

## RESEARCH ARTICLE

# Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations

Martin G. Schultz<sup>1,82</sup>, Sabine Schröder<sup>1</sup>, Olga Lyapina<sup>1</sup>, Owen R. Cooper<sup>2,3</sup>, Ian Galbally<sup>4</sup>, Irina Petropavlovskikh<sup>2,3</sup>, Erika von Schneidemesser<sup>5</sup>, Hiroshi Tanimoto<sup>6</sup>, Yasin Elshorbany<sup>7,8</sup>, Manish Naja<sup>9</sup>, Rodrigo J. Seguel<sup>10</sup>, Ute Dauert<sup>11</sup>, Paul Eckhardt<sup>12</sup>, Stefan Feigenspan<sup>11</sup>, Markus Fiebig<sup>12</sup>, Anne-Gunn Hjellbrekke<sup>12</sup>, You-Deog Hong<sup>13</sup>, Peter Christian Kjeld<sup>14</sup>, Hiroshi Koide<sup>15</sup>, Gary Lear<sup>16</sup>, David Tarasick<sup>17</sup>, Mikio Ueno<sup>15</sup>, Markus Wallasch<sup>18</sup>, Darrel Baumgardner<sup>19</sup>, Ming-Tung Chuang<sup>20</sup>, Robert Gillett<sup>4</sup>, Meehye Lee<sup>21</sup>, Suzie Molloy<sup>4</sup>, Raeesa Moolla<sup>22</sup>, Tao Wang<sup>23</sup>, Katrina Sharps<sup>24</sup>, Jose A. Adame<sup>25</sup>, Gerard Ancellet<sup>26</sup>, Francesco Apadula<sup>27</sup>, Paulo Artaxo<sup>28</sup>, Maria E. Barlasina<sup>29</sup>, Magdalena Bogucka<sup>30</sup>, Paolo Bonasoni<sup>31</sup>, Limseok Chang<sup>32</sup>, Aurelie Colomb<sup>33</sup>, Emilio Cuevas-Agulló<sup>34</sup>, Manuel Cupeiro<sup>35</sup>, Anna Degorska<sup>36</sup>, Aijun Ding<sup>37</sup>, Marina Fröhlich<sup>38</sup>, Marina Frolova<sup>39</sup>, Harish Gadhavi<sup>40</sup>, Francois Gheusi<sup>41</sup>, Stefan Gilge<sup>42,43</sup>, Margarita Y. Gonzalez<sup>44</sup>, Valerie Gros<sup>45</sup>, Samera H. Hamad<sup>46</sup>, Detlev Helmig<sup>47</sup>, Diamantino Henriques<sup>48</sup>, Ove Hermansen<sup>12</sup>, Robert Holla<sup>42</sup>, Jacques Hueber<sup>47</sup>, Ulas Im<sup>49</sup>, Daniel A. Jaffe<sup>50</sup>, Ninong Komala<sup>51</sup>, Dagmar Kubistin<sup>42</sup>, Ka-Se Lam<sup>23</sup>, Tuomas Laurila<sup>52</sup>, Haeyoung Lee<sup>53</sup>, Ilan Levy<sup>54</sup>, Claudio Mazzoleni<sup>55</sup>, Lynn R. Mazzoleni<sup>55</sup>, Audra McClure-Begley<sup>2,3</sup>, Maznorizan Mohamad<sup>56</sup>, Marijana Murovec<sup>57</sup>, Monica Navarro-Comas<sup>44</sup>, Florin Nicodim<sup>58</sup>, David Parrish<sup>2,3</sup>, Katie A. Read<sup>59</sup>, Nick Reid<sup>60</sup>, Ludwig Ries<sup>61</sup>, Pallavi Saxena<sup>62</sup>, James J. Schwab<sup>63</sup>, Yvonne Scorgie<sup>64</sup>, Irina Senik<sup>65</sup>, Peter Simmonds<sup>66</sup>, Vinayak Sinha<sup>67</sup>, Andrey I. Skorokhod<sup>68</sup>, Gerard Spain<sup>69</sup>, Wolfgang Spangl<sup>38</sup>, Ronald Spoor<sup>70</sup>, Stephen R. Springston<sup>71</sup>, Kelvin Steer<sup>72</sup>, Martin Steinbacher<sup>73</sup>, Eka Suharguniyawan<sup>74</sup>, Paul Torre<sup>75</sup>, Thomas Trickl<sup>76</sup>, Lin Weili<sup>77</sup>, Rolf Weller<sup>78</sup>, Xu Xiaobin<sup>79</sup>, Likun Xue<sup>80</sup> and Ma Zhiqiang<sup>81</sup>

In support of the first Tropospheric Ozone Assessment Report (TOAR) a relational database of global surface ozone observations has been developed and populated with hourly measurement data and enhanced metadata. A comprehensive suite of ozone data products including standard statistics, health and vegetation impact metrics, and trend information, are made available through a common data portal and a web interface. These data form the basis of the TOAR analyses focusing on human health, vegetation, and climate relevant ozone issues, which are part of this special feature.

Cooperation among many data centers and individual researchers worldwide made it possible to build the world's largest collection of *in-situ* hourly surface ozone data covering the period from 1970 to 2015. By combining the data from almost 10,000 measurement sites around the world with global metadata information, new analyses of surface ozone have become possible, such as the first globally consistent characterisations of measurement sites as either urban or rural/remote. Exploitation of these global metadata allows for new insights into the global distribution, and seasonal and long-term changes of tropospheric ozone and they enable TOAR to perform the first, globally consistent analysis of present-day ozone concentrations and recent ozone changes with relevance to health, agriculture, and climate.

Considerable effort was made to harmonize and synthesize data formats and metadata information from various networks and individual data submissions. Extensive quality control was applied to identify questionable and erroneous data, including changes in apparent instrument offsets or calibrations. Such data were excluded from TOAR data products. Limitations of *a posteriori* data quality assurance are discussed. As a result of the work presented here, global coverage of surface ozone data for scientific analysis has been significantly extended. Yet, large gaps remain in the surface observation network both in

terms of regions without monitoring, and in terms of regions that have monitoring programs but no public access to the data archive. Therefore future improvements to the database will require not only improved data harmonization, but also expanded data sharing and increased monitoring in data-sparse regions.

**Keywords:** tropospheric ozone; ground-level ozone; monitoring; database

- <sup>1</sup> Institute for Energy and Climate Research (IEK-8), Forschungszentrum Jülich, Jülich, DE
- <sup>2</sup> Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, US
- <sup>3</sup> NOAA Earth System Research Laboratory, Boulder, CO, US
- <sup>4</sup> CSIRO Climate Science Centre, Aspendale, Victoria, AU
- <sup>5</sup> Institute for Advanced Sustainability Studies, Potsdam, DE
- <sup>6</sup> National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, JP
- <sup>7</sup> NASA Goddard Space Flight Center, Greenbelt, MD, US
- <sup>8</sup> Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, US
- <sup>9</sup> Aryabhatta Research Institute of Observational Sciences, Nainital, IN
- <sup>10</sup> Environmental Department, SAGU, Santiago, CL
- <sup>11</sup> Unit II 4.2, Air Quality Assessment, German Environment Agency/Umweltbundesamt, Dessau-Roßlau, DE
- <sup>12</sup> Norwegian Institute for Air Research (NILU), Kjeller, NO
- <sup>13</sup> Air Quality Research Division, National Institute of Environmental Research, Incheon, Republic of Korea, KR
- <sup>14</sup> European Environment Agency, Copenhagen, DK
- <sup>15</sup> Japan Meteorological Agency (JMA), Tokyo, JP
- <sup>16</sup> Office of Air and Radiation (OAR) – Office of Atmospheric Programs (OAP), U.S. Environmental Protection Agency, Washington, DC, US
- <sup>17</sup> Air Quality Processes Research Section, Air Quality Research Division/Science and Technology Branch, Environment and Climate Change Canada, Toronto, CA
- <sup>18</sup> German Environment Agency/Umweltbundesamt, Langen, DE
- <sup>19</sup> Centro de Ciencias de la Atmosfera, Universidad Nacional Autónoma de México (Retired), Mexico City, MX
- <sup>20</sup> Graduate Institute of Energy Engineering, National Central University, Taoyuan city, TW
- <sup>21</sup> Dept. of Earth and Environmental Sciences, Korea University, Sungbuk-gu, Seoul, South Korea, KR
- <sup>22</sup> School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, ZA
- <sup>23</sup> Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hung Hom, Hong Kong, CN
- <sup>24</sup> Centre for Ecology and Hydrology, Environment Centre Wales, Bangor, Gwynedd, UK
- <sup>25</sup> Atmospheric Research and Instrumentation Branch, National Institute for Aerospace Technology (INTA), Ctra. Huelva – Matalascañas, Mazagón, Huelva, ES
- <sup>26</sup> CNRS/INSU, Sorbonne Université, UPMC, Université Versailles St-Quentin, LATMOS, Paris, FR
- <sup>27</sup> Ricerca sul Sistema Energetico – RSE S.p.A., 20134 Milano, IT
- <sup>28</sup> Institute of Physics, University of Sao Paulo. Rua do Matao 1371, CEP 05508-090, Sao Paulo, S.P., BR
- <sup>29</sup> National Meteorological Service, Buenos Aires Main Observatory, Buenos Aires, AR
- <sup>30</sup> Institute of Meteorology and Water Management – National Research Institute IMWM-NRI Gdynia Maritime Branch, Centre for Oceanography and Monitoring of Hydrosphere and Atmosphere, Warszawa, PL
- <sup>31</sup> National Research Council, Institute of Atmospheric Sciences and Climate, Bologna, IT
- <sup>32</sup> National Institute of Environmental Research, Incheon, Republic of Korea, KR
- <sup>33</sup> Université Clermont Auvergne, CNRS, LaMP, F-63000 Clermont-Ferrand, FR
- <sup>34</sup> Izana Atmospheric Research Center, State Meteorological Agency of Spain (AEMET), Santa Cruz de Tenerife, ES
- <sup>35</sup> Servicio Meteorológico Nacional, Ushuaia, Tierra del Fuego, AR
- <sup>36</sup> Institute of Environmental Protection – National Research Institute, Warsaw, PL
- <sup>37</sup> Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, Nanjing, CN
- <sup>38</sup> Umweltbundesamt/Federal Environment Agency Austria, Vienna, AT
- <sup>39</sup> Latvian Environment Geology and Meteorology Centre, Riga, LV
- <sup>40</sup> National Atmospheric Research Laboratory, Gadanki, IN
- <sup>41</sup> Laboratoire d'Aérodynamique, University of Toulouse, CNRS, UPS, Toulouse, FR
- <sup>42</sup> Meteorological Observatory Hohenpeissenberg, German Meteorological Service (DWD), Hohenpeissenberg, DE
- <sup>43</sup> German Meteorological Service (DWD), Research Center Human Biometeorology, Freiburg, DE
- <sup>44</sup> Atmospheric Research and Instrumentation Branch, National Institute for Aerospace Technology (INTA), Madrid, ES
- <sup>45</sup> LSCE, Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, IPSL, Gif-sur-Yvette, FR
- <sup>46</sup> School of Public Health, University of Maryland, College Park, MD, US
- <sup>47</sup> Institute of Arctic and Alpine Research (INSTAAR), University of Colorado, CO, US
- <sup>48</sup> Regional Delegation of Azores, Portuguese Institute for Sea and Atmosphere (IPMA), Ponta Delgada, Azores, PT
- <sup>49</sup> Department of Environmental Science, Aarhus University, 4000 Roskilde, DK
- <sup>50</sup> School of Science, Technology, Engineering and Math, University of Washington-Bothell, Bothell, WA, US
- <sup>51</sup> National Institute of Aeronautics and Space, Jl., Bandung 40173, ID
- <sup>52</sup> Finnish Meteorological Institute, Helsinki, FI
- <sup>53</sup> National Institute of Meteorological Sciences, Korea Meteorological Administration, Jeju, Republic of Korea, KR
- <sup>54</sup> Air Quality and Climate Change Division, Ministry of Environmental Protection, Jerusalem, IL
- <sup>55</sup> Atmospheric Science Program, Michigan Technological University, Houghton, MI, US
- <sup>56</sup> Malaysian Meteorological Department, Jalan Sultan, Petaling Jaya, MY
- <sup>57</sup> Slovenian Environment Agency, Ljubljana, SI
- <sup>58</sup> Administrația Națională de Meteorologie, Sos București-Ploiești 97, RO

## 1. Introduction

Ozone in the troposphere is relevant to human health and the environment in several respects (Cooper et al., 2014; Monks et al., 2015; Schultz et al., 2015). High ground-level ozone concentrations impact the human respiratory system and impair the growth of vegetation. Furthermore, ozone is a greenhouse gas and plays a key role in photo-oxidation processes in the troposphere. Ozone is a secondary air pollutant, i.e. it is not emitted directly but formed in the troposphere as a result of chemical reactions of precursor gases such as nitrogen oxides, carbon monoxide, and volatile organic compounds (VOC). Ozone is lost chemically through photo-dissociation, reaction with  $\text{HO}_2$  or  $\text{NO}_2$  radicals, unsaturated VOC, or halogens. It is also lost through deposition at the surface and uptake by plants, or heterogeneous reactions involving aerosol. The global average photochemical lifetime of tropospheric ozone is between 20 and 25 days (Young et al., 2013), but generally less than 5 days in the summertime surface boundary layer. The local lifetime varies considerably depending on altitude, geographic location, season, temperature, humidity, and atmospheric composition.

Surface or ground level ozone is a term used to describe the ozone mole fraction in ambient air that humans and plants experience. It is typically measured by sampling air between 2 m and 10 m above the surface. Historical observations of surface ozone mixing ratios (mole fractions) range from zero to over  $400 \text{ nmol mol}^{-1}$  (Bartel and Temple, 1952; Riveros et al., 1998; Lacasaña-Navarro et al., 1999). As documented in this article, current measurements rarely exceed  $200 \text{ nmol mol}^{-1}$  (see **Table 5** in section 5). “Zero ozone” (i.e. ozone at sub-nmol fractions) is often found in urban environments with high levels

of nitrogen oxides at night, when excess amounts of NO emitted from combustion sources react with ozone to form  $\text{NO}_2$  (for example Wang et al., 2012). At rural sites, low to very low mole fractions of ozone can be found at night due to ozone destruction at the underlying soil and plant surface (Galbally, 1968). Very low ozone mixing ratios are also seen during springtime in the Arctic troposphere where ozone gets destroyed from reactions with halogens (e.g. Tarasick and Bottenheim, 2002; Helmig et al., 2007; Simpson et al., 2007). Marine sites in clean tropical environments frequently report ozone mole fractions in the  $10\text{--}20 \text{ nmol mol}^{-1}$  range (e.g. Oltmans et al., 2006, 2012), while rural continental sites in the mid-latitudes show typical average mole fractions between 30 and  $80 \text{ nmol mol}^{-1}$  in the Northern Hemisphere, and 15 to  $25 \text{ nmol mol}^{-1}$  in the Southern Hemisphere (Galbally et al., 1986; Oltmans et al., 2012). The highest mole fractions are found in or downwind of major conurbations (e.g. Seinfeld et al., 1991). Mountain sites also generally exhibit higher mole fractions, in particular during the influence of stratospheric intrusions (e.g. Cristofanelli et al., 2006).

In spite of many years of research and substantial monitoring of surface ozone on the regional and global scales, scientists have been unable to answer the most basic questions: Which regions of the world have the greatest human and plant exposure to ozone pollution? Is ozone continuing to decline 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 ozone metrics necessary for quantifying ozone's impact on climate, human health, and crop/ecosystem productivity?

<sup>59</sup> National Centre for Atmospheric Science (NCAS), Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK

<sup>60</sup> Auckland Council, Auckland, NZ

<sup>61</sup> German Environment Agency, GAW Global Observatory Zugspitze/Hohenpeissenberg, Zugspitze, DE

<sup>62</sup> School of Environmental Sciences, Jawaharlal Nehru University, New Delhi – 110067, IN

<sup>63</sup> Atmospheric Sciences Research Center, State University of New York at Albany, Albany, NY, US

<sup>64</sup> Office of Environment and Heritage, Sydney, New South Wales, AU

<sup>65</sup> Kislovodsk High Mountain Scientific Station of A. M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, RU

<sup>66</sup> School of Chemistry, University of Bristol, Bristol, GB

<sup>67</sup> Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research (IISER) Mohali, Manauli PO, Punjab, IN

<sup>68</sup> A.M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, RU

<sup>69</sup> National University of Ireland Galway, Galway, GB

<sup>70</sup> National Institute of Public Health and the Environment, Bilthoven, NL

<sup>71</sup> Environmental and Climate Sciences Dept., Brookhaven National Laboratory, Upton, NY, US

<sup>72</sup> South Australia Environment Protection Authority, Adelaide, South Australia, AU

<sup>73</sup> Swiss Federal Laboratories for Materials Science and Technology (Empa), Dübendorf, CH

<sup>74</sup> Indonesian Agency for Meteorological Climatological and Geophysics, Jakarta, ID

<sup>75</sup> Environment Protection Authority Victoria, Melbourne, Victoria, AU

<sup>76</sup> Karlsruher Institut für Technologie, Institut für Meteorologie und Klimaforschung, Atmosphärische Umweltforschung (IMK-IFU), Garmisch-Partenkirchen, DE

<sup>77</sup> Meteorological Observation Center, China Meteorological Administration, Beijing, CN

<sup>78</sup> Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, DE

<sup>79</sup> Key Laboratory for Atmospheric Chemistry, Chinese Academy of Meteorological Sciences, China Meteorological Administration, Beijing, CN

<sup>80</sup> Environment Research Institute, Shandong University, Ji'nan, Shandong, CN

<sup>81</sup> Institute of Urban Meteorology, China Meteorological Administration, Beijing, CN

<sup>82</sup> Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Jülich, DE

Corresponding author: Martin G. Schultz  
(m.schultz@fz-juelich.de)

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/TOAR](http://www.igacproject.org/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 comprehensive tropospheric ozone assessment report based on all available surface ozone observations, the peer-reviewed literature and new analyses, and (2) generate easily accessible, documented ozone exposure and dose metrics at thousands of measurement sites around the world (urban and non-urban). The assessment report is organized as a special feature of Elementa.

Assessing the global distribution of tropospheric ozone near the surface and its trends in time is scientifically challenging, because ozone is a reactive gas with variable lifetime and consequently non-stationary distribution. Furthermore, there is inadequate data coverage in many regions of the world, combined with inhomogeneous data quality and metadata information, data access, and language issues. Prior attempts to summarize the global distribution of tropospheric ozone and its trends (e.g. Cooper et al., 2014; Sofen et al., 2016a) were limited to readily accessible data from large networks maintained by the World Meteorological Organisation, and North American, and European institutions, which introduced substantial geographical bias in the analyses. In the framework of TOAR the most comprehensive database possible of global surface ozone observations has been established at Forschungszentrum Jülich in Germany. The database contains surface ozone data sets with hourly time resolution collected from all accessible data sources worldwide, including regional or national air quality monitoring networks, multi-national programmes, and individual researchers' data. These data and the associated metadata that describe measurement sites and instrumentation have been augmented with several pieces of information from global gridded data sets. As a result, the TOAR database contains the world's largest collection of surface ozone observations in homogeneous form and allows for consistent analyses across all networks and in many world regions. While the TOAR database includes measurements from several hitherto inaccessible sources, a certain analysis bias remains due to the much denser observation networks in the Northern hemisphere mid-latitudes than anywhere else.

The TOAR database constitutes the foundation of all major analyses of surface ozone distributions and trends throughout the TOAR special feature. In particular *TOAR-Health* (Fleming et al., 2017, this issue), *TOAR-Vegetation* (Mills et al., 2017, this issue), and *TOAR-Climate* (Gaudel et al., 2017, this issue) draw heavily on data and data products from the database described in this article. In spite of its value for tropospheric ozone research, we would like to emphasize, however, that the TOAR database is not a primary data archive and has no intention to replace or substitute any existing data center for environmental

observations, nor does it have any obligation to maintain updated data records or inform countries or other legal entities in legally binding form. Establishing the TOAR data archive has only been possible through cooperation with many officially endorsed data archives, and these remain the primary repositories for the vast majority of ozone observations that are now accessible through TOAR. TOAR provides a variety of ozone metrics based on hourly observations, but not the hourly observations themselves.

This article describes the TOAR database and the ozone data products (including standard statistics as well as metrics relevant for assessing health, vegetation, and climate impacts, and trend statistics) that have been generated from hourly averages of continuous surface observations. It further demonstrates new possibilities of surface ozone analyses that have become possible through linking ozone data sets with global metadata, and it highlights the necessity and problems of *a posteriori* data quality assessment. The article structure is as follows: Section 2 summarizes the data sets that have been identified and made available for the TOAR database. Section 3 describes the different methods for public access to the TOAR database and the surface ozone data products. Section 4 details the procedures that have been applied to harmonize the ozone data and metadata in the TOAR database, including the extended metadata that were added from several global, gridded data sets. Section 5 discusses ozone quality control issues. Section 6 presents the ozone metrics data sets and demonstrates the enhanced analysis potential made possible through this work. We discuss ozone changes with (station) altitude, regional ozone differences, and seasonal cycles of ozone in different latitude bands. Finally, section 7 presents conclusions including commentary on the current state of the global surface ozone observation network and recommendations for its future development. The paper is accompanied by detailed technical documentation on: the TOAR ozone metrics and metrics data products (Supplemental Material 1), and the Jülich Open Web Interface (JOIN; Supplemental Material 2). All TOAR surface ozone data products including also standard graphics and software, are available at the PANGAEA data publishing portal (<https://doi.org/10.1594/PANGAEA.876108>). For more detailed descriptions of ozone metrics, the rationale for adopting certain metrics, and for the actual analyses of the present-day surface ozone distribution and trends with respect to health, vegetation, and climate impacts, the reader is referred to the other articles of the TOAR special feature.

## 2. Available global surface ozone observations

Surface ozone measurements commenced in the 19<sup>th</sup> century out of scientific curiosity and because it was believed "that the presence of ozone maintains health, and its absence is a cause of serious maladies" (Verdi, 1874). With the discovery of the ozone layer in the stratosphere (Fowler and Strutt, 1917) surface ozone measurements continued in the first half of the 20<sup>th</sup> century as an adjunct to explore atmospheric composition. In the early 1950's, ozone was identified as the key component of photochemical smog in Los Angeles (Haagen-Smit, 1952). This



led to the monitoring of surface ozone in the US and the subsequent discovery of photochemical smog in Australia (Galbally, 1971), the UK (Atkins et al., 1972; Derwent and Stewart, 1973), and Japan (Kondo and Akimoto, 1975). In Canada, ozone monitoring started in the early 1970s to investigate the cause of “Tobacco fleck” in southern Ontario (Cole and Katz, 1966).

The first global baseline ozone measurements at remote sites were initiated in response to the International Geophysical Year in 1957. The earliest global network of Background Air Pollution Monitoring (BAPMoN) was established under the auspices of the World Meteorological Organisation (WMO) in 1969. Some of the measurements initiated under this umbrella in the 1970s continue to the present, within the framework of WMO's Global Atmosphere Watch (GAW) programme (Schultz et al., 2015). More information on historic ozone measurements and ozone measurement techniques can be found in *TOAR-Observations* (Tarasick et al., 2017, this issue).

Since the late 1970's, air quality monitoring networks with ozone measurements have been established in several countries in Europe, North America, Australia, Japan, and South Korea. Over time, many other countries also established at least some air quality monitoring sites with ozone monitors. Often, the data from these networks are now available in near-realtime (see for example <https://aqicn.org/map/world/>). However, the rapid reporting of these data precedes the required quality control, which is needed for the purposes of TOAR. Access to quality-controlled data from the networks' archives is more difficult due to either restrictive government regulations or the lack of supra-regional data archives. In some world regions, multi-national networks or databases have been installed that complement national monitoring and data provision. Examples are the East Asia (Acid Deposition) Network (EANET) or the European Environment Agency Airbase system. Unfortunately, as described in section 5, the data from these multi-national archives are not always fully consistent with the original data reported at the regional or national level. The TOAR database maintains these “duplicate” data records as individual data series and applies a merging procedure to select the most appropriate data for analysis (see section 5).

The focus of current air quality monitoring networks generally lies in urban and suburban areas. However, in the context of acid deposition monitoring, ozone monitoring stations were also established in rural areas. These stations allow for observations of regional baseline concentrations and attribution of high ozone episodes. Examples are the US National Park Service Gaseous Pollutant Monitoring Program, the US Environmental Protection Agency Clean Air Status and Trends Network (CASTNET), the Canadian Air and Precipitation Monitoring Network (CAPMoN), the European Monitoring of Environmental Pollution (EMEP) programme, and also the Acid Deposition Monitoring Network in East Asia (EANET).

Other surface ozone observations, typically of shorter duration, have been made during field campaigns or in support of ozone impact studies on forest or agricultural vegetation. Finally, a few programs have recorded

multi-year ozone measurements from mobile platforms such as railway trains (for example TROICA: Oberlander et al., 2002; Pankratova et al., 2011), or ships (EU project APICE: Velchev et al., 2011).

For establishing the TOAR database of surface ozone concentrations we focused on data from stationary platforms, with hourly time resolution, and time series that are longer than 2 years. The vast majority of measurements were made with the UV absorption technique (see Tarasick et al., 2017, this issue). Passive sampling data were not considered due to the low time resolution. These choices were made in order to allow a globally uniform calculation of ozone metrics for the analyses in *TOAR-Metrics* (Lefohn et al., 2017, this issue), *TOAR-Health*, *TOAR-Vegetation*, and *TOAR-Climate*. Furthermore, these criteria allow characterization of at least some inter-annual variability, and to assess the robustness of the derived ozone metrics at each site. In order to achieve a well-defined data set and allow for some quality assurance of the ozone data (see section 5), the database was closed for new submissions in July 2016. Only corrections to existing data sets were accepted after this date. The most recent measurements that entered the TOAR analyses are from 2015, although most of the results presented throughout TOAR do not extend beyond 2014. In Europe, many datasets included in TOAR do not extend beyond the year 2012 due to changes in the Airbase data reporting system, which coincided with the build-up of the TOAR database and prevented inclusion of more recent Airbase data before closure of data submissions.

Data availability and accessibility of existing long-term surface ozone observations varies considerably among the countries and multi-national networks. Some networks maintain comprehensive, well-managed databases and allow open access through ftp, web downloads or interoperable web services. Such data were readily retrieved and included in the TOAR database. In cases where open access is not available or language problems prevented us from accessing or interpreting data directly, we tried to negotiate access to long-term archived data, which was particularly successful for data from Japan, and South Korea. In other regions, data collection and harmonisation remains fragmented and many different data providers must be addressed individually in order to obtain data. Major efforts were undertaken especially in Australia, South Africa, and South America to collect the available data and make them accessible to TOAR.

In some countries concerns about misuse or misinterpretation of the data prevent local authorities from openly sharing information, especially in the form of time-resolved hourly concentration values. Also, large requests for hourly data can impose a substantial work load on agency staff and they may not be in a position to engage in the necessary reprocessing of historic data if this is not a direct part of their mandate. In the case of research data, not all scientists who are involved in ozone measurements are fully supportive of the TOAR open data policy for different reasons. This may have prevented them in some cases from freely sharing their data with the TOAR database curators. We would like to note

in this context that the TOAR database contains special provisions to restrict open access to the ozone metrics at a limited number of sites if this is requested by the data provider. In such cases, only the aggregated metrics data products (see section 6) are freely available. However, it is important that these metrics are derived consistently with the metrics at all other stations in the database, so only hourly data are entered into the database.

A first attempt about 10 years ago by the first author of this article to collect and harmonize global surface ozone observations resulted in a set of about 400 data files which contained publically accessible data from four networks, namely CASTNET, EANET, EMEP, and GAW. These data were then used in the evaluation of global chemistry models, for example in the context of multi-model experiments on the hemispheric transport of air pollution (e.g. Fiore et al., 2009; Rasmussen et al., 2012). In 2015, the results of a similar ozone data collection effort were published by Sofen et al. (2016b), including data from about 6,600 sites and 8 different networks with open data archives. The authors noted the existence of other, not readily accessible ozone data, but did not see themselves in a position to acquire and process such data. The primary intention of Sofen et al. (2016b) was again to provide surface observations for the evaluation of global chemistry models, and therefore their main products are gridded data sets containing spatial averages of monthly mean or other statistically aggregated ozone concentrations. The present work benefitted from the cooperation with E. Sofen on the identification of data format and data quality issues (see discussion in sections 4 and 5).

Going beyond previous ozone data collection efforts, the TOAR initiative has worked closely with data providers from around the world in order to increase coverage beyond the regions where data are easily available from open data archives. Through the creation of regional working groups and the efforts of the many co-authors on this article, it has been possible to integrate the hourly ozone data from 9,690 stations, many of which had never been available for internationally coordinated research. Some of these data, for example from Australia and Japan, date back to the early 1980s and thus fill important gaps in our knowledge about global surface ozone during that period.

**Table 1** lists the ozone monitoring networks and data sources of the TOAR database, and **Table 2** provides a summary of the database holdings grouped by region. Regions are labelled as defined by the Task Force on Hemispheric Transport of Air Pollution (TFHTAP; Koffi et al., 2016). Large gaps remain in the global coverage of surface ozone measurements. Very few data exist in Africa, Central America, Central Asia, South Asia, and the Middle East. Where data are available from these regions, the time series are often short and sometimes appear to have some data quality issues.

**Figure 1** shows the temporal evolution of the number of ozone data records in the database by region. In general the data coverage has increased over time in all world regions. The dropoff in the most recent years is mostly due

to delays in the final processing and validation of current data. In North America, most data originate from AQS and NAPS (**Table 1**). The earliest data from AQS that are included in the TOAR database are from 1980, the earliest records from NAPS are from 1974. As we did not have NAPS data after 2013, the coverage in North America drops slightly during the final two years. In Europe, the majority of data are from Airbase with the earliest records from Great Britain dating back to 1973. Since Airbase changed their reporting system and data format for data after 2012, we could not include more recent data in the TOAR database. In East Asia, the earliest data are from Japan, dating back to 1976. Data from South Korea have been made available to us beginning in 2000. The TOAR database also includes data from 26 Chinese stations including 15 in Hong Kong, some of which date back to 1990. Unfortunately, we have not been able to obtain ozone data from the vast Chinese air quality network which commenced in 2012. Similarly, data from only 7 Taiwanese stations were included while at least 50 stations currently report ozone concentrations (AQICN, 2017).

The earliest data from Australia and New Zealand date back to 1978, while the majority of measurements in this region commenced around 1992. In Mexico and the Caribbean the network density has steadily increased since the mid-1980s. There are no other data from Central America. Data coverage decreases after 2010 in the TOAR database due to the fact that we obtained these data from individual researchers and not from official agencies and their analyses did not always include the most recent years. The earliest officially available data from Argentina and Chile are from 1994 and 1997, respectively, while ozone data from Brazil are available since 1998. Data classified as oceanic in the database are mostly from coastal sites and they are labelled OCN only because of inaccuracies in the global gridded map that was used to assign the TFHTAP region code to each station. True oceanic sites are American Samoa (GAW), Sable Island (NAPS, OTHER), Ieodo Ocean Research Station (OTHER), Ogasawara (EANET), and Minamitorishima (GAW). We note that several other island sites are not included in the OCN region. For example, Amsterdam Island (OTHER) belongs to region SAF, and Bermuda (GAW) belongs to Middle and Central America in spite of their remote locations (this has been coded in the TFHTAP gridded file). In future versions of the database, a better designation of island sites would be desirable.

The earliest data from Africa and the Middle East are from Cape Point, South Africa (GAW), which began in 1983. Assekrem, Algeria (GAW), and Amsterdam Island (OTHER) commenced in 1997 and 1995, respectively. Beginning in 2000, data from up to 20 stations are also available from Tehran, Iran. Unfortunately, these seem to have some data quality issues. From Southern Asia, 7 data sets could be obtained from India, and 1 data set from Nepal. From South East Asia, 5 data sets are from Indonesia, 4 from Thailand, 3 from Malaysia, and 1 from Vietnam. Central Asia has one data set from the station Issyk-Kul in Kyrgyzstan (GAW), and the RBU region

**Table 1:** Ozone monitoring networks as defined in the TOAR database<sup>a</sup>. DOI: <https://doi.org/10.1525/elementa.244.t1>

Network name	Short name	Regional coverage	Data center or data provider	Number of stations
Airbase	Airbase	Europe (incl. over-seas locations)	European Environment Agency <a href="https://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8">https://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8</a>	3505
University of New Hampshire Air Quality and Climate Program	Airmap <sup>b</sup>	NE US	University of New Hampshire <a href="http://www.eos.unh.edu/observatories/data.shtml">http://www.eos.unh.edu/observatories/data.shtml</a>	6
US Air Quality System	AQS	US	United States Environmental Protection Agency <a href="https://ofmext.epa.gov/AQDMRS/aqdmrs.html">https://ofmext.epa.gov/AQDMRS/aqdmrs.html</a>	2963
Australia Air Quality Network	AUSAQN <sup>b</sup>	Australia	Ian Galbally; Rob Gillett, and Suzie Molloy, CSIRO, Australia	56
The Canadian Air and Precipitation Monitoring Network	CAPMoN	Canada	Environment and Climate Change Canada, 2016. Canadian Air and Precipitation Monitoring Network (CAPMoN), Toronto, Ontario, Canada. Data file: CAPMoN_O3_ALT_to2015.csv, generated 2016-11-09.	19
US Clean Air Status and Trends Network	CASTNET	US	United States Environmental Protection Agency <a href="https://www.epa.gov/castnet">https://www.epa.gov/castnet</a>	117
Acid Deposition Monitoring Network in East Asia	EANET	East Asia	Asia Center for Air Pollution Research <a href="http://www.eanet.asia/">http://www.eanet.asia/</a>	16
European Monitoring and Evaluation Programme	EMEP	Europe	<a href="http://www.emep.int/">http://www.emep.int/</a>	196
Global Atmosphere Watch	GAW	global	World Data Center for Greenhouse Gases <a href="http://ds.data.jma.go.jp/wdcdgg">http://ds.data.jma.go.jp/wdcdgg</a>	123
Israel Air Quality Network	ISRAQN <sup>b</sup>	Israel	Levana Kordova, Israeli Ministry of Environmental Protection	12
Korea Air Quality Network	KRAQN <sup>b</sup>	South Korea	Meehye Lee, Korea University, and National Institute for Environmental Research	312
Canada National Air Pollution Surveillance	NAPS	Canada	Environment Canada <a href="http://www.ec.gc.ca/rnspa-naps/">http://www.ec.gc.ca/rnspa-naps/</a>	373
National Institute for Environmental Studies	NIES <sup>b</sup>	Japan	Hiroshi Tanimoto, National Institute for Environmental Studies, Japan	1260
Republic of South Africa Air Quality Network	RSA <sup>b</sup>	South Africa	Raeesa Moolla, University of the Witwatersrand, South Africa	22
Umweltbundesamt	UBA	Germany	S. Feigenspan (UBA) <a href="http://www.umweltbundesamt.de/en/data/current-concentrations-of-air-pollutants-in-germany">http://www.umweltbundesamt.de/en/data/current-concentrations-of-air-pollutants-in-germany</a>	613
	other <sup>b</sup>		various individuals and data from smaller national or regional networks	130

<sup>a</sup>Note that not all stations of a given network have measured ozone continuously. More information on the number of available data records during specific periods can be obtained from Figure 1. For detailed information on data availability from individual stations the TOAR database should be accessed directly via the JOIN web interface as explained in Supplemental Material 2).

<sup>b</sup>No open access, or web site not available in English; data were therefore obtained through direct personal interaction with network data managers, civil servants, or scientists.

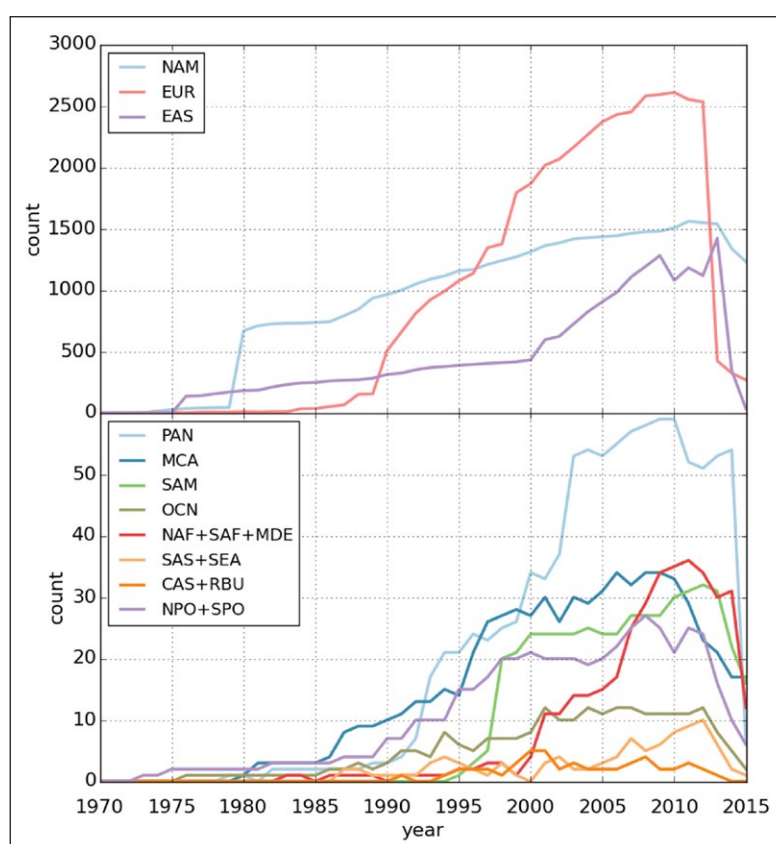
contains data from 8 stations in Russia, 1 in Armenia, and 1 in Poland (misclassified because it is located on the border with Belarus). There are a total of 26 ozone data sets from the Arctic (north of 66°N), with 8 from Finland, 5 from Norway, 4 from the USA, 4 from Canada, and the rest from Denmark, Greenland, Sweden, and Russia. The US station Barrow has the oldest data going back to 1973.

In the Antarctic region (south of 66°S), a total of 11 stations contribute data (**Table 2**), with the oldest being the Amundsen-Scott South Pole station, going back to 1975. Some of these data from the early 1970s predate the establishment of the modern UV standard, and will have used the KI method, or a chemiluminescent method calibrated to the KI standard (Tarasick et al., this issue).

**Table 2:** Number of ozone data sets in the TOAR database by world region<sup>a</sup>. DOI: <https://doi.org/10.1525/elementa.244.t2>

TFHTAP region name	Short name	Number of data sets in TOAR database
Non-Arctic, non-Antarctic Ocean	OCN	24
US and Canada (up to 66°N)	NAM	3456
Western and Eastern Europe and Turkey (up to 66°N)	EUR	4329
South Asia: India, Nepal, Pakistan, Afghanistan, Bangladesh, Sri Lanka	SAS	8
East Asia: China, Korea, Japan	EAS	1617
South East Asia	SEA	13
Pacific, Australia and New Zealand	PAN	63
Northern Africa including Sahara and Sahel	NAF	2
Sub Saharan/Sub Sahel Africa	SAF	35
Middle East: Iran, Iraq, Saudi Arabia, Oman, etc.	MDE	33
Mexico, Central America, Caribbean, Guyanas, Venezuela, Columbia	MCA	46
South America	SAM	41
Russia, Belarus, Ukraine	RBU	9
Central Asia	CAS	1
Arctic Circle (North of 66°N) and Greenland	NPO	26
Antarctic	SPO	11

<sup>a</sup>The region names and boundaries are taken from the Task Force on Hemispheric Transport of Air Pollution (TFHTAP2; Koffi et al., 2016).



**Figure 1:** Temporal evolution of the number of ozone data records in the TOAR database from various world regions. For this figure, a data record is defined as a station which has at least 3600 hours (~5 months) of valid ozone data in a given year. Regions are labelled according to the TFHTAP definitions (see Table 2). DOI: <https://doi.org/10.1525/elementa.244.f1>



### 3. Access to TOAR data

All TOAR data except for the original hourly time series are freely available for scientific and policy use. However, we want to stress that, in spite of close collaboration between the TOAR database curators and most data providers, the TOAR data products do not constitute official data and are therefore not to be used in analyses with legal implications such as exceedance monitoring.<sup>1</sup>

The actual TOAR database with the hourly observations continues to develop (for example, we recently included ozone precursor and meteorological data for European stations), while the pre-compiled data products, which have been used in other parts of the assessment, represent a frozen snapshot in order to guarantee reproducibility of results.

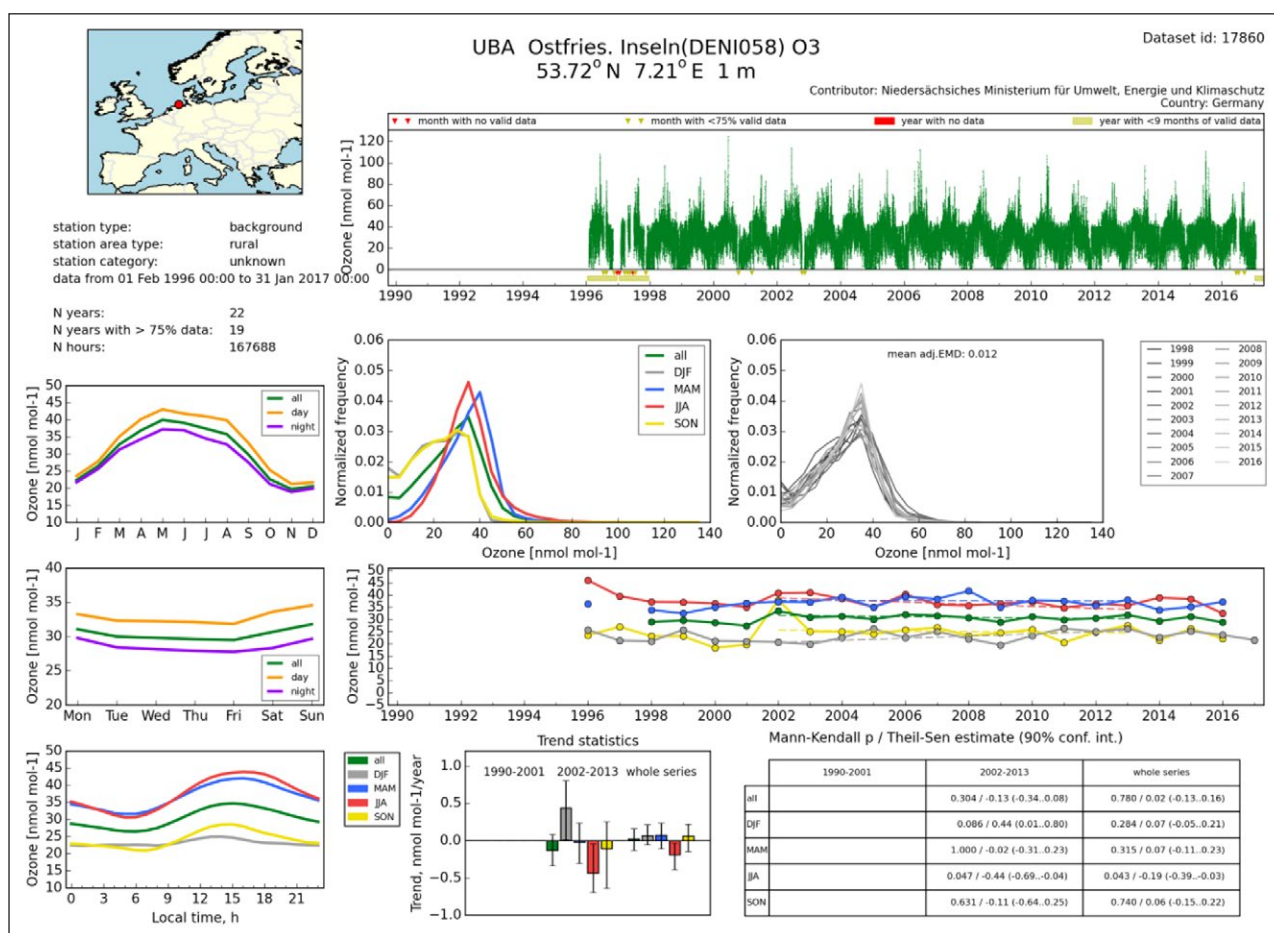
For access to the pre-compiled TOAR data products including present-day monthly, seasonal, summertime, and annual data, trend datasets, and gridded datasets, we refer the reader the data publication on PANGAEA (<https://doi.org/10.1594/PANGAEA.876108>). The metrics files are provided as simple csv files. Gridded data products intended for model evaluation purposes are provided as netcdf files. More detailed descriptions of the file formats and variables are provided in the Supplemental Material 1.

The PANGAEA data portal also contains an extensive collection of plots similar to those presented in *TOAR-Health*, *TOAR-Vegetation*, and *TOAR-Climate*.

The live database can be accessed via the interactive web interface of the Jülich Open Web Services Interface (JOIN; <https://join.fz-juelich.de>), and through the Representational State Transfer (REST) services which are also provided by JOIN. Details are provided in the JOIN user guide (Supplemental Material 2).

The JOIN web interface allows for easy access to meta-data and data from individual measurement sites. It includes several filtering options to select stations based on various metadata criteria (faceted search). Hourly, daily, monthly, seasonal, summertime, and annual data can easily be visualized as time series and, with the exception of hourly data, are also available for download as text files. The user can also generate comprehensive data summary plots, which contain a time series including data capture information, average annual, weekly, and diurnal cycles with distinction between night and day or season, respectively, frequency distributions, and trend information (see example in **Figure 2** below).

For instructions how to generate such plots, see Supplement 2. Via the REST interface of JOIN, most



**Figure 2:** Example of a standardized data summary plot provided to all individual data submitters and available through the web interface by JOIN (see text). The data summary provides information on the station location, a display of the hourly time series including information on periods with incomplete coverage, average seasonal, weekly, and diurnal cycles, frequency distributions, and a preliminary trend analysis. DOI: <https://doi.org/10.1525/elementa.244.f2>

information from the TOAR database is also available in an interoperable way. This allows for software access to TOAR data without the requirement to log into the JOIN web interface.

## 4. Procedures for collecting and harmonizing ozone data and metadata

### 4.1 Data collection

The vast majority of data in the TOAR database were obtained from a few well-managed networks which maintain their own databases and apply their own quality control measures to varying degrees (see Sofen et al., 2016b for a discussion on typical data and metadata issues). In each case we established contact with the database managers and individual data providers and asked for the best available data collection and metadata information. In many cases, this included extensive discussions of metadata and data quality issues with the data providers and archive managers. As noted above, in some regions the TOAR regional working groups and other individuals invested considerable effort to collect and harmonize ozone data sets and submit them to the database. A simple template in ASCII format was designed and shared for these submissions. We generated standardized data summary plots (**Figure 2**) from each data set that was inserted into the database and we shared these with the data providers to ensure that the processing was successful and also to prompt another critical look at the data.

### 4.2 Reported metadata and their quality control

The TOAR data submission template requested detailed metadata information, including, but not limited to the station location and altitude, station type ('traffic', 'industry', or 'background'), station type of area ('urban', 'suburban', 'rural', or 'remote'), and details about the measurement method, the data set PI, and the contributing organisation. Not all of this information is always available, and as of yet there is no standardized vocabulary applied to many of these metadata fields, limiting the ability to search for specific types of data in the database. For example, there are more than 50 different terms used to describe the most common ultraviolet absorption measurement method; in this case the information has been standardized.

We performed extensive checking of station metadata, primarily with respect to the station location. In several hundred cases we tried to manually verify the station location through Google maps (Google, 2017), and often found either incorrect or imprecise coordinates which were then corrected. Often, these checks were prompted by detecting different coordinate values for identical stations in different data repositories. In extreme cases, the station locations differed by more than 30 km from each other (for example for Somerton, GB0044R between EMEP and Airbase). In most cases the differences were on the order of a few hundred meters or less. One means of checking station coordinates was to compare the reported station altitudes with the topographic elevations from Google maps. This not only revealed many cases with highly inaccurate station elevations (for example, many of

the older AQS sites apparently had altitudes given in feet instead of meters, or they reported altitude above ground), but could also be used to correct the reported location of the site, in particular in mountainous terrain. For example, the reported station altitude of the Indian station Mt. Abu was 1680 m asl, while Google maps returned an altitude of only 1180 m at the reported coordinate values of 24.6°N, 72.7°E. The station building could be visually identified on the Google maps satellite image at the more precise location 24.653056°N, 72.779167°E. Using these coordinates to retrieve a new Google maps altitude yielded 1663 m asl, which is very close to the reported station altitude. Note, however, that differences in station altitude can also occur for sites where sampling occurs on tall towers. Some of these were identified and documented in the station\_comments attribute, but we likely missed several other towers, because information on inlet heights is generally not available.

The coordinate corrections were also sent to the data providers for verification. More recently, we have begun documenting all coordinate corrections by introducing additional database fields. For example, three different station altitude values are maintained (station\_reported\_alt, station\_google\_alt, and station\_etopo\_alt, see **Table 3** below), and we use a station\_alt\_flag to document which of these is regarded as the most trustworthy piece of information. This value is then returned as station\_alt. Similarly, we have begun to document the confidence we have in the station coordinates that are saved in the database (for details see the description of the database layout in Supplemental Material 2).

Despite these efforts many issues remain with respect to inaccuracies in the station information, and this may limit the applicability of these data for impact assessments. For example, while modern web services and GIS applications would make it feasible to relate air pollutant concentrations to the distance of pollution sources such as roads or industrial plantations, such analyses will produce erroneous results if the station coordinates are not maintained with high accuracy. Furthermore, many impact studies also require information about the sampling height of the measurement. This information is, unfortunately, currently not included in the TOAR database, because there are too few data sets that report this site characteristic.

### 4.3 Gridded metadata

In order to improve the characterization of stations and their environment, and to allow for globally consistent data aggregation by site-specific criteria, we obtained several high-resolution global gridded data products. As described above these additional metadata are also used to quality control the metadata information that is provided with the original data. **Table 3** provides an overview of the station metadata added to the data submissions through extraction from global gridded data products. Additional information can be found in the description of the database layout in Supplemental Material 2.

The global gridded products are provided in different resolutions, are valid for different years, and may contain

**Table 3:** Station metadata in the TOAR database extracted from gridded data files. DOI: <https://doi.org/10.1525/elementa.244.t3>

Variable name	Description	Original resolution	Data source
station_htap_region	region code of the Task Force Hemispheric Transport of Air Pollution (tier 1 regions). In order to avoid misclassification of coastal sites as “ocean”, the gridded data is searched for the maximum region code within a radius of 15 km around the station if the original assignment would be “ocean”.	0.1°	Koffi et al., 2016
station_etopo_alt	station altitude according to the ETOPO1 digital elevation model in m. An adjustment for coastal sites is included: if station_etopo_alt is <−5 m, the maximum altitude of the neighboring grid boxes will be used instead. Only if all neighboring grid boxes have altitudes <−5 m, the original value will be retained.	1'	Amante and Eakins, 2009: <a href="https://www.ngdc.noaa.gov/mgg/global/global.html">https://www.ngdc.noaa.gov/mgg/global/global.html</a> , accessed 05 Dec 2016.
station_etopo_relative_alt	altitude difference between station_etopo_alt and the minimum ETOP1 altitude in a radius of 5 km around the station location in m.		
station_population_density	estimated population density of the year 2010 in km <sup>−2</sup> .	2.5'	CIESIN, 2005: <a href="http://sedac.ciesin.columbia.edu/data/collection/gpw-v3">http://sedac.ciesin.columbia.edu/data/collection/gpw-v3</a> , accessed 05 Dec 2016.
station_max_population_density_5km	the maximum population density in a 5 km radius around the station location in km <sup>−2</sup> .		
station_max_population_density_25km	the maximum population density in a 25 km radius around the station location in km <sup>−2</sup> .		
station_nightlight_1km	stable nighttime lights of the year 2013 extracted from the NOAA DMSP product. The values in this data set represent a brightness index ranging from 0 to 63. Note that the sensor saturates at 63.	0.925 km	<a href="http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html">http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html</a> , accessed 05 Dec 2016.
station_nightlight_5km	the average nighttime_light value in a 5 km radius around the station location.		
station_max_nightlight_25km	the maximum nighttime_light value in a 25 km radius around the station location.		
station_climatic_zone	IPCC, 2007 classification scheme for default climate regions. If the initial climatic zone is determined as “ocean”, the maximum climatic zone value in a region of 5 km around the station is used instead. This scheme contains 12 climatic zones plus an “unclassified” category.	5'	<a href="http://eusoils.jrc.ec.europa.eu/projects/RenewableEnergy">http://eusoils.jrc.ec.europa.eu/projects/RenewableEnergy</a> , accessed 05 Dec 2016.
station_dominant_landcover	Yearly land cover type L3 from the MODIS MD12C1 collection. The year 2012 and the IGBP classification scheme (17 classes) were used.	0.05°	<a href="https://lpdaac.usgs.gov/data/set_discovery/modis/modis_products_table/mcd12c1">https://lpdaac.usgs.gov/data/set_discovery/modis/modis_products_table/mcd12c1</a> , accessed 05 Dec 2016.
station_landcover_description	a text field describing the major landcover types in a 25 km radius around the station location.		
station_rice_production	annual rice production of the year 2000 according to the Global Agro-Ecological Zones data, version 3. Units are thousand tons per grid cell.	5'	<a href="http://gaez.fao.org/Main.html">http://gaez.fao.org/Main.html</a> , accessed 05 Dec 2016.
station_wheat_production	annual wheat production of the year 2000 according to the Global Agro-Ecological Zones data, version 3. Units are thousand tons per grid cell.	5'	<a href="http://gaez.fao.org/Main.html">http://gaez.fao.org/Main.html</a> , accessed 05 Dec 2016.

(contd.)

Variable name	Description	Original resolution	Data source
station_nox_emissions	annual NO <sub>x</sub> emissions of the year 2010 from EDGAR HTAP inventory V2 in grams of NO <sub>2</sub> m <sup>-2</sup> yr <sup>-1</sup> .	0.1°	Janssens-Maenhout et al., 2015 ( <a href="http://edgar.jrc.ec.europa.eu/htap_v2/index.php?SECURE=123">http://edgar.jrc.ec.europa.eu/htap_v2/index.php?SECURE=123</a> ), accessed 05 Dec 2016.
station_omi_no2_column	5-year average (2011–2015) tropospheric NO <sub>2</sub> column value from the Ozone Monitoring Instrument (OMI) instrument on NASA AURA. Units are 10 <sup>15</sup> molecules cm <sup>-2</sup> .	0.1°	Krotkov et al., 2016

errors themselves. It is therefore not advisable to rely on one individual piece of metadata for selecting or aggregating surface ozone data. For example, we tried to correlate several “pollution indicators” (nighttime lights, population density, NO<sub>x</sub> emissions, and NO<sub>2</sub> tropospheric column densities), and found large scatter among these variables even though they are qualitatively consistent (**Figure 3**). Correlation coefficients  $r$  range from 0.46 to 0.83. The lowest correlation is found between nighttime lights and  $\log(\text{NO}_2 \text{ columns})$ , while the highest correlation is calculated for the pair of nighttime lights and  $\log(\text{population density})$ .

Nevertheless, taken together these indicators allow us to make some clear distinctions between more and less pollution-impacted environments, and with their help it has been possible for the first time to develop a globally applicable, robust station classification scheme. **Table 4** lists the criteria that were applied to mark all stations in the TOAR database as “urban”, “rural, low elevation”, “rural, high elevation”, or “unclassified”. The main intention of this classification is to identify sites which should have a clear urban, or clear rural signature. Therefore, the criteria have been chosen so that only about one half of all stations are classified as either urban or rural. All stations which do not fall in one of these categories are labelled “unclassified”. The thresholds listed in **Table 4** were determined experimentally, starting from various definitions of “urban” and “rural” obtained from web searches and varying the thresholds until we achieved the most convincing results. We verified the classification scheme manually by checking that about 100 sites on all continents, which we know to be either urban or rural, are actually classified as such, and we checked another 100 sites or so by inspecting the station location on Google maps in high resolution. This classification scheme is used extensively in *TOAR-Health* and *TOAR-Vegetation*. Examples are also shown in section 6 of this paper.

## 5. Data quality control

The assembly of so many long-term ozone measurement records invariably raises questions about the consistency and comparability of these data. Different rules and procedures for the quality assurance of the measurements and of the data management are in place in the various networks and at individual sites. Furthermore, instruments, operators, and calibration procedures may change over time which can lead to more or less visible changes in the data record (cf. Zurbenko et al., 1996).

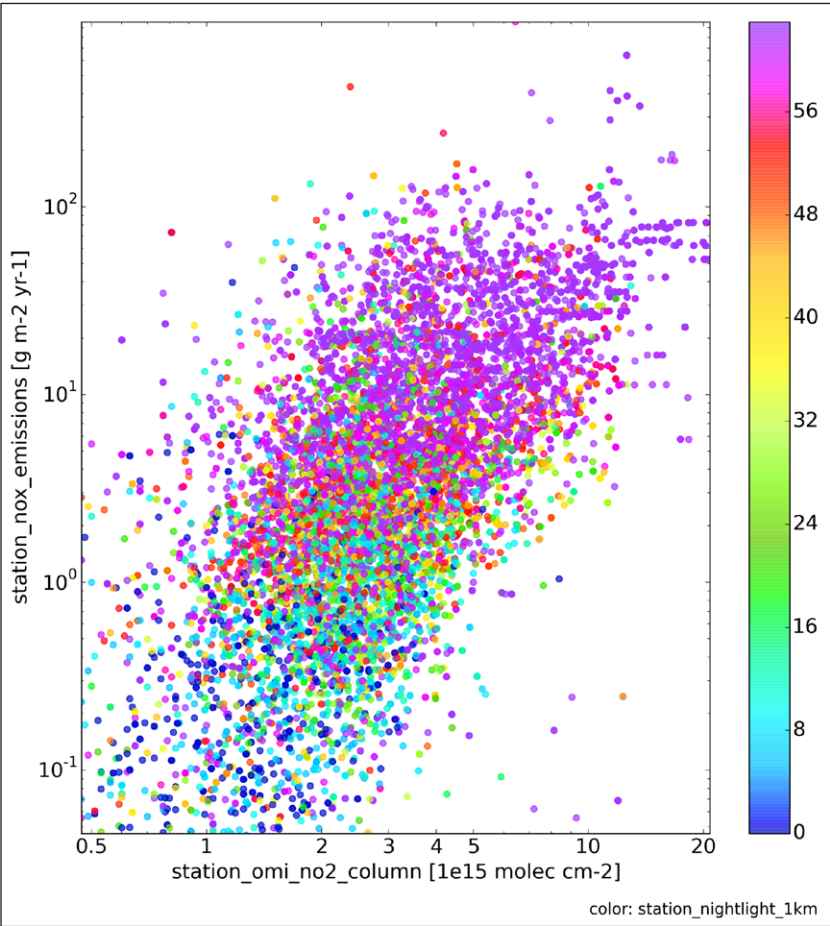
The principal sources of systematic data errors are:

- Measurement errors, i.e. errors in the set-up or operation of the instrument, calibration errors, inadequate instrument operating conditions (power failures, lack of air conditioning, etc.), or inconsistencies arising from instrument changes or maintenance;
- Sampling errors, i.e. an ill-positioned measurement site, improper set-up, or inadequate material of the inlet line, etc.;
- Data processing errors, i.e. false flagging of suspicious or erroneous measurements, neglect of documenting special conditions, such as local pollution sources, unit conversion errors, errors when applying calibration results, arithmetic errors when averaging higher frequency data to the standard hourly resolution (including the neglect of data capture criteria);
- Data submission errors, i.e. formatting errors, misinterpretation of flagging values, use of incorrect units, wrong time stamps due to incorrect time zone specifications or ambiguities with respect to stamping the beginning or end of an averaging interval.

On top of these systematic errors there are of course measurement uncertainties due to the measurement principle of the instrument, potential interferences, uncertainties of calibration, and instrument noise.

The limited information on data quality that is available in current surface ozone data sets precludes a meaningful and systematic use of metadata for the identification of potential data inaccuracies. Furthermore, such metadata would not protect against data processing and data submission errors, of which we found many in the data that were made available to TOAR. We did not keep track of all individual data errors that were identified. A rough, subjective estimate is that more than 95% of the hourly ozone values do not show any obvious quality issues. Most of the questionable data concern complete time series from some fifty sites where reported ozone mole fractions are frequently interrupted and the values, diurnal and seasonal patterns simply don't match any expectation. These series are easily identified and excluded from further analysis. About 1% of the remaining data show questionable or erroneous features during some parts of the time series. This can be individual outlier values or, for example, calibration shifts which happen throughout one year or part of a year. Even if these cases are only 1%





**Figure 3:** Correlation between global metadata variables at the locations of the stations in the TOAR database Shown are EDGAR HTAP NO<sub>x</sub> emissions and OMI tropospheric column NO<sub>2</sub> densities (both at 0.1° resolution). Colors represent the nighttime light intensity (unitless) at 1 km resolution. See Table 3 for a description of the variables displayed. DOI: <https://doi.org/10.1525/elementa.244.f3>

**Table 4:** The TOAR station classification scheme based on the gridded metadata<sup>a</sup>. DOI: <https://doi.org/10.1525/elementa.244.t4>

TOAR station category	Rule	Number of sites
urban	population_density > 15000 AND nightlight_1km >= 60 AND max_nightlight_25km = 63	3610
rural, low elevation	omi_no2_column <= 8 AND nightlight_5km <= 25 AND population_density <= 3000 AND max_population_density_5km <= 30000 AND google_alt <= 1500 AND etopo_relative_alt <= 500	2939
rural, high elevation	omi_no2_column <= 8 AND nightlight_5km <= 25 AND population_density <= 3000 AND (google_alt > 1500 OR (google_alt > 800 AND etopo_relative_alt > 500))	267
unclassified	all others	5842

<sup>a</sup>The prefix “station\_” is omitted from the variable names in the rule expressions for clarity. For details of the selected variables, see Table 3. Population density is reported in km<sup>-2</sup>, nightlight is an integer index, omi\_no2\_column is given in 10<sup>15</sup> molec. cm<sup>-2</sup>, and altitudes are in m.

of the data, several hundred time series are affected, and without accounting for such data problems, the TOAR statistics and trend analyses would lead to wrong results.

The most prevalent, obvious errors are readily exposed through visible inspection of time series, or from database queries searching for negative or anomalously high concentration values. As an example, **Table 5** lists the largest mixing ratios that are found in the data between 2008 and 2015 after we had completed our data quality inspections (see below). Each entry in **Table 5** is accompanied by a comment based on a visual inspection of the respective time series around the maximum ozone occurrence.

Four types of erroneous values were most common, and we flagged such errors in hundreds of data series: (i) extreme outliers, (ii) large negative concentration values, (iii) extended periods of very low ozone with little variability, and (iv) periods with obvious unit conversion errors. In some cases of unit errors we were able to correct the data instead of flagging them. It must be noted that identification of data errors is not always straightforward. Outliers, for example, may also arise from real events like stratospheric intrusions, and unit conversion errors may be masked by drifts in instrument sensitivity or calibration. Due to the sheer amount of data we had to process and screen, not all discovered data quality issues were systematically logged. However, in many cases erroneous or questionable data were flagged, and the data providers were contacted. Clearly, for the future development of the TOAR database, it would be desirable to develop algorithms to pre-screen the data and alert the station operators about potentially flawed time series. While we have indeed begun such developments, there has not been enough time to conduct sufficient testing, to integrate such tests into a quality control system, and to

optimize the tests in order to reduce false positive alarms and potential misses to the extent possible.

A particular data quality issue concerns very small or negative concentration values. Small negative concentrations can arise in ozone measurements due to the measurement principle of UV absorption instruments. However, assuming appropriate equipment with adequate signal to noise ratios and detection limits, and carefully maintained and calibrated instruments we would not expect mixing ratios lower than  $-1$  or  $-2$  ppbv in an hourly mean value (see Galbally and Schultz, 2013). Yet, at some stations mixing ratios of  $-5$  ppbv or even lower were reported. These are likely associated with calibration errors and cannot easily be corrected, because it is unclear if the data should be shifted or truncated. Hence, such data have to be accepted as is (they are not flagged in the TOAR database), but the user should be aware that the data uncertainty will be large. It should be noted that individual data centers or data providers handle negative concentrations differently. Without detailed knowledge of their procedures, the resulting uncertainties of the data can be only poorly assessed.

Fortunately, at least in Europe, North America, Japan, and Korea, the density of stations is high enough that unusual patterns emerge during data analysis such as the mixing ratio and trend maps that were generated for *TOAR-Health*, *TOAR-Vegetation*, and *TOAR-Climate*. Indeed, in a few dozen cases such anomalies were seen on preliminary versions of the TOAR plots and they prompted us to perform a closer inspection of the respective time series, which in turn resulted in some unit corrections and the flagging of several outlier values.

For some regions with a limited number of sites but suspicious features in individual time series, we plotted

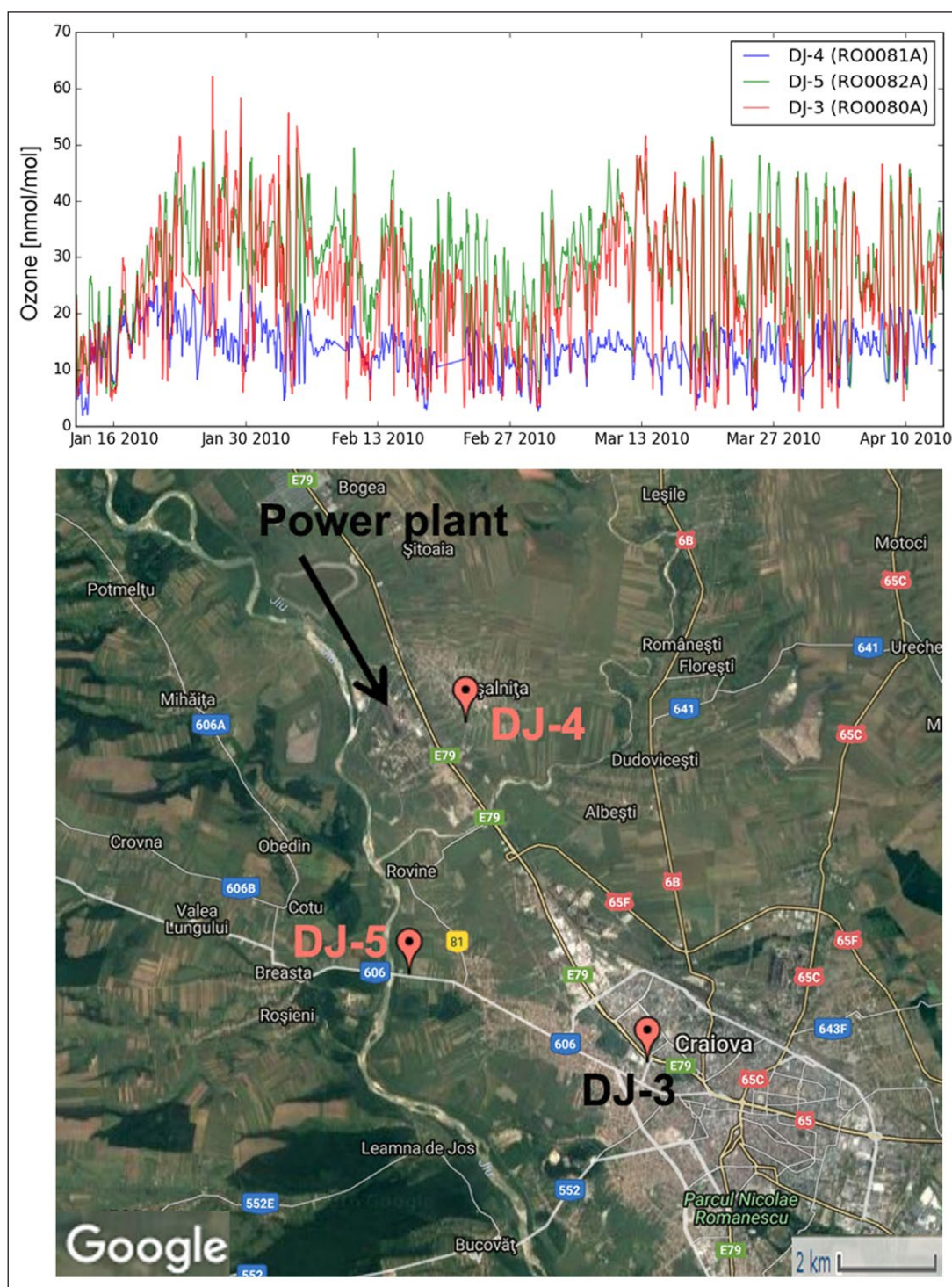
**Table 5:** Maximum ozone mixing ratios in the TOAR database of the years 2008 to 2015 after screening of time series was completed<sup>a</sup>. DOI: <https://doi.org/10.1525/elementa.244.t5>

Year	Station	max. ozone value	comment from visual inspection
2008	Cowpens, USA (45-021-0002)	222	20 Feb: very suspicious; peak looks reasonable, but after a missing value period – in original data flagged with “RL” = request exclusion
2009	unnamed (near Lincoln), Nebraska, USA (31-109-0016)	259	17 Oct: very likely an invalid outlier (another one near beginning of year) – no flags in AQS original data
2010	Saitamashikomaba, Japan (jp11204050)	223	24 Jul: looks like a real ozone episode
2011	Hendrina, South Africa (RSA019)	237	23 Aug: looks like a real peak
2012	Ichiharaiwasakinishi, Japan (jp12219190)	213	26 Jul: suspicious peak, although not completely unreasonable
2013	Horicon Wildlife Area, Wisconsin, USA (55-027-0001)	232	08 May: likely an invalid outlier – no flags in original AQS data
2014	Gongdan-Dong_Gu, South Korea (KOR437151)	215	09 July: unusually high values for this site, but reasonable peak shape
2015	Fort Mountain, Georgia, USA (13-213-0003)	269	24 April: likely a real, but untypical outlier, flagged with “IM”, “Prescribed Fire”, “Informational Only” in original AQS data

<sup>a</sup>Maximum ozone values are given in nmol mol<sup>-1</sup>.

the data from neighbouring sites together in order to check their consistency and identify the most suitable time series. One example of such analysis is shown in **Figure 4** for three Romanian stations from the EEA Airbase. In this

example, ozone data from three stations which are less than 9 km apart from each other are compared. Two of the stations (DJ-3, and DJ-5) exhibit similar patterns of variability, although the ozone mole fractions differ considerably



**Figure 4:** Comparison of an arbitrarily selected subset of hourly ozone data from three adjacent Romanian stations between October 2008 and January 2009. The three sites are named DJ-3, DJ-4, and DJ-5, respectively, and they are no more than 9 km apart from each other. All three sites are associated with EDGAR HTAP  $\text{NO}_x$  emissions of  $6.22 \text{ g m}^{-2} \text{ year}^{-1}$ , and average  $\text{NO}_2$  column of  $2.3 \cdot 10^{15} \text{ cm}^{-2}$ , as they apparently all fall into the same  $0.1^\circ$  grid cell. DJ-3 is an urban traffic site with a nighttime light intensity value of 61 (the TOAR classification of this site is “unclassified”). DJ-4 is a suburban, industrial site with a nighttime lights value of 26, and DJ-5 is a suburban background site with a nighttime lights value of 16. The TOAR classification labels both DJ-4, and DJ-5 as “rural, low elevation”. The apparent inconsistencies in the ozone data from DJ-4 are likely caused by  $\text{NO}_x$  emissions from the nearby power plant (see map). DOI: <https://doi.org/10.1525/elementa.244.f4>



during at least a third of the depicted period. Ozone variations at the third site (DJ-4) initially also follow those from the other two sites but then change their behaviour. Ozone mole fractions at DJ-4 are then substantially lower than at the other two sites and only rarely correlate with them. We excluded all data from DJ-4, because this station was clearly visible as an outlier in all TOAR analyses. In retrospect, after re-investigating this case for the preparation of this manuscript, this may have been a mistake, because the observed ozone variations and lower mole fractions at DJ-4 could be real as there is a thermal power plant about 700 m away from the station. However, even so, it is questionable if such data with very strong local influence are suitable for TOAR analyses. Future analyses of this kind could greatly benefit from the inclusion of ozone precursor data, i.e.  $\text{NO}_2$  or  $\text{NO}_x$  measurements, and from more precise  $\text{NO}_x$  emission data at finer resolution.

The TOAR database contains several “duplicate” records where data from one station had been submitted to different data archives. One would expect these time series to overlap indistinguishably on top of each other. In most cases this is indeed the case. However, we did find instances with obvious discrepancies from periods missing in one time series but not the other, outliers flagged in one data set but not the other, to different concentration values due to subtle differences in the application of data coverage rules. For example, in one network hourly values were considered valid only if two 30-minute average values were valid, whereas the presence of only one valid 30-minute value sufficed in the other network. In the case of a European mountain station (Jungfrauoch, Switzerland) the comparison of data from the different archives also led to the discovery of a unit conversion error in one of the databases. Here, different reference temperature and pressure values for the conversion from ppbv to  $\mu\text{g m}^{-3}$  were assumed by the data provider and the data center, and for a high-altitude site this has a significant effect. Note that instead of removing duplicate records from the database we established a selection and merging table (see Supplemental Material 1).

In summary, this review showed that large uncertainties remain in the reported surface ozone data from many networks and data centers. In spite of substantial efforts to identify and flag erroneous and suspicious parts of the data, there may be many time series with anomalies or inconsistencies which escaped our attention. Conversely, as demonstrated by one example above, we may have erroneously flagged legitimate data in some cases, because of insufficient information. Fortunately, as evidenced by the analyses presented in *TOAR-Health*, *TOAR-Vegetation*, and *TOAR-Climate*, the impacts of remaining data errors appear to be minor, because the concentration and trend maps show largely consistent features across wide regions. Where trends vary within a region, this may be a combination of local impacts (changes of local emission sources) and data quality issues. As discussed with the example of three Romanian stations in the context of data consistency above, detailed knowledge about individual stations and instrument performance is required in order to avoid misinterpretation of ozone trends. In order to derive

meaningful results that are robust representations of regional ozone changes, automated data filtering methods must be further developed. The objective station classification of TOAR is only a beginning in this regard.

## 6. Metrics data sets

Precompiled data sets with extensive sets of statistical quantities, including metrics defined for assessing ozone impacts on health, vegetation, and climate, and ozone trend statistics, have been made available for the TOAR report and wider research use at the PANGAEA data portal (<https://doi.org/10.1594/PANGAEA.876108>). These data products are provided in the form of easily readable ASCII files (csv format), or NetCDF (<http://www.unidata.ucar.edu/software/netcdf/>) files in case of gridded products for the evaluation of numerical models. A detailed description of the available surface ozone data products can be found in Supplemental Material 1.

The primary TOAR data products which are also used extensively in the analyses of *TOAR-Health*, *TOAR-Vegetation*, and *TOAR-Climate*, are the 5-year aggregate data sets and the trend statistics. **Table 6** lists the time periods and the conditions that must be met by an ozone series to be included in the respective metrics or trend files. Note that ozone series are here defined as merged series (see section 8 in Supplemental Material 1) where appropriate.

All aggregate metrics and trend files are available for 6 different aggregation periods during the year (**Table 7**). The extractions proceed in two steps: in the first step intermediate files are generated which contain metrics data for each individual year during the selected time interval (these are stored as “yearly statistics” on the TOAR data portal), then, in a second processing step, the aggregate or trend data sets are produced, thereby reducing the information from one line per year per station to one line per station.

All aggregated metrics and trend files preserve the complete metadata information that is provided for each site in the TOAR database. This makes it possible to relate metrics and trends to site characteristics such as population density, nighttime light intensity, etc.

**Figures 5–8** demonstrate the value of the TOAR surface ozone data products and the possibility to distinguish sites by their metadata. **Figure 5** shows annual median ozone mixing ratios versus station altitude compiled from more than 5000 data sets measured between 2008 and 2015. The data show a large spread in low altitudes ranging from around zero to above 50  $\text{nmol mol}^{-1}$ . Ozone mole fractions generally increase with altitude (e.g. Staehelin et al., 1994; Chevalier et al., 2007), although the maximum values above 3000 m are only slightly larger than those in the lowest 500 m. Rural sites tend to exhibit larger ozone mole fractions than urban sites and there is no tendency of increasing values with altitude in urban environments. This indicates that ozone in the urban environment is primarily controlled by local chemical processes and less influenced by large scale advection.

**Figure 6** shows maps of monthly mean gridded daytime ozone averages in January and July between 2010 and 2014. The data represented in these maps are



**Table 6:** Time periods for TOAR aggregate metrics and trend files and corresponding data capture requirements<sup>a</sup>. DOI: <https://doi.org/10.1525/elementa.244.t6>

Time period	Data completeness criteria
<i>aggregate metrics files</i>	
2010–2014 ("present day")	at least 3 years of data (3 "ozone seasons") available with data capture >75%
2005–2009	<i>ditto</i>
2000–2004	<i>ditto</i>
1995–1999	<i>ditto</i>
2008–2015 <sup>a</sup> ("maximum coverage")	at least one year of data with data capture >75%
<i>trend files</i>	
2005–2014	at least 7 years of data and not more than 2 years missing at either end of the interval
2000–2014 <sup>b</sup>	at least 11 years of data and not more than 2 years missing at either end of the interval
1995–2014	at least 16 years of data and not more than 2 years missing at either end of the interval
1970–2014	at least 25 years of data, allowing for variable start and end dates

<sup>a</sup>Only a reduced set of statistics has been calculated for this period, because results are less robust due to the relaxed data capture criteria.

<sup>b</sup>This is the primary focus period of the analyses in *TOAR-Health*, *TOAR-Vegetation*, and *TOAR-Climate*.

**Table 7:** Aggregation periods for which aggregate metrics and trend files are made available. In all cases a 75% data coverage criterion is applied. DOI: <https://doi.org/10.1525/elementa.244.t7>

Aggregation period	Description
annual	annual statistics
summer	summertime statistics; summer is defined as April–September in the Northern Hemisphere and October–March in the Southern Hemisphere <sup>a</sup> ; this season corresponds to the 6-month growing season of several perennial plants (see <i>TOAR-Vegetation</i> )
xsummer	extended summertime period of 7 months from April to October in the Northern Hemisphere and October to April in the Southern Hemisphere <sup>a</sup> ; this corresponds to the growing period of other perennial plants (see <i>TOAR-Vegetation</i> )
seasonal	data extractions for meteorological seasons, i.e. December–February (DJF), March–May (MAM), June–August (JJA), and September–October (SON) <sup>b</sup>
monthly	statistics are reported for each calendar month
RICE	data are extracted according to rice growing seasons in various climatic zones (see <i>TOAR-Vegetation</i> and Supplemental Material 1, section 3 for details); data are only extracted at stations located in rice growing regions (station_rice_production > 0.01 t/year)
WHEAT	as above but for wheat growing seasons

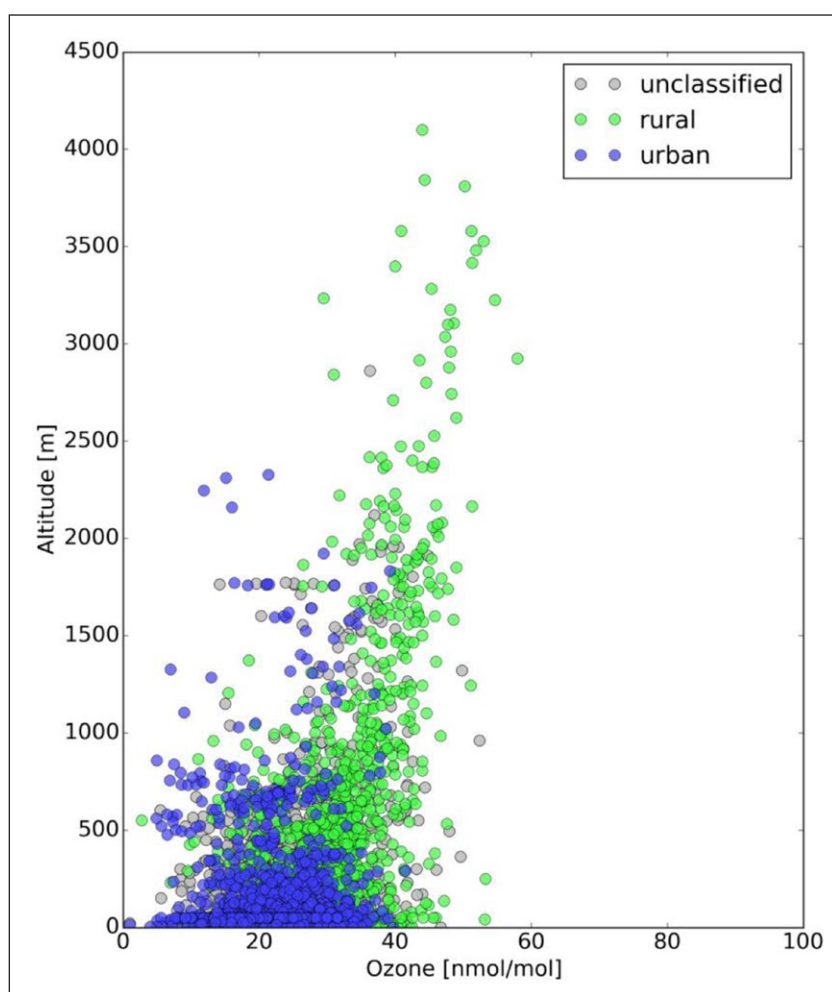
<sup>a</sup>Labelling of summer in the Southern Hemisphere uses the year in which the season begins; for example, "summer 2011" is defined as October 2011 to March 2012. Note that this extends the data extraction intervals beyond the ranges given in Table 6; "summer 2010–2014" at southern hemisphere stations actually means October 2010 to March 2015. For simplicity, there is no special treatment of tropical latitudes.

<sup>b</sup>The month of December is taken from the year before the actual year; for example "DJF 2011" ranges from December 2010 to February 2011.

intended for the evaluation of global chemistry models. The grid size of 5° × 5° longitude-latitude was chosen in order to find a compromise between maximizing the data coverage and preserving regional-scale features. Data sets with other grid resolutions are also available on the TOAR data portal. While the maximum daytime average mole fractions are seen over Europe and North America during July, the largest values actually occur over East Asia in

April and May (see **Figure 7** and map plots on the TOAR data portal for details). The lowest values are seen over the Southern Ocean and South America during winter (i.e. July).

In **Figure 7**, seasonal cycles from the gridded daytime averages are shown at selected latitudes. Arctic and Antarctic sites exhibit the largest mole fractions during winter, while mid-latitudes show maxima either in



**Figure 5:** Annual median ozone mixing ratios versus altitude at 5491 stations, stratified by TOAR classification. Data from 2008 to 2015 were included, and the average annual median value at each station is displayed. DOI: <https://doi.org/10.1525/elementa.244.f5>

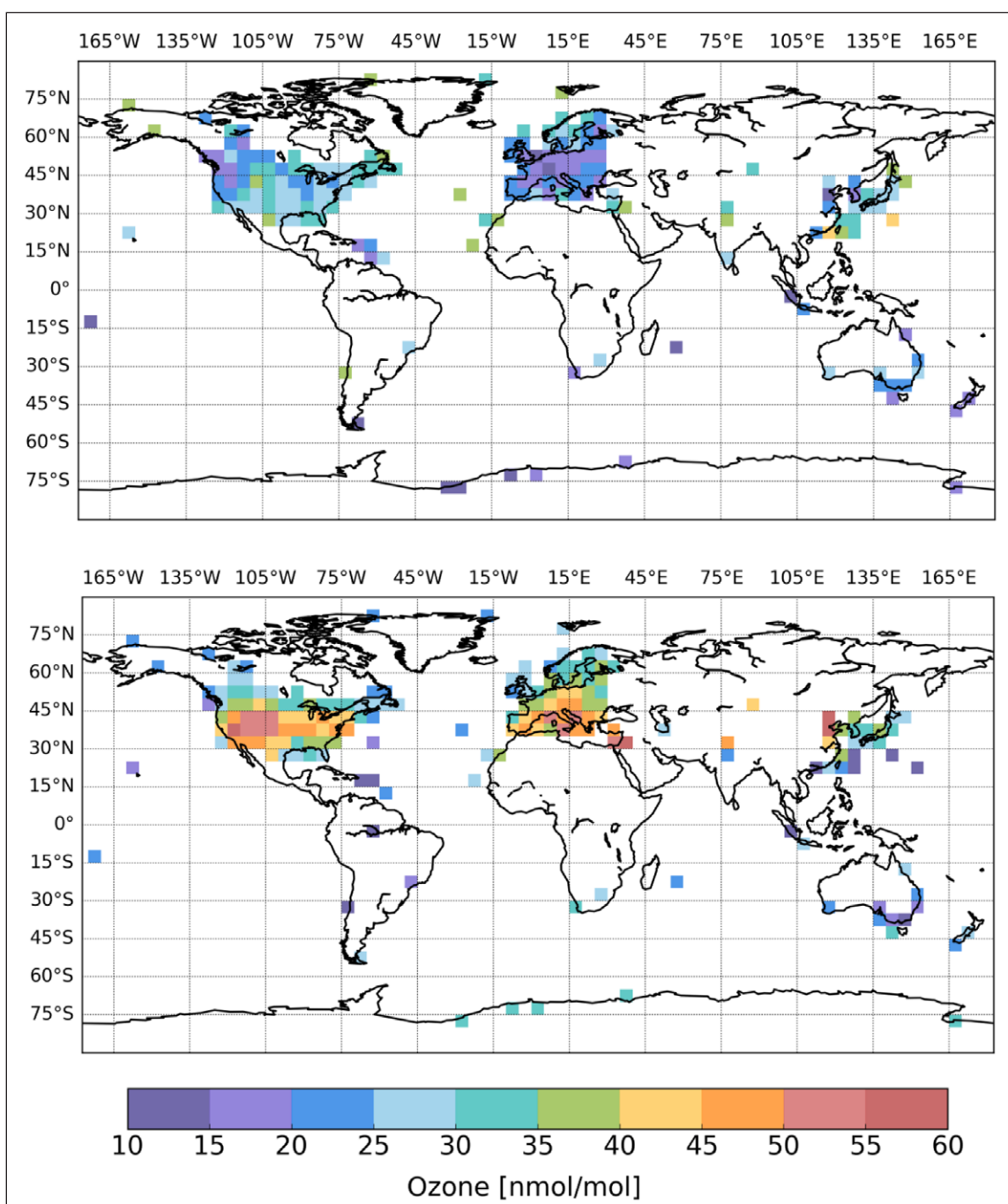
summer or spring, and tropical sites generally do not exhibit any pronounced seasonality. There are some distinct differences in the seasonal cycles of ground-level ozone in different regions (longitudes). For example, in the Northern Hemisphere subtropics ( $15^{\circ}$ – $30^{\circ}$ N), Asian stations show a clear impact of the summer monsoon (low values in June and July, see green lines). A similar effect is also seen over the Southeastern US (purple line), but it is practically absent over Europe and other parts of North America. In the temperate regions of the Northern Hemisphere ( $30^{\circ}$ – $45^{\circ}$ N and  $45^{\circ}$ – $60^{\circ}$ N) some regions exhibit a maximum in summer (July–August), while others show a springtime maximum. This has been extensively discussed in the literature (e.g. Derwent et al., 1998; Monks, 2000; Wang et al., 2003; Cooper et al., 2014). The seasonal cycle of ozone at stations in Southern Asia and Eastern Asia at  $30^{\circ}$ – $45^{\circ}$ N again indicate the influence of the Asian summer monsoon, although the minimum during June–July is less well developed than at  $15^{\circ}$ – $30^{\circ}$ N.

In **Figure 8** we compare the seasonal cycles of different metrics for all rural sites in the latitude range  $30^{\circ}$ – $45^{\circ}$ N. The 5<sup>th</sup> percentiles all maximize in spring and the ozone mole fractions are astonishingly consistent among regions.

The summertime values fall into two groups: one group shows low ozone due to the summer monsoon, particularly in Asia, while the other group shows a steady decrease of the 5<sup>th</sup> percentiles towards the winter minimum. Median values exhibit greater variability: while some regions have the maximum in summer, others show again a springtime maximum. The 95<sup>th</sup> percentiles are more consistent with respect to the seasonal cycle, with all regions displaying a broad summer maximum, except for East Asia, where the maximum occurs in April and May and it is also considerably larger than the maxima elsewhere.

## 7. Conclusions

The TOAR database constitutes the world's largest collection of surface ozone data achieved so far and is the primary source of information for the TOAR exploratory analyses of the global distribution and trends of tropospheric ozone using metrics relevant to the impact of ozone on human health, vegetation, and climate. Data from more than 9,600 sites around the world have been brought together in one single repository and thereby coverage in many world regions has substantially increased compared to previous global ozone data collections. However, important gaps still remain in several parts of



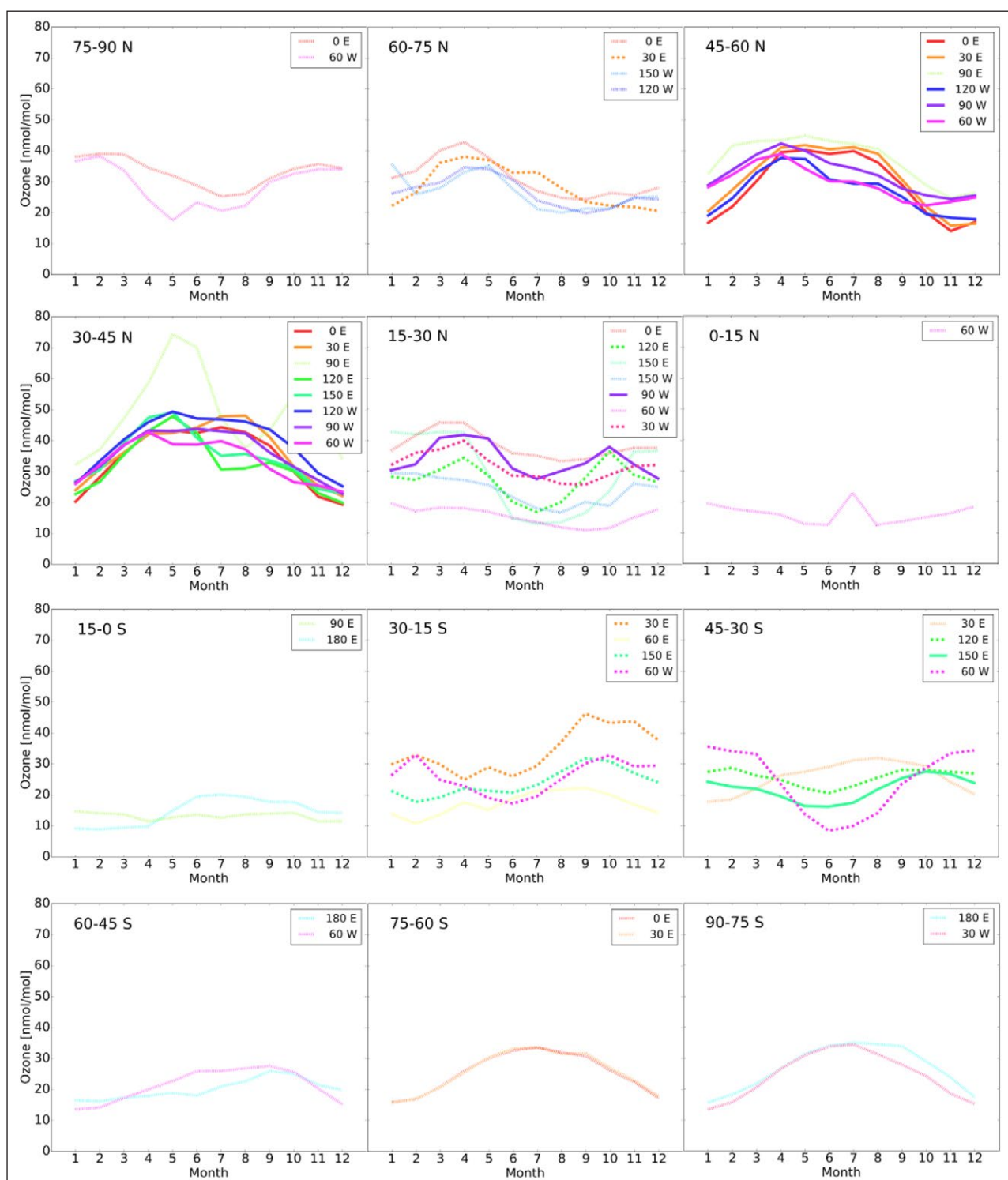
**Figure 6:** Monthly daytime (8–20 h local time) average ozone mole fractions during 2010–2014 gridded onto a  $5^\circ \times 5^\circ$  longitude-latitude grid. Top: January, bottom: July. All stations at altitudes below 2000 m and with at least 3 years of data during the interval were included. Similar maps are available for all months, for rural and urban stations, and for various metrics on the TOAR data portal. DOI: <https://doi.org/10.1525/elementa.244.f6>

the world as there are few measurements in Africa, Central and South America, Central and Southern Asia and the Middle Eastern region (or data from these regions were not accessible).

Centralized extractions of aggregated metrics data sets applying consistent filtering and statistics routines to the entire data set for all TOAR analyses are made available through the TOAR data portal at the PANGAEA publishing site. Additional methods to access TOAR data are the JOIN web interface and the associated REST services. The TOAR database allows for novel ways of analysing surface ozone

data and trends as a result of the extensive data collection, the data quality control efforts, the multi-faceted metadata, and the consistent processing of all surface ozone observations. One particular example is the first objective, globally consistent station characterisation through combination of various global, gridded data sets. Demonstrations of new surface ozone analyses are given in this paper.

An important lesson from the TOAR database building effort was the recognition of the value of intensive communication with ozone data centers and other data providers and the appreciation of their work, expressed also

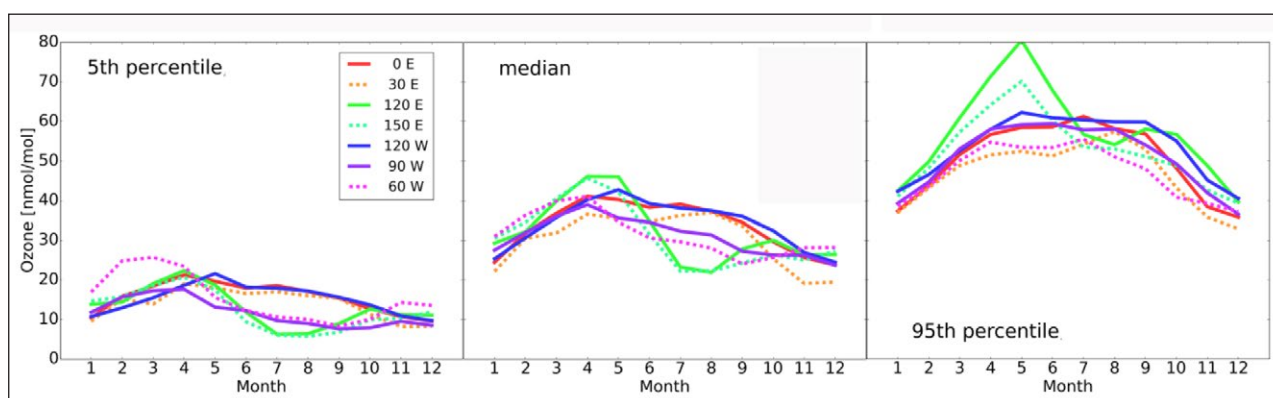


**Figure 7:** Seasonal cycles of the gridded daytime average product in  $30^\circ \times 15^\circ$  resolution for 2010–2014. Each panel contains all seasonal cycles within one  $15^\circ$  latitude band. Colors (i.e. hue) represent the longitude (in  $30^\circ$  bins). Solid lines represent grid boxes where at least 30 sites have valid data (i.e. at least 3 years with 75% coverage), dashed lines represent grid boxes with 6 to 29 sites, and dotted lines represent grid boxes with less than 6 sites. As in Figure 6, all stations at altitudes less than 2000 m and with at least 3 years of data during the interval were included. DOI: <https://doi.org/10.1525/elementa.244.f7>

by the long author list of this paper. As there is no global standardization of atmospheric composition metadata, data formats and data access methods, the data collection effort is rather complex and error-prone (see discussion in Sofen et al., 2016b). Only by intensive communication with

the original data providers has it been possible to resolve errors and ambiguities, and we hope that the feedback they have received from the TOAR database curators will help to improve the quality and consistency of the original ozone data repositories.





**Figure 8:** As Figure 7, but for the 5th percentiles, medians, and 95th percentiles at rural stations in the latitude range 30°–45°N. DOI: <https://doi.org/10.1525/elementa.244.f8>

To enhance the usability and value of surface ozone and air quality observations worldwide, the community must make further progress on the data and metadata harmonization and quality control. We recommend to managers of air quality data that: (i) data centers improve the documentation of their metadata and include more information on the measurement sites, the instruments and calibration techniques in their repositories; (ii) station coordinates be provided with better accuracy and station locations be verified in order to allow analysis of ozone data together with high resolution geographical information, thereby enabling new applications such as automatic search for nearby pollution sources; (iii) more automated quality control tools be developed and harmonized among data centers. Automated quality control tools should be based on rigorous statistical methods and may benefit from new methods of big data analytics (e.g. “deep learning”) in order to test data sets for consistency in space and time and use additional information such as meteorological fields or ozone precursor data to test the plausibility of reported ozone concentrations. Finally, we express our hope that access to quality controlled hourly observations of air quality will be facilitated through the implementation of state-of-the-art web services. At present a lot of manual intervention is needed in order to collect all available observations. Over the coming years we aim to further expand and continue updates of the TOAR database. Recent additions include ozone precursor and meteorological data from European stations. While the existence and accessibility of the TOAR database is guaranteed for several years, its further development will depend on the availability of funding.

The database and tools that have been developed in the context of TOAR constitute an important step towards a global data architecture (see for example <https://www.rd-alliance.org/about-rda/who-rda.html>) and are used as examples in presentations and discussions on this topic. This work shows a lot of potential but also reveals many issues with respect to building the fully interoperable, federated data repositories ([https://en.wikipedia.org/wiki/Federated\\_database\\_system](https://en.wikipedia.org/wiki/Federated_database_system)) of the future. In particular, it clearly demonstrates that the cooperative involvement of many people is needed in order to bring

such systems to life and to make sure that the underlying information is of known quality, robust, documented, and traceable.

### Data Accessibility Statement

General access to TOAR data is free and unrestricted through the JOIN web interface (<https://join.fz-juelich.de/>) and its associated REST service (see documentation in Supplemental Material). This applies to all data products on daily and coarser time resolution. The hourly ozone data in the TOAR database are not publically available due to restrictions imposed by individual data providers. For use of the interactive JOIN web interface, a registration is required. Many of the original hourly ozone observations are, however, available from the original data center websites (see Table 1 of this article). The TOAR data portal on PANGAEA (<https://doi.org/10.1594/PANGAEA.876108>) contains ozone statistics (including metrics for assessing health, vegetation, and climate impacts), trend estimates, and graphical material. The TOAR data portal also provides free and unrestricted access. All use of TOAR surface ozone data should include a reference to this article.

### Supplemental Files

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

- **Supplement 1.** Documentation of TOAR surface ozone data products. DOI: <https://doi.org/10.1525/elementa.244.s1>
- **Supplement 2.** Documentation of the Jülich Open Web Interface for accessing TOAR surface ozone data. DOI: <https://doi.org/10.1525/elementa.244.s2>

### Note

- <sup>1</sup> In most cases the results provided through TOAR will be very close to official statistics, but there are subtle differences in the implementation of various TOAR metrics which may change the results of such analysis. Furthermore, as described in section 5, we performed extensive additional quality control on the TOAR database records and even though we reported data flagging changes or

unit errors back to the data providers, these may not always be adopted in the original archives. Finally, some data flagging schemes can be ambiguous with respect to interpreting “unusual” data as either valid or invalid. Such data may be valid for TOAR purposes, but may be unsuitable for judging whether a given station attained the legal air quality criteria or not.

### Acknowledgements and Funding information

This work is part of the Tropospheric Ozone Assessment Report (TOAR) which was supported by the International Global Atmospheric Chemistry (IGAC) project, the National Oceanic and Atmospheric Administration (NOAA), Forschungszentrum Jülich, and the World Meteorological Organisation (WMO). Many institutions and agencies supported the implementation of the measurements, and the processing, quality assurance, and submission of the data contained in the TOAR database.

Martin Schultz, Sabine Schröder, and Olga Lyapina acknowledge funding through the Helmholtz ATMO program, support from Andreas Wahner and Astrid Kiendler-Scharr, and technical support from Michael Decker. Paulo Artaxo acknowledge funding from FAPESP and CNPq. Samera Hamad would like to thank Mr. Ali Jabir at the Iraqi Ministry of Environment, Baghdad, Iraq for his help in providing the data from the monitoring station at central Baghdad. X. Xu acknowledges funding from the National Natural Science Foundation of China (grant 41330442). Ma Zhiquiang thanks the National Science Foundation of China for funding grant 41475135. Katie Read wishes to acknowledge specific funding from the UK Natural Environmental Research Council (NERC) through the National Centre for Atmospheric Science Atmospheric Measurement Facility (NCAS AMF). Vinayak Sinha acknowledges the IISER Mohali Atmospheric Chemistry Facility and Ministry of Human Resource Development (MHRD), India for funding. Rodrigo Seguel acknowledges support from CONICYT, FONDECYT Program (Project No 11130177). James Schwab received funding from the New York State Energy Research and Development Authority. Andrey Skorokhod acknowledges funding from the Russian Science Foundation (grant no. 14-47-00049) and Department of Earth Sciences RAS (project no. 0150-2015-0053). Manish Naja acknowledge the support received from ISRO-ATCTM project and ARIES, DST for the observations. Lynn Mazzoleni and Claudio Mazzoleni wish to acknowledge funding from the U.S. National Science Foundation (grant #AGS-1110059) and the Atmospheric System Research (ASR) program of the U.S. Department of Energy (grant #DE-SC0006941). Paul Eckhardt, Markus Fiebig, Anne-Gunn Hjellbrekke and Ove Hermansen acknowledge funding from the Norwegian Environment Agency. Stephen Springston acknowledges support from the Atmospheric Radiation Measurement (ARM) Climate Research Facility program of the U.S. Department of Energy. Margarita Yela Gonzalez and Monica Navarro Comas acknowledge funding for Belgrano through grants MARACA (CGL2004-05419-C02-01/ANT), CE-ORACLEO3 (POL2006-00382) and VIOLIN (CGL2010-20353). Tao Wang acknowledges support from the Hong

Kong Polytechnic University (G-S023) and The Hong Kong Research Grants Council (PolyU 153042/15E). Ozone measurements from Gadanki, India were supported by ISRO-GBP AT-CTM project and Department of Space, Government of India. Yvonne Scorgie acknowledges the New South Wales Government as the funding source for the NSW Air Quality Monitoring Network, and the Office of Environment and Heritage and managers of the network and curators of the monitoring data set. Observation data from Pic du Midi were supported by the University Paul Sabatier, Toulouse, France, and CNRS. Valerie Gros acknowledges support from IPEV, TAAF, CNRS and CEA for measurements at Amsterdam Island. Ozone observation at Puy de Dôme was mostly supported by CNRS. Ozone observation at Donon was conducted by a French regional air quality agency, ASPA. The measurements at Jungfrauoch are run by Empa in collaboration with the Swiss Federal Office for the Environment. The monitoring data in Japan have been compiled by AEROS (Atmospheric Environmental Regional Observation System), and obtained from the Environmental Database provided by National Institute for Environmental Studies, Japan. Data from Poland were kindly provided by the Chief Inspectorate of Environmental Protection, Department of Monitoring and Information on Environment, Warsaw, Poland. Data from the Canadian Air and Precipitation Monitoring Network was kindly provided by Environment and Climate Change Canada. The Mt. Bachelor Observatory is supported by the National Science Foundation 559 (grant #AGS-1447832) and the National Oceanic and Atmospheric Administration 560 (contract #RA-133R-16-SE-0758). We thank Carolyn Jordan for help accessing the AIRMAP data. We would also like to acknowledge all former colleagues who initiated the long-term monitoring of surface ozone at various sites many years ago but are no longer active in the field, and we apologize if we failed to identify additional contributors or data set PIs.

The source of the human population data is: Gridded Population of the World: Future Estimates, for the year 2010 Data produced by Center for International Earth Science Information Network (CIESIN), Columbia University. These data were downloaded from <http://sedac.ciesin.columbia.edu/gpw> on October 19, 2011. DMSP-OLS nighttime lights of the world (V4) detected by The Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS). Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency. Data downloaded from: <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>. IPCC climatic zone data were extracted by Katrina Sharps. The rice and wheat growing regions were downloaded from the GAEZ data portal <http://gaez.fao.org/Main.html#>) and then output to ascii text format by Katrina Sharps.

### Competing interests

An Elementa EIC is a co-author on this publication. His editorial responsibilities were fully covered by the Elementa Sustainable Engineering Domain EIC and he had no influence on the peer-review process.

## Author contributions

- Contributed to conception and design: MGS, SS, OC
- Contributed to acquisition of data: all co-authors; IP, EvS, HT, YE, MN, RJS acted as regional working group leaders; UD, PE, SF, MF, AGH, YDH, PCK, HK, GL, (NM), (KS), DT, MU, and MW are curators of major environmental data archives and helped with formatting and interpreting data and metadata; MGS, and SS designed the TOAR database and ingested all data and metadata; MGS, and OL performed the quality assurance of data and metadata; DB, MTC, RG, ML, SM, RM, and TW played a major role in collecting previously unavailable data; OC, (CML), and KS helped in the provision of gridded metadata; all other authors contributed one or several data sets and helped in the formatting and interpretation of the data and metadata
- Contributed to analysis and interpretation of data: MGS, SS, OL, OC, IG
- Drafted and revised the article: MGS
- Approved the submitted version for publication: all co-authors

## References

- Amante, C** and **Eakins, BW** 2009 ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. accessed 30 April 2016. DOI: <https://doi.org/10.7289/V5C8276M>
- AQICN** 2017 <https://aqicn.org/map/#@g/23.7604/121.3464/8z>.
- Atkins, DHF, Cox, RA** and **Eggleton, AEJ** 1972 Photochemical Ozone and Sulfuric Acid Aerosol Formation in the Atmosphere Over Southern England. *Nature*, **235**: 372–376. DOI: <https://doi.org/10.1038/235372a0>
- Bartel, AW** and **Temple, JW** 1952 Ozone in Los Angeles and Surrounding Areas. *Ind. Eng. Chem.*, **44**(4): 857–861. DOI: <https://doi.org/10.1021/ie50508a043>
- Chevalier, A, Gheusi, F, Delmas, R, Ordóñez, C, Sarrat, C, Zbinden, R, Thouret, V, Athier, G** and **Cousin, JM** 2007 Influence of altitude on ozone levels and variability in the lower troposphere: a ground-based study for western Europe over the period 2001–2004. *Atmos. Chem. Phys.*, **7**: 4311–4326. DOI: <https://doi.org/10.5194/acp-7-4311-2007>
- CIESIN** 2005 Center for International Earth Science Information Network – CIESIN – Columbia University, and Centro Internacional de Agricultura Tropical – CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid, Future Estimates. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Accessed October 19, 2011. DOI: <https://doi.org/10.7927/H4ST7MRB>
- Cole, AFW** and **Katz, M** 1966 Summer Ozone Concentrations in Southern Ontario in Relation to Photochemical Aspects and Vegetation Damage. *Journal of the Air Pollution Control Association*, **16**(4): 201–206. DOI: <https://doi.org/10.1080/00022470.1966.10468463>
- Cooper, OR, Parrish, DD, Ziemke, J, Balashov, NV, Cupeiro, M**, et al. 2014 Global distribution and trends of tropospheric ozone: An observation-based review. *Elem. Sci. Anth.*, **2**: 000029. DOI: <https://doi.org/10.12952/journal.elementa.000029>
- Cristofanelli, P, Bonasoni, P, Tositti, L, Bonafe, U, Calzolari, F, Evangelisti, F, Sandrini, S** and **Stohl, A** 2006 A 6-year analysis of stratospheric intrusions and their influence on ozone at Mt. Cimone (2165 m). *J. Geophys. Res.*, **111**(D03): 306. DOI: <https://doi.org/10.1029/2005JD006553>
- Derwent, RG, Simmonds, PG, Seuring, S** and **Dimmer, C** 1998 Observation and interpretation of the seasonal cycles in the surface concentrations of ozone and carbon monoxide at Mace Head, Ireland from 1990 to 1994. *Atmos. Environ.*, **32**(2): 145–157. DOI: [https://doi.org/10.1016/S1352-2310\(97\)00338-5](https://doi.org/10.1016/S1352-2310(97)00338-5)
- Derwent, RG** and **Stewart, HNM** 1973 Elevated ozone levels in the air of Central London. *Nature*, **241**(5388): 342–343. DOI: <https://doi.org/10.1038/241342a0>
- Fiore, AM, Dentener, FJ, Wild, O, Cuvelier, C, Schultz, MG, Hess, P, Textor, C, Schulz, M, Doherty, RM, Horowitz, LW, MacKenzie, IA, Sanderson, MG, Shindell, DT, Stevenson, DS, Szopa, S, van Dingenen, R, Zeng, G, Atherton, C, Bergmann, D, Bey, I, Carmichael, G, Collins, WJ, Duncan, BN, Faluvegi, G, Folberth, G, Gauss, M, Gong, S, Hauglustaine, D, Holloway, T, Isaksen, ISA, Jacob, DJ, Jonson, JE, Kaminski, JW, Keating, TJ, Lupu, A, Marmer, E, Montanaro, V, Park, RJ, Pitari, G, Pringle, KJ, Pyle, JA, Schroeder, S, Vivanco, MG, Wind, P, Wojcik, G, Wu, S** and **Zuber, A** 2009 Multi-model Estimates of Intercontinental Source-Receptor Relationships for Ozone Pollution. *J. Geophys. Res.*, **114**(D04): 301. DOI: <https://doi.org/10.1029/2008JD010816>
- Fleming, Z**, et al. 2017 Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health, referred to as *TOAR-Health*, Under review for TOAR Special Feature.
- Fowler, A** and **Strutt, RJ** 1917 Atmospheric bands of atmospheric ozone in the spectra of sun and stars. *Proc. Roy. Soc.*, **93**: 577–583. London A. DOI: <https://doi.org/10.1098/rspa.1917.0041>
- Galbally, IE** 1968 Some measurements of ozone variation and destruction in the atmospheric surface layer. *Nature*, **218**(5140): 456–457. DOI: <https://doi.org/10.1038/218456a0>
- Galbally, IE** 1971 Surface ozone observations at Aspendale, Victoria, 1964–1970. *Atmos. Environ.*, **5**: 15–25. DOI: [https://doi.org/10.1016/0004-6981\(71\)90040-0](https://doi.org/10.1016/0004-6981(71)90040-0)
- Galbally, IE, Miller, AJ, Hoy, RD, Ammet, S, Joynt, RC** and **Attwood, D** 1986 Ozone in the surface air over the Latrobe Valley, Victoria and Cape Grim, Tasmania, Australia. *Atmos. Environ.*, **20**: 2403–2422. DOI: [https://doi.org/10.1016/0004-6981\(86\)90071-5](https://doi.org/10.1016/0004-6981(86)90071-5)



- Galbally, IE** and **Schultz, MG** (eds.) 2013 Guidelines for Continuous Measurements of Ozone in the Troposphere. GAW Report #209. Geneva, Switzerland: World Meteorological Organization. Available at: [http://www.wmo.int/pages/prog/arep/gaw/documents/Final\\_GAW\\_209\\_web.pdf](http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_209_web.pdf) Accessed on December 4, 2016.
- Gaudel, AJ**, et al. 2017 Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to climate and global model evaluation, referred to as *TOAR-Climate*. In preparation for TOAR Special Feature.
- Google** 2017 Last accessed 25<sup>th</sup> March, 2017.
- Haagen-Smit, AJ** 1952 Chemistry and Physiology of Los Angeles Smog. *Ind. Eng. Chem.*, **44**(6): 1342–1346. DOI: <https://doi.org/10.1021/ie50510a045>
- Helmig, D**, et al. 2007 A review of surface ozone in the polar regions. *Atmos. Environ.*, **41**: 5138–5161. DOI: <https://doi.org/10.1016/j.atmosenv.2006.09.053>
- IPCC** 2007 Solomon, S, Qin, D, Manning, M, Chen, Z, Marquis, M, Averyt, KB, Tignor, M and Miller, HL (eds.), Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Janssens-Maenhout, G, Crippa, M, Guizzardi, D, Dentener, F, Muntean, M, Pouliot, G, Keating, T, Zhang, Q, Kurokawa, J, Wankmüller, R, Denier van der Gon, H, Kuenen, JJP, Klimont, Z, Frost, G, Darras, S, Koffi, B and Li, M** 2015 HTAP\_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution. *Atmos. Chem. Phys.*, **15**: 11411–11432. DOI: <https://doi.org/10.5194/acp-15-11411-2015>
- Koffi, B, Dentener, F, Janssens-Maenhout, G, Guizzardi, D, Crippa, M, Diehl, T, Galmarini, S and Solazzo, E** 2016 Hemispheric Transport Air Pollution (HTAP): Specification of the HTAP2 experiments – Ensuring harmonized modelling, EUR 28255 EN, Luxembourg: Publications Office of the European Union. DOI: <https://doi.org/10.2788/725244>
- Kondo, J and Akimoto, H** 1975 Air Pollution in Tokyo: An Application of Computer Simulation, In: Pitts, JN, Jr. and Metcalf, RN (eds.), *Advances in environmental science and technology*, **5**: 371. New York, London, Wiley-Interscience.
- Krotkov, NA, McLinden, CA, Li, C, Lamsal, LN, Celarier, EA, Marchenko, SV, Swartz, WH, Bucsela, EJ, Joiner, J, Duncan, BN, Boersma, KF, Veefkind, JP, Levelt, PF, Fioletov, VE, Dickerson, RR, He, H, Lu, Z and Streets, DG** 2016 Aura OMI observations of regional SO<sub>2</sub> and NO<sub>2</sub> pollution changes from 2005 to 2015. *Atmos. Chem. Phys.*, **16**: 4605–4629. DOI: <https://doi.org/10.5194/acp-16-4605-2016>
- Lacasaña-Navarro, M, Aguilar-Garduño, C and Romieu, I** 1999 Evolución de la contaminación del aire e impacto de los programas de control en tres megaciudades de América Latina. *Salud Publica Mex*, **41**: 203–215. DOI: <https://doi.org/10.1590/S0036-36341999000300008>
- Lefohn, AS**, et al. 2017 Tropospheric Ozone Assessment Report: Global ozone metrics for climate change, human health, and crop/ecosystem research, referred to as *TOAR-Metrics*, Under review for TOAR Special Feature.
- Mills, G**, et al. 2017 Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to vegetation, referred to as *TOAR-Vegetation*. In preparation for TOAR Special Feature.
- Monks, PS** 2000 A review of the observations and origins of the spring ozone maximum. *Atmos. Environ.*, **34**(21): 3545–3561. DOI: [https://doi.org/10.1016/S1352-2310\(00\)00129-1](https://doi.org/10.1016/S1352-2310(00)00129-1)
- Monks, PS, Archibald, AT, Colette, A, Cooper, O, Coyle, M, Derwent, R, Fowler, D, Granier, C, Law, KS, Mills, GE, Stevenson, DS, Tarasova, O, Thouret, V, von Schneidmesser, E, Sommariva, R, Wild, O and Williams, ML** 2015 Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmos. Chem. Phys.*, **15**: 8889–8973. DOI: <https://doi.org/10.5194/acp-15-8889-2015>
- Oberlander, EA, Brenninkmeijer, CAM, Crutzen, PJ, Elansky, NF, Golitsyn, GS, Granberg, IG, Scharffe, DH, Hofmann, R, Belikov, IB, Paretzke, HG and van Velthoven, PFJ** 2002 Trace gas measurements along the Trans-Siberian railroad: The TROICA 5 expedition. *J. Geophys. Res.*, **107**(D14). DOI: <https://doi.org/10.1029/2001JD000953>
- Oltmans, SJ, Lefohn, AS, Harris, JM, Galbally, I, Scheel, HE, Bodeker, G, Brunke, E, Claude, H, Tarasick, D, Johnson, BJ, Simmonds, P, Shadwick, D, Anlauf, K, Hayden, K, Schmidlin, F, Fujimoto, T, Akagi, K, Meyer, C, Nichol, S, Davies, J, Redondas, A and Cuevas, E** 2006 Long-term changes in tropospheric ozone. *Atmos. Env.*, **40**: 3156–3173. DOI: <https://doi.org/10.1016/j.atmosenv.2006.01.029>
- Oltmans, SJ, Lefohn, AS, Shadwick, D, Harris, JM, Scheel, HE, Galbally, I, Tarasick, DW, Johnson, BJ, Brunke, EG, Claude, H, Zeng, G, Nichol, S, Schmidlin, F, Davies, J, Cuevas, E, Redondas, A, Naoe, H, Nakano, T and Kawasato, T** 2012 Recent tropospheric ozone changes – A pattern dominated by slow or no growth. *Atmos. Environ.*, **67**: 331–351. DOI: <https://doi.org/10.1016/j.atmosenv.2012.10.057>
- Pankratova, NV, Elansky, NF, Belikov, IB, Lavrova, OV, Skorokhod, AI and Shumsky, RA** 2011 Ozone and Nitric Oxides in the Surface Air Over Northern Eurasia According to Observational Data Obtained in TROICA Experiments. *Izvestiya, Atmospheric and Oceanic Physics*, **47**(3): 313–328. DOI: <https://doi.org/10.1134/S0001433811030108>
- Rasmussen, DJ, Fiore, AM, Naik, V, Horowitz, LW, McGinnis, SJ and Schultz, MG** 2012 Surface ozone-temperature relationships in the eastern US: A monthly climatology for evaluating



- chemistry-climate models. *Atmos. Env.*, **47**: 142–153. DOI: <https://doi.org/10.1016/j.atmosenv.2011.11.021>
- Riveros, HG, Arriaga, JL, Tejeda, J, Julian-Sanchez, A and Riveros-Rosas, H** 1998 Ozone and its precursors in the atmosphere of Mexico City. *J. Air and Waste Manage Assoc.*, **48**: 866–871. DOI: <https://doi.org/10.1080/10473289.1998.10463730>
- Schultz, MG, Akimoto, H, Bottenheim, J, Buchmann, B, Galbally, IE, et al.** 2015 The Global Atmosphere Watch reactive gases measurement network. *Elem Sci Anth*, **3**: 000067. DOI: <https://doi.org/10.12952/journal.elementa.000067>
- Seinfeld, JH, Atkinson, R, Berglun, RL, Chameides, WL, Cotton, WR, Demerjian, KL, Elston, JC, Fehsenfeld, F, Finlayson-Pitts, B, Harriss, RC, Kolb, CE, Jr., Lioy, PJ, Logan, JA, Prather, MJ, Russell, A and Steigerwald, B** 1991 Rethinking the ozone problem, National Academy Press, Washington, 524. Accessible via: <https://www.nap.edu/catalog/1889/rethinking-the-ozone-problem-in-urban-and-regional-air-pollution> accessed 31 Jan 2017.
- Simpson, WR, von Glasow, R, Riedel, K, Anderson, P, Ariya, P, Bottenheim, J, Burrows, J, Carpenter, LJ, Frieß, U, Goodsite, ME, Heard, D, Hutterli, M, Jacobi, HW, Kaleschke, L, Neff, B, Plane, J, Platt, U, Richter, A, Roscoe, H, Sander, R, Shepson, P, Sodeau, J, Steffen, A, Wagner, T and Wolff, E** 2007 Halogens and their role in polar boundary-layer ozone depletion. *Atmos. Chem. Phys.*, **7**: 4375–4418. DOI: <https://doi.org/10.5194/acp-7-4375-2007>
- Sofen, ED, Bowdalo, D and Evans, MJ** 2016a 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>
- Sofen, ED, Bowdalo, D, Evans, MJ, Apadula, F, Bonasoni, P, Cupeiro, M, Ellul, R, Galbally, IE, Girgziene, R, Luppo, S, Mimouni, M, Nahas, AC, Saliba, M and Tørseth, K** 2016b Gridded global surface ozone metrics for atmospheric chemistry model evaluation. *Earth Syst. Sci. Data*, **8**: 41–59. DOI: <https://doi.org/10.5194/essd-8-41-2016>
- Staehelin, J, Thudium, J, Buehler, R, Volz-Thomas, A and Graber, W** 1994 Trends in surface ozone concentrations at Arosa (Switzerland). *Atmos. Environ.*, **28**(1): 75–87. DOI: [https://doi.org/10.1016/1352-2310\(94\)90024-8](https://doi.org/10.1016/1352-2310(94)90024-8)
- Tarasick, D, et al.** 2017 Tropospheric Ozone Assessment Report: Tropospheric ozone observations, referred to as *TOAR-Observations*. In preparation for TOAR Special Feature.
- Tarasick, DW and Bottenheim, JW** 2002 Surface ozone depletion episodes in the Arctic and Antarctic from historical ozonesonde records. *Atmos. Chem. Phys.*, **2**(3): 197–205. DOI: <https://doi.org/10.5194/acp-2-197-2002>
- Velchev, K, Cavalli, F, Hjorth, J, Marmer, E, Vignati, E, Dentener, F and Raes, F** 2011 Ozone over the Western Mediterranean Sea – results from two years of shipborne measurements. *Atmos. Chem. Phys.*, **11**(2): 675–688. DOI: <https://doi.org/10.5194/acp-11-675-2011>
- Verdi, TS** 1874 in Board of Health of the District of Columbia, Annual Report, Chronicle Publishing Company, Washington DC, 168–178. Accessible as: <https://books.google.pt/books?id=Sjk4AAAAAMAAJ&lpg=PA170&ots=o1Cp9clgQt&dq=berigny%20james%20sedan&pg=PA168#v=onepage&q=berigny%20james%20sedan&f=false> (accessed on 24 Jan 2017).
- Wang, X, Shen, Z, Cao, J, Zhang, L, Liu, L, Li, J, Liu, S and Sun, Y** 2012 Characteristics of surface ozone at an urban site of Xi'an in Northwest China. *J. Environ. Monit.*, **14**: 116–124. DOI: <https://doi.org/10.1039/C1EM10541H>
- Wang, YH, Ridley, B, Fried, A, Cantrell, C, Davis, D, Chen, G, Snow, J, Heikes, B, Talbot, R, Dibb, J, Flocke, F, Weinheimer, A, Blake, N, Blake, D, Shetter, R, Lefer, B, Atlas, E, Coffey, M, Walega, J and Wert, B** 2003 Springtime photochemistry at northern mid and high latitudes. *J. Geophys. Res.*, **108**(D4): 8358. DOI: <https://doi.org/10.1029/2002JD002227>
- Young, PJ, Archibald, AT, Bowman, KW, Lamarque, JF, Naik, V, Stevenson, DS, Tilmes, S, Voulgarakis, A, Wild, O, Bergmann, D, Cameron-Smith, P, Cionni, I, Collins, WJ, Dalsøren, SB, Doherty, RM, Eyring, V, Faluvegi, G, Horowitz, LW, Josse, B, Lee, YH, MacKenzie, IA, Nagashima, T, Plummer, DA, Righi, M, Rumbold, ST, Skeie, RB, Shindell, DT, Strode, SA, Sudo, K, Szopa, S and Zeng, G** 2013 Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmos. Chem. Phys.*, **13**: 2063–2090. DOI: <https://doi.org/10.5194/acp-13-2063-2013>
- Zurbenko, I, Porter, PS, Rao, ST, Ku, JY, Gui, R and Eskridge, RE** 1996 Detecting Discontinuities in Time Series of Upper-Air Data: Development and Demonstration of an Adaptive Filter Technique. *J. Climate*, **9**: 3548–3560. DOI: [https://doi.org/10.1175/1520-0442\(1996\)009<3548:DDITSO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<3548:DDITSO>2.0.CO;2)

**How to cite this article:** Schultz, MG, Schröder, S, Lyapina, O, Cooper, OR, Galbally, I, Petropavlovskikh, I, von Schneidmesser, E, Tanimoto, H, Elshorbany, Y, Naja, M, Seguel, RJ, Dauert, U, Eckhardt, P, Feigenspan, S, Fiebig, M, Hjellbrekke, A-G, Hong, Y-D, Kjeld, PC, Koide, H, Lear, G, Tarasick, D, Ueno, M, Wallasch, M, Baumgardner, D, Chuang, M-T, Gillett, R, Lee, M, Molloy, S, Moolla, R, Wang, T, Sharps, K, Adame, JA, Ancellet, G, Apadula, F, Artaxo, P, Barlasina, ME, Bogucka, M, Bonasoni, P, Chang, L, Colomb, A, Cuevas-Agulló, E, Cupeiro, M, Degorska, A, Ding, A, Fröhlich, M, Frolova, M, Gadhavi, H, Gheusi, F, Gilge, S, Gonzalez, MY, Gros, V, Hamad, SH, Helmig, D, Henriques, D, Hermansen, O, Holla, R, Hueber, J, Im, U, Jaffe, DA, Komala, N, Kubistin, D, Lam, K-S, Laurila, T, Lee, H, Levy, I, Mazzoleni, C, Mazzoleni, LR, McClure-Begley, A, Mohamad, M, Murovec, M, Navarro-Comas, M, Nicodim, F, Parrish, D, Read, KA, Reid, N, Ries, L, Saxena, P, Schwab, JJ, Scorgie, Y, Senik, I, Simmonds, P, Sinha, V, Skorokhod, AI, Spain, G, Spangl, W, Spoor, R, Springston, SR, Steer, K, Steinbacher, M, Suharguniyawan, E, Torre, P, Trickl, T, Weili, L, Weller, R, Xiaobin, X, Xue, L and Zhiqiang, M 2017 Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations. *Elem Sci Anth*, 5: 58, DOI: <https://doi.org/10.1525/elementa.244>

**Managing Editor-in-Chief:** Michael E. Chang, Georgia Institute of Technology, US

**Guest Editor:** Alastair Lewis, University of York, UK

**Knowledge Domain:** Atmospheric Science

**Part of an *Elementa* Special Feature:** Tropospheric Ozone Assessment Report (TOAR)

**Submitted:** 03 April 2017    **Accepted:** 10 July 2017    **Published:** 18 October 2017

**Copyright:** © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



*Elem Sci Anth* is a peer-reviewed open access journal published by University of California Press.

**OPEN ACCESS** 