative. More quantitative approaches use station metadata. Ozone concentrations at urban stations are strongly controlled by NOx titration, and so population density (which can be seen as a proxy for NOx emissions) is often used to classify sites as urban, suburban, and rural. The intensity and nature of sources (traffic, industrial or background) can be used to refine the classification. TOAR-Surface ozone database has developed a globally consistent classification scheme based on population density, nighttime light intensity, OMI tropospheric column NO₂, and station altitude. Due to the need to define thresholds for each of these parameters, this classification based on metadata is still partly subjective.

Beyond the use of ozone monitoring station data for trends, their broader use for model evaluation and data assimilation raises the question of objectively quantifying station spatial representativeness, i.e., how a single measurement is related to its spatial surroundings (Spangl et al., 2007). Methods have been developed to characterize station representativeness with more objective criteria. Janssen et al. (2012) show that using a classification parameter based on land use improves model validation results by ~20%. Henne et al. (2010) have proposed a classification based on an explicit estimation of emissions, deposition and transport influencing a particular station. They use population density as a proxy of emissions, land cover from the Wesely (1989) dry deposition parameterization to derive deposition fluxes, and a Lagrangian trajectory approach to evaluate transport impact. Methods based solely on the characteristics of the measurements themselves, especially the diurnal profile amplitude, have also been employed (Flemming et al., 2005; Tarasova et al., 2007; Gauert et al., 2014). Urban stations exhibit larger diurnal amplitude (due to strong night time ozone loss and strong daily photochemical production) while remote and high elevation stations show much flatter diurnal profiles. Joly and Peuch (2012) have refined this approach by adding other parameters such as the weekend effect to describe the station characteristics. These methods are used to evaluate models with objectively classified station data (Marécal et al., 2015), and to build assimilation systems of observational data. Solazzo and Galmari (2015) have proposed an alternative approach employing spectral analysis of the ozone time series, and correlation analysis of different spectral components. They find that the area of representativeness is generally very non-isotropic and quite heterogeneous (as also shown by the catchment areas of Henne et al. (2010)). Noting that certain spectral components of ozone variability showed discontinuities between countries (Europe) and networks (North America), they discard those as noise, and note an improvement in evaluated model performance of ~5%. Schutgens et al. (2016) have shown how high sub-grid variability, such as is prominent with ozone observations, can result in imperfect matches between individual stations and the regionally averaged values.

All these methods show grouping significantly different from the metadata approach, and also demonstrate both the value and challenge of representative station classifications. These findings also suggest that trend analyses using the large number of observations available from surface ozone sites benefit from applying objective classification methods (TOAR-Surface ozone database; Fleming et al., 2018, hereinafter TOAR-Health).

In the free troposphere local effects are less important and representativeness areas are larger. Typical correlation lengths in the free troposphere, for individual measurements, are about 500 km (Liu et al., 2009; Nastrom, 1977). However, the distances between ground-based observing sites are usually larger than this (Figures 8, 15, 16, 21), and as in the case of surface data, sites are unevenly distributed. Observations are also less frequent, so the ozone distribution is in general undersampled, both in space and time. Several authors have noted that this raises representativeness issues (Logan et al., 2012; Tilmes et al.,...
examined individual ozone soundings. Similarly, Eriksson and Chen (2002) found vertical correlation lengths of 2–5 km in ozonesonde data, while Sofieva et al. (2004) found vertical correlation lengths of ~1 km for small-scale deviations from a smoothed profile.

Like spatial representativeness, temporal autocorrelation in ozone exists on a variety of timescales. Hourly averaged ozone measurements are uncorrelated if more than a few days apart (Galbally et al. 1986; Liu et al. 2009; Lehman et al., 2004), but monthly averages nevertheless show significant autocorrelation (e.g. Tarasick et al., 2005, Oltmans et al., 2006). The temporal persistence, or autocorrelation can differ by location and can be closely linked to weather patterns and variations in sources and sinks. Frequent observations can allow quantification of temporal autocorrelation at most relevant timescales.

A number of research groups have developed techniques to optimize and evaluate proposed network changes that can be applied to tropospheric ozone monitoring. Spatial coherence has been studied by McBratney (1981) and Yost (1982) who identified “areas of influence.” Dantzig et al. (1963) and Cressie (1985) pioneered research efforts on optimization of networks, work that has been carried forward with more respect for the specific goals of environmental monitoring by Nychka and Salzman (1998), among others. Weatherhead et al. (2017) have addressed the challenge of designing monitoring systems when realistic constraints, including financial budget constraints, must be considered.

6. Conclusions and recommendations for design of a future global observational program

From the earliest measurements in the 19th century, both measurement methods and the portion of the globe observed have evolved and changed significantly. The historical methods have different uncertainties and biases, and the data records differ with respect to coverage (space and time), information content, and representativeness. There are significant uncertainties with the 19th and early 20th century measurements with regard to representative- ness, and interfering gases. SO, levels in particular appear to have been quite high in urban areas, and may have negatively biased urban ozone measurements. There is therefore no unambiguous evidence of very low ozone values in the 19th century.

There are 49 and 11 sets of measurements of surface ozone, by KI and spectroscopic methods respectively, made before the mid-1970s and suitable for this historical analysis. Values of ozone absorption coefficients used before 1960 varied, however, and caused ozone to be underestimated by up to 11%. Overall, the 60 available datasets during 1896–1975 indicate an ozone mole fraction in the well-mixed unpolluted boundary layer that lies in the range 22 to 26 nmol mol\(^{-1}\). Comparison with modern measurements from the TOAR database suggests that surface ozone has increased by 32–71%, with large uncertainty, in rural air in the temperate and polar zones of the northern hemisphere, and by much smaller amounts in the southern hemisphere. This estimate depends much more on the modern region chosen for comparison than on
Table 12: Scientific tasks, goals, and requirements for future tropospheric ozone monitoring. DOI: https://doi.org/10.1525/elementa.376.t12

<table>
<thead>
<tr>
<th>Scientific task or question</th>
<th>Goals and Requirements</th>
<th>Station location</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-term tropospheric ozone monitoring</td>
<td>Detection of long term ozone distribution changes, ozone transport changes. Need decadal stability of ~1 nmol mol(^{-1}). Vertical profiling important.</td>
<td>Multiple sites in different regions and land use classifications. Choice of sites should be guided by objectively quantified station spatial representativeness.</td>
<td>Current global network is unevenly distributed and covers only ~25% of the globe (TOAR Surface ozone database). Sites with long-term records are very important.</td>
</tr>
<tr>
<td>Air quality model validation</td>
<td>Moderate accuracy and precision, preferably 3–5% level. Need vertical resolution of ~0.2 km or better. Need hourly time resolution, at least for short (campaign) periods. Flux measurements.</td>
<td>Multiple sites in different regions. Choice of sites should be guided by objectively quantified station spatial representativeness. Collocated profile measurements of other species desirable.</td>
<td>Measurement campaigns at multiple sites are desirable. Measurements of surface deposition fluxes for different environments are needed (Hardacre et al., 2015; Bariteau et al., 2010; Luhar et al. 2017, 2018).</td>
</tr>
<tr>
<td>Chemical data assimilation</td>
<td>Moderate accuracy and precision, preferably 3–5% level. Vertically-resolved measurements desirable. Daily or better time resolution.</td>
<td>Many sites in different regions. Choice of sites should be guided by objectively quantified site spatial representativeness. Satellite, surface monitor, aircraft data.</td>
<td>Can we increase the impact of sparse measurements? Aircraft, lidar, ozonesondes have small measurement errors, relative to model error. Data impact should therefore be significant.</td>
</tr>
<tr>
<td>Satellite ozone data validation</td>
<td>High accuracy and high precision, preferably 2–3% level. Profile (free tropospheric) information required.</td>
<td>Location should represent different observational conditions (latitude, ozone profiles, etc.) and preferably have related measurements (surface O(_3), total O(_3), aerosol)</td>
<td>Data quality of prime importance; periodic re-evaluation needed.</td>
</tr>
<tr>
<td>How do ozone levels in the free troposphere affect levels in the planetary boundary layer (PBL)?</td>
<td>Measurement campaigns with vertical sounding at a resolution down to a few hours – lidar, satellite, sonde and other met measurements, possibly at multiple sites.</td>
<td>Sites in different latitude bands. Sites with multi-year measurement records are of value for background climatology.</td>
<td>Important to interpreting satellite measurements, which are primarily sensitive to ozone above the PBL (Crawford and Pickering, 2014; Martins et al. 2015).</td>
</tr>
</tbody>
</table>

(Contd.)
<table>
<thead>
<tr>
<th>Scientific task or question</th>
<th>Goals and Requirements</th>
<th>Station location</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model processes: Can we properly model and quantify long range transport (e.g. Stohl and Trickl, 1999), and stratosphere-troposphere transport events and their effects on ozone levels in both the free troposphere and PBL? How well is plume mixing represented? (Osman et al., 2016; Trickl et al., 2014, 2016; Eastham and Jacob, 2017)</td>
<td>Measurement campaigns with lidar, global satellite observations, mesoscale or ozonesonde (daily launches).</td>
<td>Sites in different latitude bands. Sites near tropopause breaks. Sites with multi-year measurement records are of value for background climatology.</td>
<td>Global flux is moderately well-estimated by models; regional flux varies (Stohl et al., 2003a, b) and is not well characterized by scattered observations. Models show great improvement in capturing tropopause folding events (e.g. Trickl et al., 2010, 2011; He et al., 2011; Langford et al., 2015; Lin et al., 2015a), in part due to better spatial resolution, but effects on surface ozone may be underestimated (Zhang et al., 2011; Lin et al., 2012a, b; Hess and Zbinden, 2013; Zanis et al., 2014; Lefohn et al., 2014; Akritidis et al., 2016).</td>
</tr>
<tr>
<td>Climate change effects: How will temperature increases affect ozone photochemistry, transport? Will more forest fires increase photochemical ozone production? Will the Brewer-Dobson circulation increase stratospheric ( \mathrm{O}_3 ) transport to the troposphere? What is causing the observed trend in Arctic surface ozone depletion events (ODE)?</td>
<td>Detection of long term ozone distribution changes, ozone transport changes. Need vertical resolution of ~0.2 km or better, to separate PBL, mid-troposphere and UTLS processes. Need global coverage, and long term stability of ~1 nmol mol(^{-1}). Long term measurements in the background atmosphere of ozone precursors.</td>
<td>Several sites in different latitude bands. Sites with long-term measurement records are preferable.</td>
<td>Climate change is expected to increase planetary wave activity and so produce an increased Brewer-Dobson circulation (e.g. Butchart et al., 2006; Butchart, 2014). In the lower troposphere increases will be offset by losses due to reaction with water vapour to produce OH (Stevenson et al., 2013). At some Arctic sites a long-term increase is observed in surface ODEs (Tarasick et al., 2014), possibly related to changes in snow and sea ice cover (Simpson et al., 2007).</td>
</tr>
<tr>
<td>Oil and gas extraction: Will increased activity in oil and/or gas fracking release NO(_x) and VOCs, leading to high ozone formation?</td>
<td>Measurement campaigns with surface monitors, measurements of other species; possibly at multiple sites. Vertical profiling would be useful.</td>
<td>Near extraction areas. Sites with multi-year measurement records are of value for background climatology.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the historical data, as when some of the historical datasets judged less reliable are omitted the results are quite similar. The ±20% range of the estimated increase comes primarily from the variability of present day surface ozone. Data representativeness thus seems to be the more important source of uncertainty.

Based on a more limited, but completely independent set of data, free tropospheric ozone appears to have also changed by a similar amount, in the mid-latitudes of the Northern Hemisphere. In spite of the extensive efforts to identify and evaluate early ozone data records, other data may be available that could pass the selection criteria from this study.

Representativeness, especially for surface sites, is a potential source of significant biases, which are difficult to quantify. Recent research into objective methods of determining areas of representativeness has made valuable progress in reducing this source of uncertainty.

The great majority of validation and intercomparison studies of free tropospheric ozone measurement methods are undertaken with ECC ozonesondes. ECC sondes have been compared to UV-absorption measurements in a number of intercomparison studies. The sondes show a modest (~1–5%) high bias in the troposphere, with an uncertainty of 5%, but no evidence of an instrument change with time. Other methods – Umkehr, lidar, FTIR and UV instruments on commercial aircraft – all show modest low biases relative to the ECCs, and so, using ECC sondes as a transfer standard, all appear to agree to within 1σ with the UV-absorption standard.

Relative to the UV standard, BM sondes show a 20% increase in sensitivity to tropospheric ozone from 1970–1995. The KC sondes show a smaller increase of 5–10%. In combination with the gradual shift of the global network to ECC sondes, this will, if uncorrected, introduce an erroneous positive trend in the free troposphere, to analyses based on sondes data.

Satellite biases are often larger than those of other free tropospheric measurement systems, ranging between −10% and +20%, and standard deviations are large: about 10–30%, versus 5–10% for sondes, aircraft instruments, lidar and ground-based FTIR. Although measurement drift has been examined extensively for satellite measurements of stratospheric ozone (Harris et al., 2015; Hubert et al., 2016), there is relatively little information on temporal changes of bias for satellite measurements of tropospheric ozone. This is an evident area of concern, and one that must be addressed if satellite retrievals are used for trend studies (TOAR-Climate, Gaudel et al., 2018).

The importance of ECC sondes as a transfer standard for satellite validation means that more effort should be placed on understanding and reducing their uncertainties. The overall accuracy of the global ozonesonde network has improved: at many important sites the historical record has been homogenized, by correcting for known changes in station records (e.g. Tarasick et al., 2016; van Malderen et al., 2016; Witte et al., 2017, 2019; Sterling et al., 2018, Thompson et al., 2018). In addition, spatial inhomogeneity has been reduced by adopting strict standard operating procedures (Smit and ASOPOS panel, 2011). These continuing efforts should proceed in tandem with research to better quantify systematic and random uncertainty in ECC data and to understand changes therein. The global network is also unevenly distributed (Figures 15 and 16), and so additional sites, in southern midlatitudes, North Africa, Asia and other areas with limited coverage, are recommended, possibly as a measurement campaign (Thompson et al., 2011; Lebelveld et al., 2002).

Planning future observations of ozone will need to make careful use of known spatial and temporal coherence. Decisions concerning spatial choices and temporal frequency need to be made with consideration for measurement accuracy and co-location with other observations, including NOx, windspeed and direction, and other relevant information needed to understand both ozone and its sources. The integration or merging of data from different platforms, which has had little attention to date, can improve coverage.

Although tropospheric ozone monitoring has evolved from sporadic measurements at a few locations to extensive, well-calibrated networks with formal international collaboration (e.g. Schultz et al., 2015), as well as global satellite observations, it is not comprehensive, nor evenly distributed. It is recommended that the design of the global observational program in the future be guided by several current and emerging scientific issues (Table 12). Each method of observation has its inherent advantages and limitations, and so different techniques will continue to complement and support each other. For example, satellite observations are likely to be of great importance to ozone data assimilation, but ozonesonde and lidar profiles are required for satellite product evaluation. It is to be hoped that commercial aircraft monitoring will be expanded, to close monitoring gaps with reliable, well-calibrated measurements. Some modest improvements in the distribution of ground-based observing sites could yield significant benefits in global coverage. International cooperation and data sharing will be of paramount importance, as the TOAR project has demonstrated.

Data Accessibility Statement

No new measurements were made for this review article. All datasets discussed in the text were obtained from the published scientific literature. Lidar, FTIR, Umkehr and ozonesonde data used in this publication are publicly available, from the World Ozone and UV Data Centre (http://www.woudc.org) and the Network for the Detection of Atmospheric Composition Change (http://www.ndacc.org), and MOZAIC/IAGOS data at (http://www.iagos.org). Satellite data products are available at the URLs listed in Supplemental Material, Text S-12.

Supplemental files

The supplemental files for this article can be found as follows:

- **Text S-1:** Instrumental method: Gas phase titration (GPT). DOI: https://doi.org/10.1525/elementa.376.s1
- **Text S-2:** Comparison of modern surface ozone observations with historical records. DOI: https://doi.org/10.1525/elementa.376.s1
• **Text S-3.** Data selection identified as questionable under Criterion 4. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-4.** A note on units and nomenclature. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-5.** Standard deviations used in representing data sets. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-6.** Historical measurements of urban and suburban ozone in northern Europe. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-1.** Some historical measurements of urban and suburban ozone in northern Europe. DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-1.** Some historical measurements of urban and suburban ozone in northern Europe. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-2.** Historical measurements of rural background surface ozone at low altitudes in northern temperature regions. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-7.** Notes on Tables 4–6, 8, S1. DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-2.** 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. DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-3.** 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). DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-8.** Analysis of changes in surface ozone. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-9.** Umkehr retrieval algorithms. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-3.** (left) Umkehr averaging kernels. (right) Uncertainties for standard Umkehr levels. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-10.** Analysis of ozonesonde intercomparison results. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-4.** Left: Average estimated uncertainty of Umkehr averaging kernels. (right) Historical measurements of rural back. 

• **Figure S-5.** Published intercomparison results of ECC sondes without adjustments. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-6.** As Figure S5, for the upper troposphere. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-7.** Intercomparison results of BM sondes, without adjustments. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-8.** As Figure S7, for the upper troposphere. DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-9.** Sites and characteristics existing of tropospheric ozone lidar systems. DOI: https://doi.org/10.1525/elementa.376.s1

• **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. DOI: https://doi.org/10.1525/elementa.376.s1

• **Figure S-10.** Average (1994–2012) relative differences (%) between trajectory-mapped MOZART/IGOS profile data and trajectory-mapped ozonesonde data. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-11.** FTIR retrieval and information content. DOI: https://doi.org/10.1525/elementa.376.s1

• **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). DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-10.** Sites and characteristics existing of tropospheric FTIR stations. DOI: https://doi.org/10.1525/elementa.376.s1

• **Text S-12.** Tropospheric ozone satellite and residual measurements: Instrument and method details. DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-11.** Comparison of retrieval algorithms and datasets. DOI: https://doi.org/10.1525/elementa.376.s1

• **Table S-12.** Comparison of GOME, OMI, TES and IASI specifications. DOI: https://doi.org/10.1525/elementa.376.s1

**Acknowledgements**

The authors thank the many observers who, over many years, obtained the ozone measurements used in this study. Their careful work has stood the test of time and is gratefully acknowledged.

**Funding information**

C. Vigouroux was supported financially by the EU H2020 project GAIA-Clim (No 640276). M. Steinbacher acknowledges funding from the GAW Quality Assurance/Science
Activity Centre Switzerland (QA/SAC-CH), which is supported by MeteoSwiss and Empa. OHP observations are funded by the NDACC French program. The Laboratoire Inter-universitaire des Systèmes Atmosphériques (LISA) acknowledges the support from CNES (Centre National des Etudes Spatiales)/TOSCA (Terre Océan Surface Continentale Atmosphère), PNTS (Programme National de Télédétection Spatiale) and ANR (Agence Nationale de la Recherche – project: ANR-15-CE04-0005) for the development and production of ozone observations from IASI+GOME-2 and IASI. The MLS, OMI and TES projects are supported by the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Aura Program. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. The National Center for Atmospheric Research is sponsored by the National Science Foundation. J.W. Hannigan is supported under contract by NASA.

Competing interests
The authors have no competing interests to declare.

Author contributions
• All authors contributed to interpretation, revision of all aspects of the paper
• Other authors contributed text, analysis and expert comment in their areas of particular expertise: historiographic datasets (IEG, TJW, SJO, JS, MS); statistical analysis (KLC); ground-based methods (IEG, MS); Umkehr (IP); ozonesondes (DWT, AMT, MO, JL, SJO, BH); lidar (GA, TL, TT, MGM); FTIR (CV, JH, OG); aircraft (JS, VT, MO); representativeness (MS, EW, ML, GF); satellite methods (JZ, XL, AG, HW, JC, GD, JLN), stratospheric intrusions (PZ), as well as comments throughout the paper
• All authors contributed to interpretation, revision of several drafts, and approved the submitted version for publication

References


Chung, YS and Dann, T. 1985. Observations of stratospheric ozone at the ground level in Regina, Canada. Atmos. Environ. 19: 157–162. DOI: https://doi.org/10.1016/0004-6981(85)90147-7


Cooper, OR, Parrish, DD, Stohl, A, Trainer, M, Nédélec, P, Thouret, V, Cammas, JP, Oltmans, SJ, Johnson, BJ, Tarasick, D, LeBlanc, T, McDermid, I, Jaffe, D,


Dufour, G, Eremenko, M, Orphal, J and Flaud, J-M. 2010. IASI observations of seasonal and day-to-day variations of tropospheric ozone over three highly populated areas of China: Beijing, Shanghai, and Hong Kong. Atmos. Chem. Phys. 10: 3787–3801. DOI: https://doi.org/10.5194/acp-10-3787-2010

Eastham, SD and Jacobs, DJ. 2017. Limits on the ability of global Eulerian models to resolve intercontinental transport of chemical plumes. Atmos. Chem. Phys. 17: 2543–2553. DOI: https://doi.org/10.5194/acp-17-2543-2017


Gaudel, A, Ancellet, G and Godin-Beekmann, S. 2015. Analysis of 20 years of tropospheric ozone vertical profiles by lidar and ECC at Observatoire de Haute Provence (OHP) at 44°N, 6.7°E. *Atmos. Environ.* 113: 78–89. DOI: https://doi.org/10.1039/b305727a


previous measurements around 293 K. Atmos. Meas. Tech. 7: 609–624. DOI: https://doi.org/10.5194/amt-7-609-2014


Henne, S, Brunner, D, Folini, D, Solberg, S, Klausen, J and Buchmann, B. 2010. Assessment of parameters describing representativeness of air quality in-situ measurement sites. Atmos. Chem. Phys. 10: 3561–3581. DOI: https://doi.org/10.5194/acp-10-3561-2010


Hering, WS and Dütsch, HU. 1965. Comparison of chemiluminescent and electrochemical ozonesonde
observations. J. Geophys. Res. 70(22): 5483–5490. DOI: https://doi.org/10.1029/JZ070i022p05483


Langford, AO; Senff, CJ; Alvarez, RJ; Brioude, J; Cooper, OR; Holloway, JS; Lin, MY; Marchbanks, RD; Pierce, RB; Sandberg, SP; Weickmann, AM and Williams, EJ. 2015. An overview of the 2013 Las Vegas Ozone Study (LVOS): Impact of stratospheric intrusions and long-range transport on surface air quality. Atmos. Environ. DOI: https://doi.org/10.1016/j.atmosenv.2014.08.040


Leblanc, T; Brewer, MA; Wang, PS; Granados-Muñoz, MJ; Strawbridge, KB; Travis, M; Firaniski, B; Sullivan, JT; McGee, TJ; Surnick, GK; and Twigg, LW. 2018. Validation of the TOLNet lidars: the Southern California Ozone Observation Project (SCOOP). Atmos. Meas. Tech. 11(11): 6137–6162. DOI: https://doi.org/10.5194/amt-11-6137-2018


Lefohn, AS; Malley, CS; Smith, L; Wells, B; Hazucha, M; Simon, H; Naik, V; Mills, G; Schultz, MG; Paolletti, E; De Marco, A; Xu, X; Zhang, L; Wang, T; Neufeld, HS; Musselman, RC; Tarasick, D; Brauer, M; Fung, Z; Tang, H; Kobayashi, K; Sicard, P; Solberg, S; and Gerosa, G. 2018. Tropospheric ozone assessment report: Global ozone metrics for climate change, human health, and crop/ecosystem research. Elem Sci Anth. 6(1): 28. DOI: https://doi.org/10.1525/elementa.279


Lelieveld, J; Berresheim, H; Bormann, S; Crutzen, P; Dentener, FJ; Fischer, H; Feichter, J; Flatau, PJ; Hield, J; Holzinger, R; Kornmann, R; Lawrence, MG; Levin, Z; Markowicz, KM; Mihalopoulos, N; Minikin, A; Ramanathan, V; De Reus, M; Roelefsl, GJ; Scheeren, HA; Schiare, J; Schlager, H; Schultz, M; Siegmund, P; Steil, B; Stephanou, EG; Stier, P; Traub, M; Warneke, C; Williams, J and Ziereis, H. 2002. Global air pollution crossroads over the Mediterranean. Science. Oct 25. 298(5594): 794–9. DOI: https://doi.org/10.1126/science.1075457


Lin, M; Fiore, AM; Horowitz, LW; Langford, AO; Oltmans, SJ; Tarasick, D and Reider, HE. 2015a. Climate variability modulates western US ozone air quality in spring via deep stratospheric intrusions. Nature Communications. 6: 7105. DOI: https://doi.org/10.1038/ncomms8105

Lin, M; Horowitz, LW; Cooper, OR; Tarasick, DW; Conley, S; Iraci, LT; Johnson, B; Leblanc, T; Petropavlovskikh, I and Yates, EL. 2015b. Revisiting the evidence of increasing springtime ozone mixing ratios in the free troposphere over western North America. Geophys. Res. Lett. 42. DOI: https://doi.org/10.1002/2015GL065311

Lin, MAM; Fiore, OR; Cooper, LW; Horowitz, AO; Langford, H; Levy, II; Johnson, BJ; Naik, V; Oltmans, SJ and Senff, CJ. 2012a. Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions. J. Geophys. Res. 117: D00V22. DOI: https://doi.org/10.1029/2012JD018151


Mann, HB and Whitney, DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics.* 18(1): 50–60. DOI: https://doi.org/10.1214/aoms/1177730491


Paneth, FA and Glückauf, E. 1941. Measurement of atmospheric ozone by a quick electrochemical...
method. Nature 147: 614–615. DOI: https://doi.org/10.1038/147614a0


Sofen, ED, Bowdalo, D and Evans, MJ. 2016. How to most effectively expand the global surface ozone observing network. Atmos. Chem. Phys. 16: 1445–1457. DOI: https://doi.org/10.5194/acp-16-1445-2016


Tiwari, VS and Sreedharan, CR. 1973. Ozone concentration studies and ozone flux measurements near the ground at Poona. PAGEOPH. 106: 1124. DOI: https://doi.org/10.1007/BF00881066


deep stratospheric intrusions? Atmos. Chem. Phys. 14(18): 9941–9961. DOI: https://doi.org/10.5194/acp-14-9941-2014


Wargan, K, Pawson, S, Olsen, MA, Witte, JC, Douglass, AR, Ziemke, JR, Strahan, SE and Nielsen, JE. 2015. The global structure of upper...


Welch, BL. 1947. The generalization of “Student’s” problem when several different population variances are involved. *Biometrika.* **34**(1–2): 28–35. DOI: https://doi.org/10.1093/biomet/34.1-2.28

Wesely, M. 1989. Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmos. Environ.* **23**: 1293–1304. DOI: https://doi.org/10.1016/0004-6981(89)90153-4


