An improved Trajectory-mapped Ozonesonde dataset for the Stratosphere and Troposphere (TOST): update, validation and applications

Zhou Zang[1], Jane Liu[1], David Tarasick[2], Omid Moeini[2], Jianchun Bian[3], Jinqiang Zhang[3], Anne M. Thompson[4,5], Roeland Van Malderen[6], Herman G.J. Smit[7], Ryan M. Stauffer[4], Bryan J. Johnson[8] and Debra E. Kollonige[4,9]

[1]Department of Geography and Planning, University of Toronto, Toronto, Canada
[2]Environment and Climate Change Canada, Toronto, Canada
[3]Key Laboratory of Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[4]Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[5]University of Maryland Baltimore County, Baltimore, MD, USA
[6]Royal Meteorological Institute of Belgium, Brussels, Belgium
[7]Institute for Energy and Climate Research: Troposphere (IEK-8), Research Centre Juelich (FZJ), Juelich, Germany.
[8]NOAA/ESRL Global Monitoring Division, Boulder, Colorado, USA
[9]Science Systems and Applications, Inc., Lanham, MD, USA

Correspondence: Jane Liu (janejj.liu@utoronto.ca)

Abstract

A global-scale horizontally- and vertically-resolved ozone climatology provides detailed insights into ozone variability. Here, the seasonal, annual and decadal-monthly Trajectory-mapped Ozonesonde dataset for the Stratosphere and Troposphere (TOST) ozone climatology is improved and updated from 1970-2021 on a grid of 5° × 5° × 1 km (latitude, longitude, and altitude) from the surface to 26 km by the geometric coordinate and from the surface to 20 hPa in 26 pressure levels by the pressure coordinate, with the most recent ozonesonde data re-evaluated following the ASOPOS-2 guidelines (GAW Report No. 268, 2021). Comparison between

ozonesonde and trajectory-derived ozone shows good agreement in each decade, altitude, and station, with relative differences (RD) of 2-4% in the troposphere and 0.5% in the stratosphere. TOST also aligns well with aircraft, the Satellite Aerosol and Gas Experiment (SAGE) and the Microwave Limb Sounder (MLS) datasets. The updated TOST improves data coverage in all latitude bands and altitudes and reduces RD by 14-17% compared to the previous version, employing twice as many ozonesonde profiles and an updated trajectory simulation model. Higher uncertainties in TOST are where data are sparse, i.e., southern high latitudes, tropics and pre-1980s, and where variability is high, i.e., at the surface and upper troposphere and lower stratosphere (UTLS). Caution should therefore be taken when using TOST in these spaces and times. TOST captures global ozone distributions and temporal variations, showing an overall non-significant change in lower stratospheric ozone after 1998. TOST offers users a long record, global coverage, and high vertical resolution.

1. Introduction

Ozone is an important oxidant photochemically linked to the hydroxyl radical in the troposphere, with detrimental effects on crop productivity, natural ecosystems and human health (Fleming et al., 2018; Harmens et al., 2018; Mills et al., 2018; Vicedo-Cabrera et al., 2019). Tropospheric ozone is the third largest greenhouse gas contributing to radiative forcing, particularly in the upper troposphere (Gulev et al., 2021; Szopa et al., 2021; Forster et al., 2021). The global ozone distribution and its long-term changes at different altitudes, longitudes, and latitudes are critical to understanding global ozone variability and its forcing on climate change. While the ozone trends themselves can indicate the impact of changes in climatic dynamics (Hassler et al., 2008) and chemistry, including the effect of the Montreal Protocol (Steinbrecht et al., 2017), long-term horizontally- and vertically-resolved ozone are needed for prescribing,

evaluating and refining ozone simulations in climate models (Hassler et al., 2018), and to quantify changes in radiative forcing and projecting reliable future climate scenarios (Nowack et al., 2015).

Balloon-borne ozonesondes are the principal source of trend-quality long-term records of ozone profiles below ~18 km (Tarasick et al., 2021). In addition, lidar records also provide long-term tropospheric ozone profiles, such as the Observatoire de Haute Provence lidar and the Jet Propulsion Laboratory Table Mountain lidar (Ancellet and Beekmann, 1997; McDermid et al., 2002). However, the horizontal and temporal coverages of both ozonesondes and lidars are limited by the sparse distribution of the stations (less than 100 worldwide for ozonesondes and 9 lidars from the Tropospheric Ozone Lidar Network) and their low observation frequency (1-3 times per week for ozonesondes; 1-5 times per week for lidars) (McDermid et al., 2002; Liu et al., 2013a; Chouza et al., 2019; Ancellet et al., 2022). The In-Service Aircraft for a Global Observing System (IAGOS) program has measured ozone profiles worldwide since 1994 via the instruments onboard a number of commercial aircraft, with high sampling frequency at some airports (Thouret et al., 1998). However, sampling is unevenly distributed both spatially and temporally because the flights are constrained by commercial airlines' operation schedules. Satellite observations have the advantage of providing ozone data on a global scale with consistent quality. However, it is still challenging to retrieve tropospheric ozone through the large stratospheric ozone burden (Bhartia, 2002). Satellite data can provide total column ozone retrievals which yet are not vertically resolved. The satellite ozone profiles have limited vertical sensitivity and the sensitivity decreases strongly toward the surface (Liu et al., 2010; Keppens et al., 2015). The direct retrieval from nadir-viewing instruments typically provides 1-2 pieces of independent information vertically in the troposphere (Tarasick et al., 2019b). Large retrieval errors occur when retrieval sensitivity is low, as the solution relies heavily on the a priori

(Keppens et al., 2015). In addition, single space instruments are of limited lifetime, while long-term studies on ozone require combining measurements from different instruments, which could introduce uncertainty related to the differences among different instruments (Rahpoe et al., 2015). A number of studies have developed long-term (since the 1980s) ozone climatologies by combining ozone measurements with ozonesondes and multiple satellite instruments (McPeters et al., 2007; McPeters and Labow, 2012; Hassler et al., 2018; Bodeker et al., 2021; Bognar et al., 2022), but these datasets are generally zonally-averaged. Chemistry–climate models are also developed to provide ozone fields in 3 dimensions in latitude, longitude, and altitude, especially for long-term, global-scale simulations (Eyring et al., 2010; Chen et al., 2018); these models present our best understanding of processes controlling ozone variations but still suffer from large uncertainties associated with emission inventories, parameterizations, radiation transport schemes, and simulation of the atmospheric circulations (Young et al., 2018; Wild et al., 2020; Griffiths et al., 2021; Zeng et al., 2022). Some advanced models can improve global tropospheric ozone in 3 dimensions by assimilating the satellite data to enhance the modeling accuracy (Miyazaki et al., 2020a; Colombi et al., 2021). However, in addition to the aforementioned sources of uncertainties, such assimilations still rely heavily on the sufficiency and spatial-temporal continuity of the satellite data (Huijnen et al., 2020; Miyazaki et al., 2020b).

Liu et al. (2013a, b) constructed a long-term 3-dimensional global-scale ozone dataset using a trajectory-mapping approach, extending sparse ozonesonde measurements and filling gaps in the spatial domain with backward and forward trajectory simulations. The trajectory-mapping method assumes the ozone mixing ratio in the same air parcel along each trajectory is constant for several days, which is reasonable given that the lifetime of ozone in most of the troposphere and lower stratosphere ranges from days to months, varying with season and altitude (Han et al., 2019; Prather and Zhu, 2024). The constructed global dataset is independent of satellite

measurements and photochemical modeling. The trajectory mapping can outperform conventional statistical interpolation methods (Stohl et al., 2001) because it is based on sound principles of ozone lifetime and wind-driven air movement. The trajectory-derived ozone data cover higher latitudes (to 90ºN and 90ºS) and a longer time period (since the 1960s) (Liu et al., 2013b). The Trajectory-mapped Ozonesonde dataset for the Stratosphere and Troposphere (TOST) Version 1 (TOST-v1, Liu et al. 2013a, b) is available from 1965-2012 at the World Ozone and UV Data Centre (WOUDC, https://woudc.org/archive/products/ozone/vertical-ozone-profile/ozonesonde/1.0/tost/, last access: Sept 28, 2024), and has been successfully applied in model evaluation (Skeie et al., 2020; Badia et al., 2021), ozone and climate trend studies (Polvani et al., 2017; Gaudel et al., 2018; Gulev et al., 2021), as a background ozone climatology (Xu et al., 2018; Moeini et al., 2020), and for tropospheric ozone burden estimation (Griffiths et al., 2021). For users' convenience, the remaining gaps after trajectory mapping were further filled with a linear combination of spherical functions and provided as "smoothed" data in TOST-v1. Yet, the smoothed data should be used with caution; otherwise, misinterpretation of the smoothed data can be problematic (Chipperfield et al., 2022).

There have been several important developments since the publication of the first version of TOST data in 2013 (TOST-v1, Liu et al. 2013a, b). An improved version of TOST, namely TOST-v2, is necessary for the following reasons. Firstly, there are some 50,000 new ozone profiles, many from newly established ozonesonde stations (see Section 2.1). These new ozonesonde data permit updating TOST, providing 3-dimensional ozone data with larger areal coverages and longer periods up to 2021. Secondly, data from many ozonesonde stations have been updated to higher-quality versions. An important source of uncertainty in TOST-v1 is possible biases in station records due to instrument changes and/or changes in operating procedures. Homogenized time series are now available from the Harmonization and Evaluation

126 of Ground Based Instruments for Free Tropospheric Ozone Measurements (HEGIFTOM) project
127 for over 40 ozonesonde stations (Table S1). For these records, biases due to instrument changes,
128 sensing solution, and preparation changes have been corrected, to reduce the overall uncertainty
129 from 10-20% to 5-10% (Smit and Thompson, 2021). This effort to improve data quality also
130 uncovered an apparent change of bias at stations flying one type of sonde (Stauffer et al., 2020;
131 2022); 14 global ozonesonde stations (the bolded stations in Table S1) have shown an apparent
132 drop-off of 2-4 % in stratospheric ozone and total ozone column since circa 2013, due to a
133 possible instrument artifact. This is the subject of ongoing research (e.g.
134 https://gml.noaa.gov/annualconference/abstracts/78-230424-A.pdf, last access: Sept 28, 2024).
135 For these stations, ozone measurements above 40 hPa (~20 km) are not recommended for trend
136 calculations. We need therefore to exclude data above 40 hPa for the affected profiles in
137 constructing TOST. Thirdly, version 4.9 of the Hybrid Single-Particle Lagrangian Integrated
138 Trajectory (HYSPLIT) model (Draxler and Hess, 1998) used for trajectory simulation has been
139 improved and updated to version 5.2. Here, we address the mentioned issues and construct an
140 improved and updated TOST using the most state-of-the-art HYSPLIT and the most updated
141 ozonesonde data. Fourthly, while Liu et al. (2013a, b) validated TOST-v1 with ozonesonde data
142 at 20 selected stations, TOST-v2 is validated against the ozonesonde data at all 141 stations
143 individually with the trajectory-mapped approach omitting the input from the station being tested.
144 In addition, comparisons are made with the IAGOS measurements in the troposphere, and with
145 two limb-viewing satellite measurements, the Satellite Aerosol and Gas Experiment (SAGE) and
146 the Microwave Limb Sounder (MLS), in the stratosphere. This more comprehensive validation
147 and associated uncertainty analysis demonstrates the improved quality of TOST-v2, and also
148 provides some caveats for users of TOST.
149 In the following, Section 2 describes the datasets, including ozonesonde, satellite, and

aircraft data and the trajectory-mapping methodology. Section 3 shows the validation results, comparisons with satellite and aircraft observations, a summary of uncertainties in TOST-v2, and improvements in TOST-v2. Based on TOST-v2, we characterize global ozone variations in the troposphere and stratosphere, and show stagnant ozone variation in the lower stratosphere since the late 1990s in Section 4. Finally, conclusions are drawn in Section 5.

2. Data and Methods

2.1 Ozonesonde data

Ozonesonde data over 1970-2021 at 141 ozonesonde stations worldwide (Figure 1) were downloaded from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC, https://woudc.org/archive/Archive-NewFormat/OzoneSonde_1.0_1/), or where available, homogenized data from Southern Hemisphere ADditional OZonesondes (SHADOZ, https://doi.org/10.57721/SHADOZ-V06, last access: Sept 28, 2024) and HEGIFTOM (https://hegiftom.meteo.be/datasets/ozonesondes, last access: Sept 28, 2024). The homogenized ozonesonde stations from HEGIFTOM include ozonesonde stations from the SHADOZ network (Thompson et al., 2017; Witte et al., 2017; 2018), the Canadian network (Tarasick et al., 2016), the US network (Sterling et al., 2018), the Network for the Detection of Atmospheric Composition Change (NDACC) and several individual stations (Van Malderen et al., 2016; Witte et al., 2019; Ancellet et al., 2022), with an overall accuracy of 3-5% in both the stratosphere and troposphere. Ozonesonde data from the Beijing Nanjiao Meteorological Observatory (116.47°E, 39.81°N) in Beijing, China, are provided by the Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences. The ozone profiles at Beijing are measured by the Brewer-Mast type GPSO3 ozonesonde and the IAP electrochemical concentration cell (ECC) ozonesonde, which are in fair agreement with commercial ECC ozonesondes (Wang et al., 2003; Xuan et al., 2004;

Bian et al., 2007) in both laboratory and field experiments (Zhang et al., 2021; Zeng et al., 2023). In total, data from 43 more stations were used in TOST-v2 than in TOST-v1 (Liu et al., 2013b).

Figure 1a provides an overview of the distribution of the ozonesonde stations, the number of profiles, and the beginning year for every station. Most of the stations with data before the 1980s are located in North America, Europe, and East Asia. The majority of the stations in the Southern Hemisphere started measurement in the 1990s or later, and so the Southern Hemisphere contains a smaller number of ozone profiles than in the Northern Hemisphere. Figure 1b shows that the total number of ozonesonde profiles per year has almost doubled since the 1990s and reached a maximum in the late 2000s with over 3000 profiles per year. Since then, the amount of ozonesonde profiles available on the WOUDC site has declined slightly to 2000-3000 profiles per year. The average annual number of profiles per station slightly increased since the 1990s and has stabilized at about 40 profiles per year.

All the ozonesonde profiles were processed into 1-km vertical resolution by integrating and averaging the ozone mixing ratio in 1-km layers from the sea level. The ozonesonde data above 26 km were excluded as the data above this height show large uncertainties at mid- and high-latitudes (Fioletov et al., 2006).

2.2 Trajectory simulation

Forward and backward trajectories in four days were simulated every 6 hours using the version 5.2 HYSPLIT model (Stein et al., 2015). HYSPLIT was driven by the reanalysis of hourly meteorological data from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR), which has a horizontal resolution of 2.5° by 2.5° in latitude and longitude and 17 vertical levels from the surface to 10 hPa (Kalney et al., 1996). The length of the trajectories influences the spatial coverage and accuracy of the ozone

198 mapping. Generally, uncertainties increase rapidly along the trajectories, with typical errors of
199 about 100–200 km day^{-1} (Stohl, 1998). Trajectories have horizontal uncertainties of 350–400 km
200 after 3 days and 600-1000 km after 4 days in the Northern Hemisphere (Engström and
201 Magnusson, 2009). Trajectories show typical vertical deviations of about 200, 800, and 1000 m
202 after 2, 4, and 6 days in the stratosphere, and even greater uncertainties in the troposphere (Stohl
203 and Seibert, 1998). Therefore, to limit trajectory errors, 4-day trajectories were used herein,
204 following previous studies (Tarasick et al., 2010; Liu et al., 2013 a, b).
205
206 **2.3 Three-dimensional ozone mapping based on ozonesonde profiles and trajectories**
207 Ozone mixing ratios from each sounding at the 26 levels were assigned to the corresponding
208 forward and backward trajectory paths. These ozone values at positions every 6 hours along the
209 4-day backward and forward trajectories (32 positions for each level for both forward and
210 backward trajectories) were averaged in bins of 5º latitude and 5º longitude, for each 1-km
211 altitude for every month. This bin size corresponds both to the typical uncertainties of 4-day
212 trajectories discussed above, and to the typical ozone correlation length (500-1500 km) in the
213 troposphere and the stratosphere (Liu et al., 2009). To produce the mapping with pressure
214 altitudes, we also averaged the 4-day backward and forward trajectories in bins of 5º latitude and
215 5º longitude for every month, using the pressure altitudes generated by HYSPLIT trajectories.
216 The 26 pressure altitudes are 950, 850, 750, 650, 550, 450, 400, 350, 300, 250, 225, 200, 175,
217 150, 125, 100, 90, 80, 70, 60, 50, 40, 35, 30, 25, 20 hPa, following the pressure coordinates in
218 European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA).
219 Ozonesonde profiles in both the troposphere and stratosphere were used. In addition, the
220 ozonesonde data at the 26 levels are separated into the troposphere and stratosphere from the
221 measured ozonesonde temperature profile, following the WMO definition of the tropopause. In

this way, two datasets are constructed using ozonesonde data in the troposphere (troposphere-only) and in the stratosphere (stratosphere-only). The troposphere-only and stratosphere-only fields are helpful to calculate the tropopause in modeling studies, as the average tropopause height is usually not specified in the full-atmosphere field.

Based on this mapping, TOST-v2 was constructed at 26 altitude levels in two altitude coordinates (by geometric levels and pressure levels), from two altitude starting levels (altitude above sea level and altitude above ground level), for three temporal resolutions (in the seasonal mean for each year, the annual means for each year from 1970 to 2021, and monthly means for each decade from the 1970s to the 2010s) and with three types of data fields (trop-stra, troposphere-only, and stratosphere-only) for users' convenience (Table 1). In TOST-v2, we also generated the corresponding datasets that show ozone variation at 3 percentile levels (25, 50 and 75th). Examples presented in this paper all use TOST at geometric coordinates with altitudes above sea level. For this coordinate system, both ozonesonde profiles and mapped data necessarily begin at the altitude of the surface, leaving the levels below the topography of the Earth's surface as null if the levels are above the sea level.

Errors in the mapped data can come from trajectory errors, and from ignoring ozone chemistry (production and loss) along the transport pathway and deposition in the surface layer (Liu et al., 2013a). Differences between the results of backward and forward trajectory mapping can provide a measure of these errors, since in the absence of such errors the results of forward-only and backward-only trajectory mapping should be identical. Therefore, mappings from the forward-only and backward-only trajectories were compared as an initial quality check. Figure S1 shows monthly means (January and July) in 2000 at 3-4 km and 19-20 km, for forward-only and backward-only mapping. In general, the differences between the two mappings are commonly less than 15% and have no distinct pattern, indicating that trajectory errors are not

246 dependent on the direction of the trajectories. These modest differences between forward-only
247 and backward-only trajectory-mapped ozone fields also validate the reliability of this trajectory-
248 mapping method; both backward and forward trajectories, therefore, were combined in TOST to
249 achieve better averages and higher spatial coverage.
250
251 **2.4 Validations of TOST**
252 To comprehensively validate TOST, we compare TOST with actual ozonesonde, satellite,
253 and aircraft observations. Multiple metrics were used to indicate the level of agreement between
254 the TOST and other data. We used correlation coefficient (R) to present the agreement of the two
255 compared data, and linear fitting coefficient with the intercept set to 0 to show the overall
256 tendency of overestimation or underestimation. We also used relative difference (RD) to
257 represent the relative difference between the two compared data, and used bias and root mean
258 square error/difference (RMS) to show the absolute difference between the two compared data.
259 Details of the metrics can be found in Section S1.
260
261 **2.4.1 Ozonesonde profiles for validation**
262 We first validate TOST by comparing the actual ozone profile at each of the ozonesonde
263 stations with the trajectory-derived ozone profile for that station without the input of that station
264 itself. This method is computationally intensive, as the trajectory mapping must be re-calculated
265 (with data for all stations except one), for each ozonesonde station, but it directly tests the
266 reliability of deriving ozone concentrations at a location by integrating the contributions via
267 trajectories from surrounding sites, which is the essential assumption of the trajectory-mapping
268 method. We refer to this set of data that selectively excludes the local data at each station as
269 "Traj-derived" throughout this paper.

2.4.2 Satellite ozone profile for comparison with TOST in the stratosphere

TOST is further compared with two well-known satellite limb sounder datasets, the Satellite Aerosol and Gas Experiment (SAGE) and the Microwave Limb Sounder (MLS).

SAGE II was launched into a 57-degree inclination orbit on board Earth Radiation Budget Satellite (ERBS), and was in operation from 1984–2005. Using the highly accurate solar occultation technique, SAGE can resolve ozone vertical variation in the stratosphere and the middle-upper troposphere at 1-km vertical resolution (Kent et al., 1993), with the highest accuracy over the 20-45 altitudes (Cunnold et al., 1996). Here we use the Version 7.0 SAGE II ozone mixing ratio (https://sage.nasa.gov/missions/about-sage-ii/, last access: Sept 28, 2024) in the 1980s and 1990s for the comparison.

The MLS, onboard the Aura satellite, can measure stratospheric ozone profiles with a vertical resolution of about 3 km. MLS observes microwave radiances that are both emitted and absorbed by the atmosphere. The retrieval is more complex, and uses the optimal estimation approach. Here we use the Version 5.0 MLS ozone mixing ratio (https://disc.gsfc.nasa.gov/datasets/ML2O3_005/summary?keywords=ML2O3_005, last access: Sept 28, 2024) in the 2000s and 2010s for the comparison.

2.4.3 Aircraft ozone profiles for comparison with TOST in the troposphere

The IAGOS network (https://www.iagos.org/, last access: Sept 28, 2024) has been measuring ozone profiles worldwide since 1994 via dual-beam ultraviolet absorption monitors onboard commercial aircraft (Petzold et al., 2015), with an accuracy of about ± (2 nmol mol^{-1} + 2%) (Nédélec et al., 2015). Ozone monitors are calibrated annually to a reference analyser at the Bureau Internationale des Poids et Mesures (BIPM), and also compared every 2 hours to an in-

294 flight ozone calibration source. Generally good agreement is found between IAGOS profiles and
295 ozonesondes, with positive biases for the sondes of 5-10% (Tilmes et al., 2012; Zbinden et al.,
296 2013; Staufer et al., 2013, 2014; Tanimoto et al., 2015; Tarasick et al., 2019b), making IAGOS
297 ozone suitable for the validation of TOST. Here, the IAGOS ozone profiles were processed into 1
298 km layers from sea level and averaged into bins of 5º latitude and 5º longitude for each month. In
299 total, all IAGOS ozone data from 310 airports were used for the comparison (Table S2). Then,
300 the processed IAGOS ozone profiles were matched with the TOST ozone for the corresponding
301 grids to examine the performance of TOST in the troposphere.
302
303 **3. Validations and comparisons of TOST**
304 **3.1 Validations with ozonesonde observations**
305 First, we show the overall comparison in monthly mean ozone concentrations between
306 ozonesonde and trajectory-derived values without the inputs of the stations being tested (Traj-
307 Derived), from all the existing stations at all altitude levels. Note that the actual TOST dataset
308 would be better than "Traj-Derived ozone", especially at the sampling locations because the
309 input of the local station is included in the TOST dataset. Because of the large range of ozone
310 concentrations in the troposphere and stratosphere (0-6000 ppb), we divide the altitude levels
311 into three to present the overall accuracy of TOST in the lower troposphere (ozone
312 concentrations below 50 ppb), the upper troposphere (ozone concentrations between 50 and150
313 ppb) and the stratosphere (ozone concentrations over 150 ppb).
314 Figure 2a-f shows the overall ozone comparisons between ozonesonde (Sonde-Observed)
315 and Traj-Derived ozone in the entire study period (Figure 2a-c) and each decade (Figures 2d-f).
316 Each dot in Figure 2a-c represents the paired ozone concentrations from Traj-Derived and
317 Sonde-Observed values in each month at each latitude-longitude-altitude grid-cell, and the color

318 indicates the density of the dots. Overall, in the lower troposphere (Figure 2a), the Sonde-
319 Observed and Traj-Derived ozone concentrations agree well, with an R of 0.69 and an RMS of
320 7.5 ppb, a low bias (0.7 ppb) and RD (1.8%). The linear fit for the entire study period shows a
321 slope of 0.99. In the upper troposphere (Figure 2b), the agreement between the Sonde-Observed
322 and Traj-Derived ozone concentration is moderately lower, with a linear fitting coefficient of
323 1.01 and RMS of 21.1 ppb, and higher bias (2.9 ppb) and RD (4.0%) than those in the lower
324 troposphere. This lower agreement in the upper troposphere owes to the greater influence of
325 stratosphere-to-troposphere (STE) in the upper troposphere, where trajectories by the Lagrangian
326 dispersion model (such as HYSPLIT) show substantially increased deviations due to the strong
327 turbulence and convection (Stohl et al., 2002). The positive bias may imply that STE is slightly
328 overestimated in HYSPLIT. In the stratosphere (Figure 2c), the overall agreement between the
329 Sonde-Observed and Traj-Derived ozone concentrations has a linear fitting coefficient of 0.97
330 and an RMS of 416.9 ppb. The RD is only 0.5%, indicating higher reliability of Traj-Derived in
331 the stratosphere.

332 This validation method compares ozonesonde station data with Traj-Derived ozone, i.e., the
333 ozone found by averaging trajectories that come from other stations. Before the 1990s, RDs were
334 smaller than 0 and Rs were smaller than 0.60 (Figure 2d and 2e) in the lower troposphere,
335 indicating a tendency to underestimate the Traj-Derived ozone in the lower troposphere. After the
336 1990s, owing to the additional ozonesonde measurements provided by SHADOZ in the tropics,
337 the underestimation of Traj-Derived ozone in the lower troposphere is greatly reduced and the Rs
338 increased to > 0.71 (Figure 2d and 2e). Similarly, with the additional ozonesonde measurements
339 after the 1990s, the Rs in the upper troposphere increased from < 0.50 to > 0.58 (Figure 2d). In
340 all decades, the agreement between Sonde-Observed and Traj-Derived ozone in the stratosphere
341 is the best, with Rs of ~0.97. The RD in each decade is small (-0.3% - 1.4%), indicating no

342 systematic underestimation or overestimation in the stratospheric Traj-Derived data. However, in
343 the upper troposphere, ozone concentrations are slightly overestimated by Traj-Derived ozone
344 data concentrations, with RD of 0.6-4.5%.

345 Figure 3 examines how the RD between the ozonesondes and Traj-Derived ozone values
346 varies with altitude, presenting the frequency distributions of RD across all stations, at every
347 other altitude level and in each decade. The RD distributions are based on the monthly ozone
348 concentration difference between the actual ozonesonde and Traj-Derived data from all the
349 existing stations at the corresponding altitude level and decade. The distributions of RD show
350 little skewness in these altitudes and decades, indicating no systematic bias during the study
351 period. The overall interquartile ranges (25-75th), denoted by thick red lines with the widths
352 given in red values, indicate that RD is between -30 to 30% and the widths of interquartile ranges
353 are between 9 to 61%, with the lowest interquartile ranges (-10 to 10%) and widths (9 to 30%) of
354 RD in the stratosphere and lower-middle troposphere. Higher interquartile ranges of RD appear
355 in the 13-19 km altitude range, where the upper troposphere-lower stratosphere (UTLS) region is
356 located, and are due to the large vertical gradients of ozone concentrations in the UTLS and the
357 variability of the tropopause (Millan et al., 2023). The surface (boundary layer) ozone, however,
358 shows a positive bias of the median of up to 12%, in all decades, suggesting that TOST, which
359 neglects ozone chemistry and deposition processes, tends to overestimate ozone concentrations
360 there.

361 Figure 4 exemplifies comparisons in vertical profiles between Sonde-Observed and Traj-
362 Derived ozone profiles at individual stations in different seasons. Four stations with sufficient
363 data coverage (>15 years) were selected from the Antarctic coastal region (Syowa), Europe
364 (Hohenpeissenberg), North America (Boulder), and East Asia (Beijing). The decadal mean
365 (1990s and 2000s) profiles in January and July are used to compare the performance of Traj-

Derived ozone profiles in boreal winter and summer. In general, the Traj-Derived profiles can capture the vertical ozone variation in different seasons, with good correlation ($R > 0.99$) and high accuracy (bias < 100 ppb, RD < 10%) in comparison to the actual ozonesonde profiles. The comparison at Syowa shows a larger bias, but much of this is due to the fact that in the 1990s this station launched the Japanese KC-79 carbon-iodine sonde, while other stations in the Southern Hemisphere launched ECC sondes; the Traj-Derived profiles would therefore be expected to be 10-20% higher in the troposphere and about 5% higher in the lower stratosphere (Smit and Kley, 1998). The excellent agreement in tropospheric ozone at Hohenpeissenberg is likely due to frequent and dense European ozonesonde observations; similar cases also are seen at Uccle, Payerne, and Praha. Larger discrepancies are shown near the planetary boundary layer (PBL) and UTLS, as the simulated trajectories over these regions have more uncertainties (Stohl and Seibert, 1998; Sicard et al., 2019), and ozone chemistry and deposition are potentially important in the PBL at time scales similar to that of the longer trajectories (four days) (Prather et al., 2023; Prather and Zhu, 2024).

3.2 Comparisons with satellite observations in the stratosphere

To compare with satellite data, we first validated the Traj-Derived ozone profiles against ozonesonde measurements. The corresponding validation was conducted for the satellite data of SAGE and MLS in the same period, location, and altitude. The sets of ozonesonde, Traj-Derived and satellite data were selected only when all three datasets were available in the same month, decade, and grid cell, so to ensure that both the Traj-Derived and satellite data could be independently evaluated by the ozonesondes. As MLS data at altitudes below 261 hPa are not recommended (Livesey et al., 2022), we only compare both satellite datasets above this altitude, i.e., ~16-26 km. Figure 5 shows the timeseries of the vertical variation of monthly RD from 16-

26 km between Traj-Derived and SAGE ozone profiles from 1985-2005 (Figure 5c and 5f), and between Traj-Derived and MLS ozone from 2005-2019 (Figures 5i and 5l). SAGE ozone data are reliable above 20 km (Kremser et al., 2020), having a mean RD of about -10-10% above this altitude, similar to that of Traj-Derived ozone. Between 16 and 20 km, SAGE ozone concentrations are lower than the Traj-Derived ozone by 5 to 10% (Figure 5f), as Traj-Derived ozone overestimates the ozonesondes by 9 to 15% (Figure 5e). Over the MLS period from 2005 to 2019, TOST ozone at all altitudes between 16 and 26 km agrees with actual ozonesondes better than during the SAGE period (Figures 5h and 5k vs. Figures 5b and 5e). Accordingly, the Traj-Derived ozone concentrations show good agreement with MLS ozone above 22 km, but are lower than MLS ozone below 20 km (Figures 5i and 5l), as MLS generally overestimates ozone concentrations below 20 km (Figures 5g and 5j).

Figure S3 compares the RMS of Traj-Derived and satellite ozone in different latitude zones from 16-26 km. Compared to SAGE in the 1990s, the Traj-Derived ozone has comparable RMSs in the Northern Hemisphere, yet higher RMSs in the Southern Hemisphere, due to the fewer ozonesonde stations there. MLS ozone also shows lower RMSs in the Southern Hemisphere, but higher RMSs in the Northern Hemisphere.

Table S3 summarizes the evaluation of both Traj-Derived and satellite ozone against the ozonesondes over 16-26 km. The Traj-Derived and SAGE ozone values show a high correlation ($R = 0.95$ or greater in all cases) with actual ozonesondes. The Traj-Derived ozone data show RDs of -1% to +2% in the 1980s and 1990s, and just -0.3% to +0.4% in the 2000s and 2010s, while the SAGE ozone data shows RDs of -4% to +0.5% and the MLS data shows RDs of -2% to +11%.

It is expected that TOST would outperform satellite instruments in measurements below the tropopause, as satellite measurements are hampered by the large stratospheric ozone burden that

satellite instruments must look through. Yet, our comparisons suggest that even above 15 km, where SAGE and MLS are considered most reliable (Wang et al., 2002; Kremser et al., 2020; Livesey et al., 2022), TOST can provide comparable or better accuracy.

3.3 Comparisons with aircraft observations in the troposphere

We compare TOST ozone with the IAGOS dataset in the lower troposphere from 1994-2021 (Figure 6). Note that this comparison is between the full TOST (not Traj-derived) and IAGOS datasets here. TOST ozone values are generally higher than IAGOS with a mean bias of 2.2 ppb and R of 0.49, but RDs (5.8%) and RMS (8.8 ppb) are low. The linear fit has a slope of 1.03. The two ozone datasets employ different measurement techniques and atmospheric sampling (Petetin et al., 2018). Previous studies have reported that IAGOS ozone values are systematically lower than ozonesonde values, typically by 5-10% in the troposphere (Tilmes et al., 2012; Zbinden et al., 2013; Staufer et al., 2013, 2014; Tanimoto et al., 2015; Tarasick et al., 2019b). The comparisons in Figure 6 are consistent with these earlier estimates, as the RD indicates that IAGOS measurements average 6% lower than TOST.

3.4 Uncertainty analysis

As noted in Section 2.3, the ozone concentrations in a grid in a month are determined by the ozone concentrations along all the trajectories passing through that grid cell in that month. Therefore, an estimate of the random uncertainty of TOST may be obtained from the standard error of the mean in each grid cell. Note that this may not be a true estimate of the standard error, as some cells may contain more than one value from an individual trajectory so these values are not independent and the standard error calculation is biased low.

For convenience, given the large range of ozone concentrations between the stratosphere and

438 troposphere, we use the ratio of the standard error to the mean in each of the grid cells, SE/Mean,
439 expressed in %, to estimate spatial patterns of the uncertainty. The standard error is proportional
440 to the variability of the ozone values in a grid cell (i.e. the standard deviation) and inversely
441 proportional to the square root of the number of data values. Thus in general, the more
442 trajectories passing a grid cell, the more data samples for that cell and the lower the standard
443 error for that cell. For each cell, we also calculated the number of independent samples, i.e., the
444 trajectory originated from a single ozonesonde altitude was counted only once in a grid cell when
445 the trajectory passes that cell regardless of how long the trajectory stays in that cell. Figure 7
446 shows the SE/Mean and the number of independent samples in January and July of the 2000s at
447 3-4 km and 19-20 km. Generally, the Southern Hemisphere shows higher SE/Mean values (>
448 10%) than the Northern Hemisphere (< 6%), which reflects the large number (>100) of ozone
449 soundings in the Northern Hemisphere, especially over North America and Europe. However,
450 near the equator, despite the higher sampling rate, the SE/Mean still is as high as 15%. Compared
451 to the stratospheric level (19-20 km), the tropospheric level (3-4 km) shows an overall higher
452 SE/Mean. SE/Mean varies less with season in the stratosphere than in the troposphere. For
453 example, at 3-4 km, the SE/Mean in January is generally <7% but becomes >10% in July in the
454 Northern Hemisphere, and vice versa in the Southern Hemisphere. This is likely due to more
455 vertical motion in the PBL (Stohl and Seibert, 1998; Sicard et al., 2019) so that ozone in some
456 bins comes from multiple altitude levels, as well as increased photochemistry and biomass
457 burning. Stratospheric intrusions to the lower troposphere are more frequent in boreal spring and
458 summer than in winter (Terao et al., 2008; Greenslade et al., 2017), and can be responsible for
459 much of the variability at 3-4 km (Tarasick et al., 2019a).

460 To provide an overview of the uncertainties of TOST in different altitudinal and latitudinal
461 zones, as well as in different seasons and decades, we calculated the RD of the monthly ozone

462 mixing ratio between ozonesonde and Traj-Derived ozone over 1970-2021 (Figure 8). Among
463 altitudes, the highest RD values appear at 9-10 km over the tropopause region, and the second
464 highest RD at the surface, while the lowest RD values are in the lower-middle troposphere (3-6
465 km) and stratosphere (19-26 km), consistent with Figure 3. By season (Figure 8b), the RD varies
466 slightly with a lower value in JJA and SON than in other seasons. There is considerable variation
467 in RDs with latitude (Figure 8c); the RDs in the southern high latitudes (90°-60°S) and the
468 northern tropics (0°-30°N) are higher than in other latitudinal zones. This could reflect higher
469 horizontal gradients of ozone (e.g. stations in or outside the ozone hole) in the southern high
470 latitudes or biases between ECC sondes and other types (the Indian and Japanese sondes) in the
471 northern tropics. After the 1990s, the RDs are reduced markedly compared to the 1980s and
472 1970s (Figure 8d), likely related to the improved data coverage in the later periods. This
473 overview provides caveats regarding where (surface and UTLS, the northern high latitudes and
474 tropics) and when (before the 1990s) more caution is advised when using TOST.
475
476 **3.5 Improvements in the new version**
477 The improvements in TOST-v2 are attributed to the increased amount and improved quality
478 of ozonesonde data, as well as the improved trajectory simulation and ozone mapping. Because
479 more ozonesonde stations and more ozonesonde data have become available since the 1990s or
480 2000s (Table S1), more ozone profiles were used in constructing TOST-v2, leading to improved
481 data density. Table S4 summarizes the data coverage, the number of ozonesonde stations and
482 ozonesonde profiles used for TOST-v2 and TOST-v1. The data coverage is defined as the ratio of
483 the number of grid cells with valid annual means to the total number of grid cells in the
484 corresponding latitudinal zone. The number of ozonesonde stations, compared to Liu et al.
485 (2013b), increases in all latitudes by ~50%, and the total number of ozonesonde profiles used is

doubled. Data coverage increases as well, in all latitude bands, by 5-15% (Table S4) and in all altitudes by a maximum of 10% (Figure S4).

In addition to the data density, the data quality was also improved in TOST-v2. Figure 9a-b shows the distributions of ozone concentrations in TOST-v2 and TOST-v1 at the lowest level (0-1 km) for the 2000s. Over the Antarctic, gaps are observed only in TOST-v2. This is more reasonable for the sea-level data because the altitude over the Antarctic is over 1 km (Figure S5a), where trajectories should not appear at 0-1 km. Therefore, the spatial distributions of ozone are clearly improved with this topography correction in TOST-v2, which could be attributed to the updated terrain file since HYSPLIT v5.0 (https://www.arl.noaa.gov/hysplit/hysplit-model-updates/). Over the eastern Pacific, marked with an ellipse in Figure 9a, b, TOST-v1 shows higher ozone concentrations than TOST-v2 by 30% (Figure 9c). Compared to the ozonesonde measurement at 0-1 km in the 2000s in these two regions (Davis station for the Antarctic and Easter Island station for the eastern Pacific), TOST-v2 agrees better with ozonesondes than TOST-v1, indicating better representation of ozone distributions (Figure S5b).

With reference to spatial distributions at 19-20 km in the 2000s, Figure 9d-e shows that in the Antarctic and the tropical eastern Pacific, TOST-v1 values show higher concentrations than TOST-v2 (Figure 9f). Figure S5c compares ozone concentrations from ozonesonde, TOST-v2, and TOST-v1 at 19-20 km in the 2000s at an Antarctic station (Syowa) and a tropical station (Bogota). Compared to TOST-v1, TOST-v2 ozone values show a better agreement with the ozonesonde measurement. The difference between TOST-v2 ozone and ozonesonde measurements is 10% and 29% in Syowa and Bogota stations, while in TOST-v1, ozone concentrations at these stations show 24% and 39% differences (Figure S5c).

In summary, TOST has been improved in TOST-v2 with higher spatial coverage, improved description of ozone spatial distributions, and a better agreement with ozonesonde measurements

in both the troposphere and stratosphere. Furthermore, TOST-v2 provides additional information that shows ozone variations in 3 percentile levels (25, 50 and 75th). TOST-v2 is also generated in a pressure altitude coordinate for users' convenience.

4. Global ozone spatial-temporal variations observed from TOST

4.1 Ozone spatial variations in the troposphere and stratosphere

As a 3-dimensional ozone dataset, TOST can depict both horizontal and vertical ozone distributions, as well as long-term ozone variation. Figure 10 shows distributions of decadal mean TOST ozone at 3-4 km and 19-20 km in four seasons of the 2000s. At 3-4 km in the troposphere, ozone concentrations are higher over the continent in the Northern Hemisphere, especially in MAM and DJF (>50 ppb), reflecting the ozone production from the photochemical reactions of anthropogenic and natural emissions. In addition, the continental outflow from the southern US (in MAM) and the biomass-burning-produced ozone in southern Africa (in JJA and SON) are well captured and in agreement with satellite observations (Fishman et al., 1990; Ebojie et al., 2016). At 19-20 km in the stratosphere (Figure 10e-h), ozone concentrations are higher near the poles than in the tropics, due to the impact of the Brewer–Dobson circulation. The North Pole has higher ozone concentrations than the South Pole in DJF and MAM, and vice versa in JJA and SON, reflecting the seasonality of the Brewer-Dobson circulation. Also at 19-20 km, the ozone concentrations are lower over Asia in JJA (Figure 10f) than in other seasons, reflecting the transport of ozone by Asian summer monsoon from the tropics (Gettelman et al., 2004; Bian et al., 2020).

Although trajectory mapping fills in much of the global spatial domain, large gaps can still be found, particularly in the tropics, where ozone soundings are less dense. Since some applications require a default ozone value at all grid cells, a gap-filled and smoothed ozone

dataset is also provided for the decadal mean ozone in each month and the annual mean ozone, by fitting the maps at each level with a linear combination of spherical functions (Liu et al., 2013b). As shown in Figures 10i-p, small-scale variations and extreme values are reduced in the smoothed ozone fields, while broad patterns of the ozone distribution are retained, making these smoothed maps valuable for qualitative visualization of the spatial, seasonal, and decadal variations in ozone at different altitudes. They should, however, be used for any kind of quantitative analysis with great caution, as these data, where gaps exist in the unsmoothed TOST dataset, may be interpolated from the original measurement that is far in distance and thus the degree to which they represent the true ozone value should be carefully examined. For example, erroneous conclusions have been inferred from the smoothed TOST-v1 over the tropics, with very limited observations before 1998, where the smoothed data were mostly interpolated from higher latitudes (Chipperfield et al., 2022). In addition, smoothing, as noted, filters small-scale variations and extreme values, and retains ozone variations on large scales, which should be born in mind when using the smoothed data. The smoothed dataset has not been quantitatively evaluated in any way.

Figure 11a-d shows the latitude-altitude distribution of TOST ozone in each season averaged over 1970-2021. The steep changes in ozone concentration from <100 to >500 ppb in the vicinity of the tropopause (the black lines in Figure 11a-d, calculated from the NCEP/NCAR reanalysis) are well captured. Due to the Brewer-Dobson circulation, ozone concentrations above the tropopause increase with latitude from the tropics to the poles, which is also well reflected in the latitude-altitude distribution. TOST ozone concentrations around 12-13 km are higher in spring (600-800 ppb) than in the other seasons (< 500 ppb) over northern midlatitudes (45-60°N), which reflects the stronger Brewer-Dobson circulation in spring (Holton et al., 1995). Figure 11e shows the monthly mean TOST ozone time series from 1970 to 2021, averaged over 30-70°N at each

558 level. Clear seasonal cycles are well captured every year.

559

560 **4.2 Long-term trend in the lower stratospheric ozone**

561 One of the advantages of TOST is its long-term coverage, which enables investigation of
562 variations in ozone back to the 1970s. While analysis of long-term ozone trends in the
563 troposphere with TOST is underway in separate projects, we show here an application of TOST
564 data for studying lower stratospheric ozone changes. Following the implementation of the
565 Montreal Protocol and its amendments, recent studies have found an increase in upper
566 stratospheric ozone since the late 1990s (Chipperfield et al., 2017; Szeląg et al., 2020; Dunn et
567 al., 2023). However, the lower stratospheric ozone trend remains highly uncertain (Ball et al.,
568 2020). Quantifying lower stratospheric ozone trends depends largely on the quality of the
569 observational datasets (Li et al., 2023). While the trend is commonly analyzed with individual
570 ozonesonde time series, it is challenging to assess how well individual long-term station changes
571 represent regional or global variations. Combining data from sparse and widely separated
572 ozonesonde sites involves implicit assumptions about their representativeness. With
573 meteorological trajectory mapping, each original ozonesonde measurement is assigned a
574 trajectory which describes its representativeness, and the TOST averages are therefore weighted
575 according to the representativeness of each measurement. While this is subject to trajectory
576 errors and the fact that coverage is incomplete (Table S3), unless trajectory errors are non-
577 random, it should produce a better result than simple averaging of sonde station data by
578 geographic region.

579 Figure 12 shows the annual mean of ozone anomalies at 21-22 km and 24-25 km from 1970
580 to 2021; the area-weighted averages were taken over grid cells from 30°-70°N all with valid data
581 throughout all years, i.e. ~ 70% of grid cells in the latitudinal zone. The 3-year running means

582 are also shown with the time series. The ozone time series at both levels captures the clear ozone
583 depletion before the early 1990s and the slow recovery in the latter part of the 1990s. In addition,
584 these updated TOST time series show that stratospheric ozone since 2000 has changed little.
585 There are non-significant trends in the ozone concentrations at 21-22 km (0.5±0.6 %/decade) and
586 24-25 km (-0.2±0.9 %/decade) from 1998 to 2021, indicating little change of lower stratospheric
587 ozone, despite the fact that 25 years have passed since peak stratospheric chlorine. Recent studies
588 using merged satellite data suggested that the decrease in the lower stratospheric ozone is
589 offsetting the increase in the upper stratosphere (Ball et al., 2018, 2019; Szelag et al., 2020; Li et
590 al., 2023), which is responsible for the flat trend in the total column ozone since the late 1990s.
591 However, in the Northern Hemisphere mid-latitudes, TOST indicates no significant trend in the
592 lower stratospheric ozone after the late 1990s. The difference between the TOST and satellite-
593 based data calls for further in-depth studies on the stratospheric ozone trend, especially in the
594 lower stratosphere. Such observations of the variation in stratospheric ozone are essential to
595 verifying the expected stratospheric ozone recovery under the Montreal Protocol and to
596 understanding the feedbacks among dynamical, thermal, and ozone variability.
597

598 **5. Conclusions**

599 An improved TOST dataset has been constructed from 1970 to 2021 based on the updated
600 ozonesonde profiles at 141 ozonesonde stations from WOUDC, SHADOZ, HEGIFTOM and
601 NDACC (Table 1). The updated TOST was derived by combining the 4-day forward and
602 backward trajectories from each ozonesonde profile, which were driven by the most state-of-the-
603 art HYSPLIT model (v5.2) and NCEP reanalysis data (NNRP-1). Similar to TOST-v1, ozone
604 concentrations in each season, in each year (1970-2021) and in each month of a decade (January
605 to December from the 1970s to 2010s) are provided in 3-dimensional grids of 5º × 5º × 1 km

606 (latitude, longitude, and altitude). For users' convenience, the corresponding spatially-smoothed
607 datasets are also provided for qualitative visualization, model initialization, and other
608 applications. TOST is provided in geometric coordinates from both sea level and ground level at
609 26 layers; separate ozone climatology datasets are generated using ozonesonde profiles in both
610 the troposphere and stratosphere, only in the troposphere, and only in the stratosphere. Statistics
611 including standard error and independent number of samples are also provided. In TOST-v2, the
612 corresponding datasets are also provided in a pressure altitude coordinate. In addition to the
613 seasonal, annual, or decadal-monthly means, the corresponding datasets for ozone variations at 3
614 percentile levels (25, 50 and 75th) are also provided.

615 Comprehensive validation of TOST-v2 was conducted. At all the ozonesonde stations used,
616 trajectory-derived ozone profiles without the input of the station itself were compared with the
617 corresponding ozonesonde profiles at the stations. The overall comparison between the
618 ozonesonde and trajectory-derived ozone shows good agreement in both the troposphere (R =
619 0.56-0.69, RD = 2-4%) and stratosphere ($R = 0.97$, RD = 0.5%) in each decade and in all
620 decades' mean (Figure 2). The frequency distribution of RD at different altitudes shows
621 interquartile ranges of RD between -30 to 30%, with the lowest interquartile ranges of RD (-10
622 to 10%) in the stratosphere and lower troposphere, and no systematic bias except in the surface
623 layer (Figure 3). The patterns of ozone profiles at individual stations are also well captured and
624 quantified, with $R > 0.76$ and RD of 2-8% (Figure 4). Larger discrepancies are shown near the
625 PBL and UTLS, especially for coastal stations where the trajectory-derived ozone may be biased
626 by trajectories from the continent (Tarasick et al., 2010).

627 The comparison between TOST and satellite data, i.e., SAGE in the 1980s and 1990s, and
628 MLS in the 2000s and 2010s, illustrates that TOST data have comparable accuracy with the
629 satellite data in the stratosphere, while in the troposphere TOST is markedly superior (Figure 5).

In different latitude zones and decades, TOST performs comparably with SAGE and MLS data as well (Figure S3 and Table S3). TOST-v2 was also directly compared to MOZAIC-IAGOS ozone profiles over the period 1994-2021 from the surface to 5 km. Despite the systematic difference between MOZAIC-IAGOS and ozonesonde measurements, the two ozone datasets agree well in the monthly mean of 1994-2021 for the lower troposphere (RD =6%, Figure 6).The uncertainties of TOST are largely dependent on the availability of ozonesonde data. Higher uncertainties are found before the 1990s, as global coverage is sparse in the tropics before SHADOZ. Higher uncertainties also appear at southern high latitudes and in the northern tropics (Figure 8), likely because of greater ozone variability there, although biases between ozonesonde types may also contribute. TOST data at the PBL and UTLS have twice the RDs compared to other altitude levels; the former is due to more small-scale processes in the PBL while the latter is related to the large ozone gradient and the dynamic variation of the tropopause. In addition, the smoothed dataset should be used for quantitative analysis with great caution, as it has not been quantitatively evaluated in any way.

Compared to the previous version of TOST (TOST-v1, Liu et al. 2013a and b), TOST-v2 is mainly improved in two aspects. Firstly, the record is extended to 2021 and data coverage is increased by as much as 15%, as more ozone profiles and 43 additional ozonesonde stations are used in constructing the new version of TOST. Secondly, the spatial distribution of ozone has better agreement with ozonesonde measurements in both the troposphere and stratosphere over regions of Antarctica and the eastern Pacific, with RD decreased by > 50%. Here we suggest future work for TOST improvement in three ways: applying more sophisticated meteorological reanalysis data to generate trajectories, using varying trajectory lengths based on the lifetime of ozone at different altitudes, and filling remaining gaps through reliable and effective gap-filling approaches. Moreover, including a tropopause referenced ozone climatology in the TOST dataset

would help better comparisons with model data in the UTLS.

TOST can capture global ozone distributions in the troposphere and stratosphere (Figures 10 and 11), showing horizontal and vertical variations, the continental outflow, and the gradient of ozone concentrations near the tropopause. TOST can also reflect the seasonal variations in ozone concentrations near the vicinity of the tropopause. Over the Northern Hemispheric mid-latitudes, the time series of the updated TOST shows a stagnant recovery but an overall non-significant trend in lower stratospheric ozone after 1998 (Figure 12), which is different from the decreasing trend reported in satellite-based data (Ball et al., 2018, 2019; Szelag et al., 2020; Li et al., 2023). This finding suggests more in-depth studies of stratospheric ozone trends, especially in the lower stratosphere.

It is anticipated that TOST-v2 dataset can benefit future studies, owing to its long record, global coverage, and high vertical resolution. We expect that it will be a useful dataset for trend studies, especially in the free troposphere, and also in the stratosphere, given the excellent long-term stability of the global ozonesonde network (Stauffer et al., 2022). We caution, however, that users should keep in mind the assumptions and limitations of the data product as described here.

Author contribution

J. L. and D.T. conceptualized and designed this study. Z.Z. performed data process, analysis, and composed the first draft. Z. Z., J. L., and D.T. revised the manuscript with input from all the co-authors. All the coauthors contributed substantially to this study in making ozonesonde measurements, processing, calibrating, and archiving the ozonesonde data, and providing constructive and valuable suggestions to and comments on the manuscript. All the coauthors approved the submission of this paper.

Code and data availability

The ozonesonde data used in this study can be obtained from the WOUDC (https://woudc.org/archive/Archive-NewFormat/OzoneSonde_1.0_1/), SHADOZ (https://doi.org/10.57721/SHADOZ-V06) and HEGIFTOM (https://hegiftom.meteo.be/datasets/ozonesondes). The trajectory model HYSPLIT (Version 5.2) is from the NOAA Air Resources Laboratory (http://www.arl.noaa.gov/ready.html), driven by the NCEP/NCAR reanalysis data from the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, at https://www.ready.noaa.gov/data/archives/reanalysis/. The aircraft data can be accessed from IAGOS network (https://www.iagos.org/). The two satellite data for comparison, the SAGE II (Version 7.0) and the MLS (Version 5.0), are obtained from https://sage.nasa.gov/missions/about-sage-ii/ and https://disc.gsfc.nasa.gov/datasets/ML2O3_005/summary?keywords=ML2O3_005, respectively. We are in the process of making the TOST available on the WOUDC website. TOST data currently are available at https://doi.org/10.5281/zenodo.13984482.

Competing interests

The authors declare that they have no conflict of interest.

Special issue statement

This article is part of the special issue “Tropospheric Ozone Assessment Report Phase II (TOAR-II) Community Special Issue (ACP/AMT/BG/GMD inter-journal SI)”.

Acknowledgements

We thank many for their dedication to WOUDC, SHADOZ, and HEGIFTOM, making ozonesonde data accessible. We also thank SAGE II and MLS team for their ozone data for

comparison. We acknowledge the HYSPLIT team for the trajectory model. Z. Z. and J. L. acknowledge the financial support from Natural Science and Engineering Council of Canada (Grant No. RGPIN-2020-05163); J. B. and J. Z. from the National Natural Science Foundation of China (Grant No. 42293321). We are grateful to Drs. Michael Prather, Owen Cooper, and an anonymous reviewer for their constructive comments and suggestions, which helped improve this manuscript.

References

Ancellet, G. and Beekmann, M.: Ozone Monitoring and Measurements, in: Tropospheric Ozone Research: Tropospheric Ozone in the Regional and Sub-regional Context, Springer, 207-306, 1997.

Ancellet, G., Godin-Beekmann, S., Smit, H. G., Stauffer, R. M., Van Malderen, R., Bodichon, R., and Pazmiño, A.: Homogenization of the Observatoire de Haute Provence electrochemical concentration cell (ECC) ozonesonde data record: comparison with lidar and satellite observations, Atmospheric Measurement Techniques, 15, 3105-3120, 2022.

Badia, A., Iglesias-Suarez, F., Fernandez, R. P., Cuevas, C. A., Kinnison, D. E., Lamarque, J. F., Griffiths, P. T., Tarasick, D. W., Liu, J., and Saiz-Lopez, A.: The role of natural halogens in global tropospheric ozone chemistry and budget under different 21st century climate scenarios, Journal of Geophysical Research: Atmospheres, 126, e2021JD034859, 2021.

Ball, W. T., Chiodo, G., Abalos, M., Alsing, J., and Stenke, A.: Inconsistencies between chemistry–climate models and observed lower stratospheric ozone trends since 1998, Atmospheric chemistry and physics, 20, 9737-9752, 2020.

Ball, W. T., Alsing, J., Staehelin, J., Davis, S. M., Froidevaux, L., and Peter, T.: Stratospheric ozone trends for 1985–2018: sensitivity to recent large variability, Atmospheric Chemistry and Physics, 19, 12731-12748, 2019.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., and Anderson, J.: Evidence for a continuous decline in lower stratospheric ozone

offsetting ozone layer recovery, Atmospheric Chemistry and Physics, 18, 1379-1394, 2018.
Bhartia, P. K.: OMI Algorithm Theoretical Basis Document Volume II, OMI Ozone, 2002.
Bian, J., Gettelman, A., Chen, H., and Pan, L. L.: Validation of satellite ozone profile retrievals using Beijing ozonesonde data, Journal of Geophysical Research: Atmospheres, 112, 2007.
Bian, J., Li, D., Bai, Z., Li, Q., Lyu, D., and Zhou, X.: Transport of Asian surface pollutants to the global stratosphere from the Tibetan Plateau region during the Asian summer monsoon, National Science Review, 7, 516-533, 2020.
Bodeker, G. E., Nitzbon, J., Tradowsky, J. S., Kremser, S., Schwertheim, A., and Lewis, J.: A global total column ozone climate data record, Earth System Science Data, 13, 3885-3906, 2021.
Bognar, K., Tegtmeier, S., Bourassa, A., Roth, C., Warnock, T., Zawada, D., and Degenstein, D.: Stratospheric ozone trends for 1984–2021 in the SAGE II–OSIRIS–SAGE III/ISS composite dataset, Atmospheric Chemistry and Physics, 22, 9553-9569, 2022.
Chen, X., Liu, Y., Lai, A., Han, S., Fan, Q., Wang, X., Ling, Z., Huang, F., and Fan, S.: Factors dominating 3-dimensional ozone distribution during high tropospheric ozone period, Environmental Pollution, 232, 55-64, 2018.
Chipperfield, M. P., Bekki, S., Dhomse, S., Harris, N. R., Hassler, B., Hossaini, R., Steinbrecht, W., Thiéblemont, R., and Weber, M.: Detecting recovery of the stratospheric ozone layer, Nature, 549, 211-218, 2017.
Chipperfield, M. P., Chrysanthou, A., Damadeo, R., Dameris, M., Dhomse, S. S., Fioletov, V., Frith, S. M., Godin-Beekmann, S., Hassler, B., and Liu, J.: Comment on “Observation of large and all-season ozone losses over the tropics” [AIP Adv. 12, 075006 (2022)], AIP Advances, 12, 2022.
Chouza, F., Leblanc, T., Brewer, M., and Wang, P.: Upgrade and automation of the JPL Table Mountain Facility tropospheric ozone lidar (TMTOL) for near-ground ozone profiling and satellite validation, Atmospheric Measurement Techniques, 12, 569-583, 2019.
Colombi, N., Miyazaki, K., Bowman, K. W., Neu, J. L., and Jacob, D. J.: A new methodology for inferring surface ozone from multispectral satellite measurements, Environmental Research Letters, 16, 105005, 2021.
Cunnold, D., Wang, H., Chu, W., and Froidevaux, L.: Comparisons between Stratospheric Aerosol and Gas Experiment II and microwave limb sounder ozone measurements and

761 aliasing of SAGE II ozone trends in the lower stratosphere, Journal of Geophysical
762 Research: Atmospheres, 101, 10061-10075, 1996.
763 Draxler, R. R. and Hess, G.: An overview of the HYSPLIT_4 modelling system for trajectories,
764 Australian meteorological magazine, 47, 295-308, 1998.
765 Dunn, R. J., Miller, J. B., Willett, K. M., Gobron, N., Ades, M., Adler, R., Alexe, M., Allan, R. P.,
766 Anderson, J., and Anneville, O.: Global climate, Bulletin of the American Meteorological
767 Society, 104, S11-S145, 2023.
768 Ebojie, F., Burrows, J., Gebhardt, C., Ladstätter-Weißenmayer, A., Von Savigny, C., Rozanov, A.,
769 Weber, M., and Bovensmann, H.: Global tropospheric ozone variations from 2003 to 2011
770 as seen by SCIAMACHY, Atmospheric Chemistry and Physics, 16, 417-436, 2016.
771 Engström, A. and Magnusson, L.: Estimating trajectory uncertainties due to flow dependent
772 errors in the atmospheric analysis, Atmospheric Chemistry and Physics, 9, 8857-8867, 2009.
773 Eyring, V., Cionni, I., Bodeker, G. E., Charlton-Perez, A. J., Kinnison, D. E., Scinocca, J. F.,
774 Waugh, D. W., Akiyoshi, H., Bekki, S., and Chipperfield, M. P.: Multi-model assessment of
775 stratospheric ozone return dates and ozone recovery in CCMVal-2 models, Atmospheric
776 Chemistry and Physics, 10, 9451-9472, 2010.
777 Fioletov, V., Tarasick, D., and Petropavlovskikh, I.: Estimating ozone variability and instrument
778 uncertainties from SBUV (/2), ozonesonde, Umkehr, and SAGE II measurements:
779 Short-term variations, Journal of Geophysical Research: Atmospheres, 111, 2006.
780 Fishman, J., Watson, C. E., Larsen, J. C., and Logan, J. A.: Distribution of tropospheric ozone
781 determined from satellite data, Journal of Geophysical Research: Atmospheres, 95, 3599-
782 3617, 1990.
783 Fleming, Z. L., Doherty, R. M., Von Schneidemesser, E., Malley, C. S., Cooper, O. R., Pinto, J. P.,
784 Colette, A., Xu, X., Simpson, D., and Schultz, M. G.: Tropospheric Ozone Assessment
785 Report: Present-day ozone distribution and trends relevant to human health, Elem Sci Anth,
786 6, 12, 2018.
787 Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D.,
788 Mauritsen, T., Palmer, M., and Watanabe, M.: The Earth's energy budget, climate feedbacks,
789 and climate sensitivity, 2021.
790 Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C.,
791 Coheur, P.-F., Cuesta, J., and Cuevas, E.: Tropospheric Ozone Assessment Report: Present-

792 day distribution and trends of tropospheric ozone relevant to climate and global atmospheric
793 chemistry model evaluation, Elementa: Science of the Anthropocene, 6, 2018.
794 Gettelman, A., Kinnison, D. E., Dunkerton, T. J., and Brasseur, G. P.: Impact of monsoon
795 circulations on the upper troposphere and lower stratosphere, Journal of Geophysical
796 Research: Atmospheres, 109, 2004.
797 Greenslade, J. W., Alexander, S. P., Schofield, R., Fisher, J. A., and Klekociuk, A. K.:
798 Stratospheric ozone intrusion events and their impacts on tropospheric ozone in the
799 Southern Hemisphere, Atmospheric Chemistry and Physics, 17, 10269-10290, 2017.
800 Griffiths, P. T., Murray, L. T., Zeng, G., Shin, Y. M., Abraham, N. L., Archibald, A. T., Deushi,
801 M., Emmons, L. K., Galbally, I. E., Hassler, B. L.W. Horowitz, J. Keeble, J. Liu, O. Moeini,
802 V. Naik, F.M. O'Connor, D. Tarasick, S. Tilmes, S.T. Turnock, O. Wild, P.J. Young and P.
803 Zanis: Tropospheric ozone in CMIP6 simulations, Atmospheric Chemistry and Physics, 21,
804 4187-4218, 2021.
805 Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., Gong, D.,
806 Kaufman, D. S., Nnamchi, H. C., and Quaas, J.: Changing state of the climate system, 2021.
807 Han, H., Liu, J., Yuan, H., Wang, T., Zhuang, B., and Zhang, X.: Foreign influences on
808 tropospheric ozone over East Asia through global atmospheric transport, Atmospheric
809 Chemistry and Physics, 19, 12495-12514, 2019.
810 Harmens, H., Hayes, F., Mills, G., Sharps, K., Osborne, S., and Pleijel, H.: Wheat yield
811 responses to stomatal uptake of ozone: Peak vs rising background ozone conditions,
812 Atmospheric environment, 173, 1-5, 2018.
813 Hassler, B., Bodeker, G., and Dameris, M.: A new global database of trace gases and aerosols
814 from multiple sources of high vertical resolution measurements, Atmospheric Chemistry
815 and Physics, 8, 5403-5421, 2008.
816 Hassler, B., Kremser, S., Bodeker, G. E., Lewis, J., Nesbit, K., Davis, S. M., Chipperfield, M. P.,
817 Dhomse, S. S., and Dameris, M.: An updated version of a gap-free monthly mean zonal
818 mean ozone database, Earth System Science Data, 10, 1473-1490, 2018.
819 Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.:
820 Stratosphere-troposphere exchange, Reviews of Geophysics, 33, 403-439, 1995.
821 Huijnen, V., Miyazaki, K., Flemming, J., Inness, A., Sekiya, T., and Schultz, M. G.: An
822 intercomparison of tropospheric ozone reanalysis products from CAMS, CAMS interim,

TCR-1, and TCR-2, Geoscientific model development, 13, 1513-1544, 2020.

Kalney, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437-472, 1996.

Kent, G., Winker, D., Osborn, M., and Skeens, K.: A model for the separation of cloud and aerosol in SAGE II occultation data, Journal of Geophysical Research: Atmospheres, 98, 20725-20735, 1993.

Keppens, A., Lambert, J.-C., Granville, J., Miles, G., Siddans, R., van Peet, J. C., Van Der A, R., Hubert, D., Verhoelst, T., and Delcloo, A.: Round-robin evaluation of nadir ozone proFIle retrievals: Methodology and application to MetOp-A GOME-2, Atmospheric Measurement Techniques, 8, 2093-2120, 2015.

Kremser, S., Thomason, L. W., and Bird, L. J.: Simplified SAGE II ozone data usage rules, Earth System Science Data, 12, 1419-1435, 2020.

Li, Y., Dhomse, S. S., Chipperfield, M. P., Feng, W., Bian, J., Xia, Y., and Guo, D.: Quantifying stratospheric ozone trends over 1984–2020: a comparison of ordinary and regularized multivariate regression models, Atmospheric Chemistry and Physics, 23, 13029-13047, 2023.

Liu, G., Tarasick, D. W., Fioletov, V. E., Sioris, C. E., and Rochon, Y. J.: Ozone correlation lengths and measurement uncertainties from analysis of historical ozonesonde data in North America and Europe, Journal of Geophysical Research: Atmospheres, 114, 2009.

Liu, G., Liu, J., Tarasick, D., Fioletov, V., Jin, J., Moeini, O., Liu, X., Sioris, C., and Osman, M.: A global tropospheric ozone climatology from trajectory-mapped ozone soundings, Atmospheric Chemistry and Physics, 13, 10659-10675, 2013a.

Liu, J., Tarasick, D., Fioletov, V., McLinden, C., Zhao, T., Gong, S., Sioris, C., Jin, J., Liu, G., and Moeini, O.: A global ozone climatology from ozone soundings via trajectory mapping: a stratospheric perspective, Atmospheric Chemistry and Physics, 13, 11441-11464, 2013b.

Liu, X., Bhartia, P., Chance, K., Spurr, R., and Kurosu, T.: Ozone profile retrievals from the Ozone Monitoring Instrument, Atmospheric Chemistry and Physics, 10, 2521-2537, 2010.

Livesey, N., Read, W., Wagner, P., Froidevaux, L., Santee, M., Schwartz, M., Lambert, A., Millán

Valle, L., Pumphrey, H., and Manney, G.: Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS) version 5.0 x level 2 and 3 data quality and description document Version 5.0–1.1 a (Tech. Rep.), Jet Propulsion Laboratory, California Institute of Technology. Retrieved from https://mls. jpl. nasa. gov/data/v5-0_data_quality_document. pdf, 2022.

McDermid, I. S., Godin, S. M., and Walsh, T. D.: Lidar measurements of stratospheric ozone and intercomparisons and validation, Applied Optics, 29, 4914-4923, 1990.

McPeters, R. D. and Labow, G. J.: Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, Journal of Geophysical Research: Atmospheres, 117, 2012.

McPeters, R. D., Labow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, Journal of Geophysical Research: Atmospheres, 112, 2007.

Millan, L. F., Manney, G. L., Boenisch, H., Hegglin, M. I., Hoor, P., Kunkel, D., Leblanc, T., Petropavlovskikh, I., Walker, K., and Wargan, K.: A Multi-Parameter Dynamical Diagnostics for Upper Tropospheric and Lower Stratospheric Studies, EGUsphere, 1-41, 2023.

Mills, G., Pleijel, H., Malley, C. S., Sinha, B., Cooper, O. R., Schultz, M. G., Neufeld, H. S., Simpson, D., Sharps, K., and Feng, Z.: Tropospheric Ozone Assessment Report: Present-day tropospheric ozone distribution and trends relevant to vegetation, Elem Sci Anth, 6, 47, 2018.

Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis, Atmospheric Chemistry and Physics, 20, 931-967, 2020a.

Miyazaki, K., Bowman, K., Sekiya, T., Eskes, H., Boersma, F., Worden, H., Livesey, N., Payne, V. H., Sudo, K., and Kanaya, Y.: Updated tropospheric chemistry reanalysis and emission estimates, TCR-2, for 2005–2018, Earth System Science Data, 12, 2223-2259, 2020b.

Moeini, O., Tarasick, D. W., McElroy, C. T., Liu, J., Osman, M. K., Thompson, A. M., Parrington, M., Palmer, P. I., Johnson, B., and Oltmans, S. J.: Estimating wildfire-generated ozone over North America using ozonesonde profiles and a differential back trajectory technique, Atmospheric Environment: X, 7, 100078, 2020.

Nédélec, P., Blot, R., Boulanger, D., Athier, G., Cousin, J.-M., Gautron, B., Petzold, A., Volz-

885 Thomas, A., and Thouret, V.: Instrumentation on commercial aircraft for monitoring the
886 atmospheric composition on a global scale: the IAGOS system, technical overview of ozone
887 and carbon monoxide measurements, Tellus B: Chemical and Physical Meteorology, 67,
888 27791, 2015.
889 Nowack, P. J., Luke Abraham, N., Maycock, A. C., Braesicke, P., Gregory, J. M., Joshi, M. M.,
890 Osprey, A., and Pyle, J. A.: A large ozone-circulation feedback and its implications for
891 global warming assessments, Nature Climate Change, 5, 41-45, 2015.
892 Petetin, H., Sauvage, B., Smit, H. G., Gheusi, F., Lohou, F., Blot, R., Clark, H., Athier, G.,
893 Boulanger, D., and Cousin, J.-M.: A climatological view of the vertical stratification of RH,
894 O3 and CO within the PBL and at the interface with free troposphere as seen by IAGOS
895 aircraft and ozonesondes at northern mid-latitudes over 1994–2016, Atmospheric Chemistry
896 and Physics, 18, 9561-9581, 2018.
897 Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A., Gallagher, M., Hermann,
898 M., Pontaud, M., Ziereis, H., and Boulanger, D.: Global-scale atmosphere monitoring by in-
899 service aircraft–current achievements and future prospects of the European Research
900 Infrastructure IAGOS, Tellus B: Chemical and Physical Meteorology, 67, 28452, 2015.
901 Polvani, L. M., Wang, L., Aquila, V., and Waugh, D. W.: The impact of ozone-depleting
902 substances on tropical upwelling, as revealed by the absence of lower-stratospheric cooling
903 since the late 1990s, Journal of Climate, 30, 2523-2534, 2017.
904 Prather, M. J. and Zhu, X.: Lifetimes and timescales of tropospheric ozone: Global metrics for
905 climate change, human health, and crop/ecosystem research, Elementa: Science of the
906 Anthropocene, 12, 2024.
907 Prather, M. J., Guo, H., and Zhu, X.: Deconstruction of tropospheric chemical reactivity using
908 aircraft measurements: the Atmospheric Tomography Mission (ATom) data, Earth System
909 Science Data, 15, 3299-3349, 2023.
910 Rahpoe, N., Weber, M., Rozanov, A., Weigel, K., Bovensmann, H., Burrows, J., Laeng, A., Stiller,
911 G., Von Clarmann, T., and Kyrölä, E.: Relative drifts and biases between six ozone limb
912 satellite measurements from the last decade, Atmospheric Measurement Techniques, 8,
913 4369-4381, 2015.
914 Sicard, M., Granados-Muñoz, M. J., Alados-Arboledas, L., Barragán, R., Bedoya-Velásquez, A.
915 E., Benavent-Oltra, J. A., Bortoli, D., Comerón, A., Córdoba-Jabonero, C., and Costa, M. J.:

916 Ground/space, passive/active remote sensing observations coupled with particle dispersion
917 modelling to understand the inter-continental transport of wildfire smoke plumes, Remote
918 Sensing of Environment, 232, 111294, 2019.
919 Skeie, R. B., Myhre, G., Hodnebrog, Ø., Cameron-Smith, P. J., Deushi, M., Hegglin, M. I.,
920 Horowitz, L. W., Kramer, R. J., Michou, M., and Mills, M. J.: Historical total ozone
921 radiative forcing derived from CMIP6 simulations, Npj Climate and Atmospheric Science, 3,
922 32, 2020.
923 Smit, H. G. and Kley, D.: JOSIE: The 1996 WMO International intercomparison of ozonesondes
924 under quasi flight conditions in the environmental simulation chamber at Jülich,
925 Proceedings of the XVIII Quadrennial Ozone Symposium, 971-974, 1998.
926 Smit, H. and Thompson, A.: Ozonesonde Measurement Principles and Best Operational Practices:
927 ASOPOS 2.0 (Assessment of Standard Operating Procedures for Ozonesondes), WMO,
928 World Meteorological Organization, GAW Report, 173, 2021.
929 Staufer, J., Staehelin, J., Stübi, R., Peter, T., Tummon, F., and Thouret, V.: Trajectory matching of
930 ozonesondes and MOZAIC measurements in the UTLS – Part 1: Method description
931 and application at Payerne, Switzerland, Atmospheric Measurement Techniques., 6, 3393-
932 3406, 10.5194/amt-6-3393-2013, 2013.
933 Staufer, J., Stähelin, J., Stübi, R., Peter, T., Tummon, F., and Thouret, V.: Trajectory matching of
934 ozonesondes and MOZAIC measurements in the UTLS–Part 2: Application to the global
935 ozonesonde network, Atmospheric Measurement Techniques, 7, 241-266, 2014.
936 Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Tarasick, D. W., Van Malderen, R., Smit, H.
937 G., Vömel, H., Morris, G. A., Johnson, B. J., and Cullis, P. D.: An examination of the recent
938 stability of ozonesonde global network data, Earth and Space Science, 9, e2022EA002459,
939 2022.
940 Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Witte, J. C., Tarasick, D. W., Davies, J.,
941 Vömel, H., Morris, G. A., Van Malderen, R., and Johnson, B. J.: A post-2013 dropoff in
942 total ozone at a third of global ozonesonde stations: Electrochemical concentration cell
943 instrument artifacts? Geophysical Research Letters, 47, e2019GL086791, 2020.
944 Stein, A., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M., and Ngan, F.: NOAA's
945 HYSPLIT atmospheric transport and dispersion modeling system, Bulletin of the American
946 Meteorological Society, 96, 2059-2077, 2015.

947 Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A.,
948 Degenstein, D., Damadeo, R., and Zawodny, J.: An update on ozone profile trends for the
949 period 2000 to 2016, Atmospheric Chemistry and Physics, 17, 10675-10690, 2017.
950 Sterling, C. W., Johnson, B. J., Oltmans, S. J., Smit, H. G., Jordan, A. F., Cullis, P. D., Hall, E. G.,
951 Thompson, A. M., and Witte, J. C.: Homogenizing and estimating the uncertainty in
952 NOAA's long-term vertical ozone profile records measured with the electrochemical
953 concentration cell ozonesonde, Atmospheric Measurement Techniques, 11, 3661-3687, 2018.
954 Stohl, A.: Computation, accuracy and applications of trajectories—A review and bibliography,
955 Atmospheric Environment, 32, 947-966, 1998.
956 Stohl, A., Eckhardt, S., Forster, C., James, P., Spichtinger, N., and Seibert, P.: A replacement for
957 simple back trajectory calculations in the interpretation of atmospheric trace substance
958 measurements, Atmospheric Environment, 36, 4635-4648, 2002.
959 Stohl, A., James, P., Forster, C., Spichtinger, N., Marenco, A., Thouret, V., and Smit, H. G.: An
960 extension of Measurement of Ozone and Water Vapour by Airbus In-service Aircraft
961 (MOZAIC) ozone climatologies using trajectory statistics, Journal of Geophysical Research:
962 Atmospheres, 106, 27757-27768, 2001.
963 Stohl, A. and Seibert, P.: Accuracy of trajectories as determined from the conservation of
964 meteorological tracers, Quarterly Journal of the Royal Meteorological Society, 124, 1465-
965 1484, 1998.
966 Szeląg, M. E., Sofieva, V. F., Degenstein, D., Roth, C., Davis, S., and Froidevaux, L.: Seasonal
967 stratospheric ozone trends over 2000–2018 derived from several merged data sets,
968 Atmospheric Chemistry and Physics, 20, 7035-7047, 2020.
969 Szopa, S., Naik, V., Adhikary, B., Artaxo, P., Berntsen, T., Collins, W., Fuzzi, S., Gallardo, L.,
970 Kiendler-Scharr, A., and Klimont, Z.: Short-lived climate forcers Climate Change 2021: The
971 Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of
972 the Intergovernmental Panel on Climate Change, Climate Change 2021: The Physical
973 Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the
974 Intergovernmental Panel on Climate Change, 817-922, 2021.
975 Tanimoto, H., Zbinden, R. M., Thouret, V., and Nédélec, P.: Consistency of tropospheric ozone
976 observations made by different platforms and techniques in the global databases, Tellus B:
977 Chemical and Physical Meteorology, 67, 27073, 2015.

978 Tarasick, D., Carey-Smith, T., Hocking, W., Moeini, O., He, H., Liu, J., Osman, M., Thompson,
979 A., Johnson, B., and Oltmans, S.: Quantifying stratosphere-troposphere transport of ozone
980 using balloon-borne ozonesondes, radar windprofilers and trajectory models, Atmospheric
981 Environment, 198, 496-509, 2019a.
982 Tarasick, D.W., J. Davies, H.G.J. Smit and S.J. Oltmans, A re-evaluated Canadian ozonesonde
983 record: measurements of the vertical distribution of ozone over Canada from 1966 to 2013,
984 Atmospheric Measurement Techniques, 9, 195-214, 2016.Tarasick, D.W., I. Galbally, O.R.
985 Cooper, M.G. Schultz, G. Ancellet, T. LeBlanc, T.J. Wallington, J. Ziemke, X. Liu, M.
986 Steinbacher, J. Stähelin, C. Vigouroux, J. Hannigan, O. García, G. Foret, P. Zanis, E.
987 Weatherhead, I. Petropavlovskikh, H. Worden, J.L. Neu, M. Osman, J. Liu, M. Lin, M.
988 Granados-Muñoz, A.M. Thompson, S.J. Oltmans, J. Cuesta, G. Dufour, V. Thouret, B.
989 Hassler, A.M. Thompson and T. Trickl, TOAR- Observations: Tropospheric ozone from
990 1877 to 2016, observed levels, trends and uncertainties, Elementa: Science of the
991 Anthropocene, 7(1), p.39, 2019b.
992 Tarasick, D., Jin, J., Fioletov, V., Liu, G., Thompson, A., Oltmans, S., Liu, J., Sioris, C., Liu, X.,
993 and Cooper, O.: High-resolution tropospheric ozone fields for INTEX and ARCTAS from
994 IONS ozonesondes, Journal of Geophysical Research: Atmospheres, 115, 2010.Tarasick,
995 D.W., H.G.J. Smit, A.M. Thompson G.A. Morris, J.C. Witte, J. Davies, T. Nakano, R. van
996 Malderen, R.M. Stauffer, T. Deshler, B.J. Johnson, R. Stübi, S.J. Oltmans and H. Vömel,
997 Improving ECC ozonesonde data quality: Assessment of current methods and outstanding
998 issues. Earth and Space Science, 8, e2019EA000914, 2021.
999 Terao, Y., Logan, J. A., Douglass, A. R., and Stolarski, R. S.: Contribution of stratospheric ozone
1000 to the interannual variability of tropospheric ozone in the northern extratropics, Journal of
1001 Geophysical Research: Atmospheres, 113, 2008.
1002 Thompson, A. M., Witte, J. C., Sterling, C., Jordan, A., Johnson, B. J., Oltmans, S. J., Fujiwara,
1003 M., Vömel, H., Allaart, M., and Piters, A.: First reprocessing of Southern Hemisphere
1004 Additional Ozonesondes (SHADOZ) ozone profiles (1998–2016): 2. Comparisons with
1005 satellites and ground-based instruments, Journal of Geophysical Research: Atmospheres,
1006 122, 13,000-013,025, 2017.
1007 Thouret, V., Marenco, A., Logan, J. A., Nédélec, P., and Grouhel, C.: Comparisons of ozone
1008 measurements from the MOZAIC airborne program and the ozone sounding network at

eight locations, Journal of Geophysical Research: Atmospheres, 103, 25695-25720, 1998.
Tilmes, S., Lamarque, J.-F., Emmons, L., Conley, A., Schultz, M., Saunois, M., Thouret, V., Thompson, A., Oltmans, S., and Johnson, B.: Ozonesonde climatology between 1995 and 2011: description, evaluation and applications, Atmospheric Chemistry and Physics, 12, 7475-7497, 2012.
Van Malderen, R., Allaart, M. A., De Backer, H., Smit, H. G., and De Muer, D.: On instrumental errors and related correction strategies of ozonesondes: possible effect on calculated ozone trends for the nearby sites Uccle and De Bilt, Atmospheric Measurement Techniques, 9, 3793-3816, 2016.
Vicedo-Cabrera, A. M., Sera, F., Liu, C., Armstrong, B., Milojevic, A., Guo, Y., Tong, S., Lavigne, E., Kyselý, J., and Urban, A.: Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries, bmj, 368, 2020.
Wang, G., Kong, Q., Xuan, Y., Wan, X., Chen, H., and Ma, S.: Development and application of ozonesonde system in China, Advances in Earth Science, 18, 471, 2003.
Wang, H. J., Cunnold, D. M., Thomason, L. W., Zawodny, J. M., and Bodeker, G. E.: Assessment of SAGE version 6.1 ozone data quality, Journal of Geophysical Research: Atmospheres, 107, ACH 8-1-ACH 8-18, 2002.
Wild, O., Voulgarakis, A., O'Connor, F., Lamarque, J.-F., Ryan, E. M., and Lee, L.: Global sensitivity analysis of chemistry–climate model budgets of tropospheric ozone and OH: exploring model diversity, Atmospheric Chemistry and Physics, 20, 4047-4058, 2020.
Witte, J. C., Thompson, A. M., Schmidlin, F., Northam, E. T., Wolff, K. R., and Brothers, G. B.: The NASA Wallops Flight Facility digital ozonesonde record: Reprocessing, uncertainties, and dual launches, Journal of Geophysical Research: Atmospheres, 124, 3565-3582, 2019.
Witte, J. C., Thompson, A. M., Smit, H. G., Vömel, H., Posny, F., and Stübi, R.: First reprocessing of Southern Hemisphere ADditional OZonesondes profile records: 3. Uncertainty in ozone profile and total column, Journal of Geophysical Research: Atmospheres, 123, 3243-3268, 2018.
Witte, J. C., Thompson, A. M., Smit, H. G., Fujiwara, M., Posny, F., Coetzee, G. J., Northam, E. T., Johnson, B. J., Sterling, C. W., and Mohamad, M.: First reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) profile records (1998–2015): 1. Methodology and evaluation, Journal of Geophysical Research: Atmospheres, 122, 6611-

6636, 2017.

Xu, W., Xu, X., Lin, M., Lin, W., Tarasick, D., Tang, J., Ma, J., and Zheng, X.: Long-term trends of surface ozone and its influencing factors at the Mt Waliguan GAW station, China–Part 2: The roles of anthropogenic emissions and climate variability, Atmospheric Chemistry and Physics, 18, 773-798, 2018.

Xuan, Y., Ma, S., Chen, H., Wang, G., Kong, Q., Zhao, Q., and Wan, X.: Intercomparisons of GPSO3 and Vaisala ECC ozone sondes, Plateau Meteorology, 23, 394-399, 2004.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J. L., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M. T., and Zeng, G.: Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Elementa: Science of the Anthropocene, 6, 10.1525/elementa.265, 2018.

Zbinden, R., Thouret, V., Ricaud, P., Carminati, F., Cammas, J.-P., and Nédélec, P.: Climatology of pure tropospheric profiles and column contents of ozone and carbon monoxide using MOZAIC in the mid-northern latitudes (24° N to 50° N) from 1994 to 2009, Atmospheric Chemistry and Physics, 13, 12363-12388, 2013.

Zeng, G., Morgenstern, O., Williams, J. H., O'Connor, F. M., Griffiths, P. T., Keeble, J., Deushi, M., Horowitz, L. W., Naik, V., and Emmons, L. K.: Attribution of stratospheric and tropospheric ozone changes between 1850 and 2014 in CMIP6 models, Journal of Geophysical Research: Atmospheres, 127, e2022JD036452, 2022.

Zeng, Y., Zhang, J., Li, D., Liao, Z., Bian, J., Bai, Z., Shi, H., Xuan, Y., Yao, Z., and Chen, H.: Vertical distribution of tropospheric ozone and its sources of precursors over Beijing: Results from~ 20 years of ozonesonde measurements based on clustering analysis, Atmospheric Research, 284, 106610, 2023.

Zhang, J., Li, D., Bian, J., Xuan, Y., Chen, H., Bai, Z., Wan, X., Zheng, X., Xia, X., and Lü, D.: Long-term ozone variability in the vertical structure and integrated column over the North China Plain: Results based on ozonesonde and Dobson measurements during 2001–2019, Environmental Research Letters, 16, 074053, 2021.

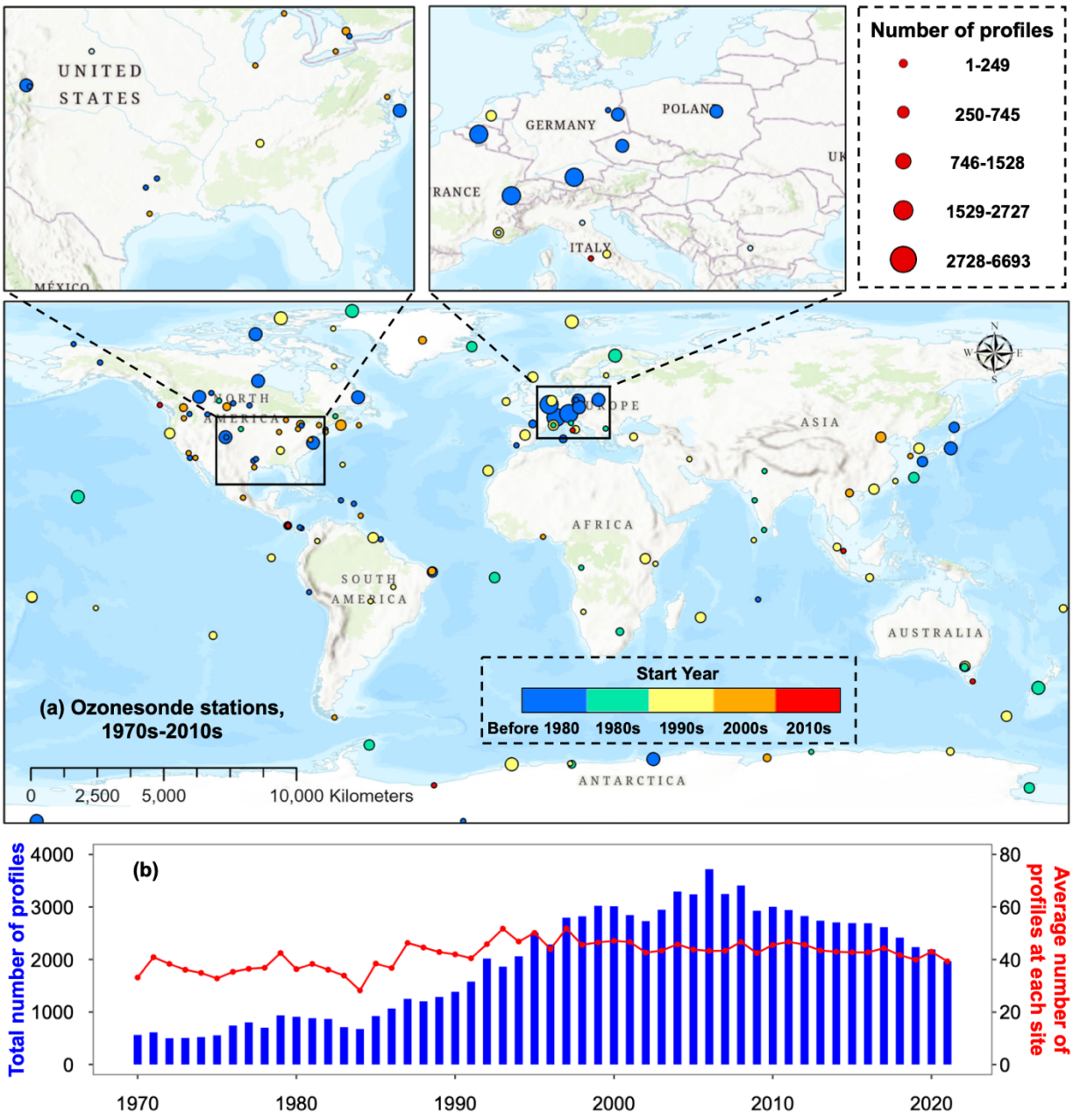


Figure 1. (a) Global distribution of ozonesonde stations used in this study to construct TOST-v2. Station details are provided in Table S1. The size and color of the dots indicate the total number of sounding profiles and the start year of the measurement time series. (b) The total number of profiles per year (left y-axis, blue bars) and the average number of profiles per site and per year (right y-axis, red dots and line) from 1970 to 2021.

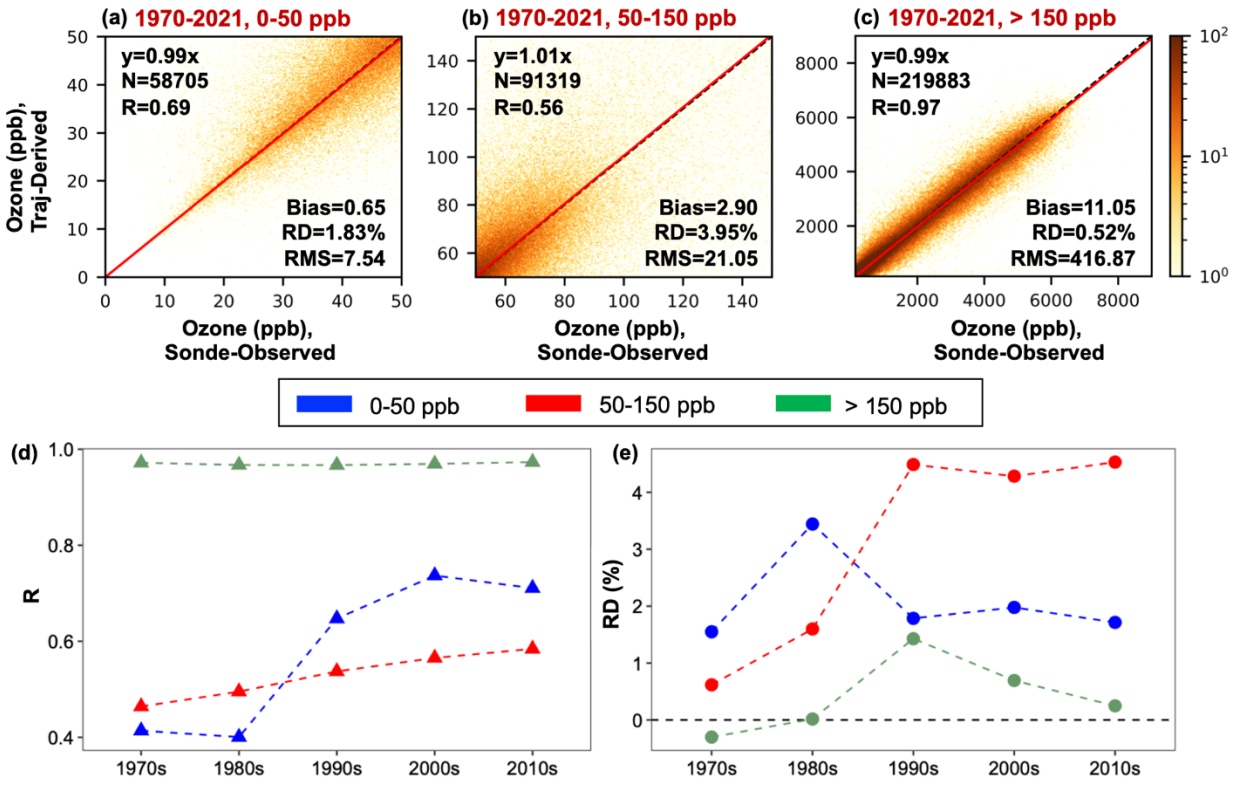


Figure 2. (a-c) Comparison of monthly mean tropospheric ozone mixing ratios from ozonesondes (Sonde-Observed) and trajectory-derived TOST data (Traj-Derived) for the study period at 0-50 ppb, 50-150 ppb, and >150 ppb. Solid red lines represent the linear fitting line (with the intercept set to 0) and dashed black lines denote the 1:1 axis. N is the total number of data points, R is the correlation coefficient, Bias is the overall average difference in monthly means [Traj-Derived ozone - Sonde-Observed ozone, in ppb], RD is the relative difference in % [100 × (Traj-Derived ozone - Sonde-Observed ozone)/ Sonde-Observed ozone], and RMS is the root mean square difference in ppb. Note that Traj-Derived ozone at each station is derived without input from the station itself; that is, Traj-Derived represents an ensemble of 141 separate computations of TOST, each one withholding a single validation station. (d-e) the R and RD between the Traj-Derived ozone and Sonde-Observed ozone by decade. The dashed black line in (e) denotes where the RD is 0.

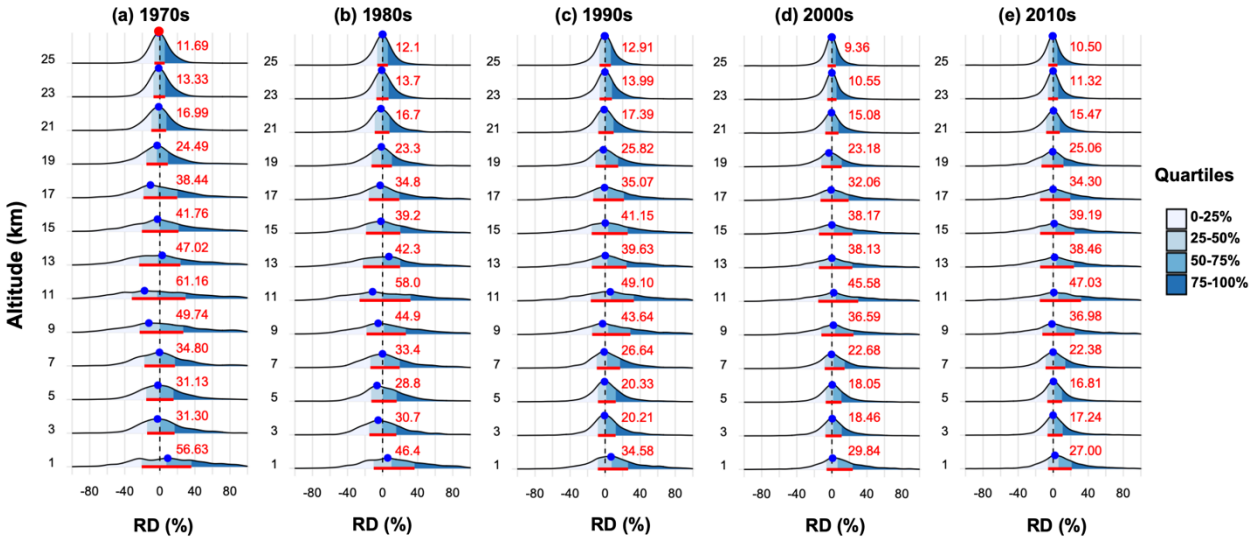


Figure 3. Frequency distribution of relative difference (RD) of the monthly ozone mixing ratios between ozonesonde and Traj-Derived data by every other altitude in the 1970s, 1980s, 1990s, 2000s and 2010s (y-axis: frequency in %, x-axis: RD in %), with the colors denoting the 4 quartiles of RD. The dashed line indicates zero difference in RD. The blue dot represents the maximum frequency. The thick red line denotes the width of distribution at 25-75 percentile, with the corresponding width of the distribution value in red.

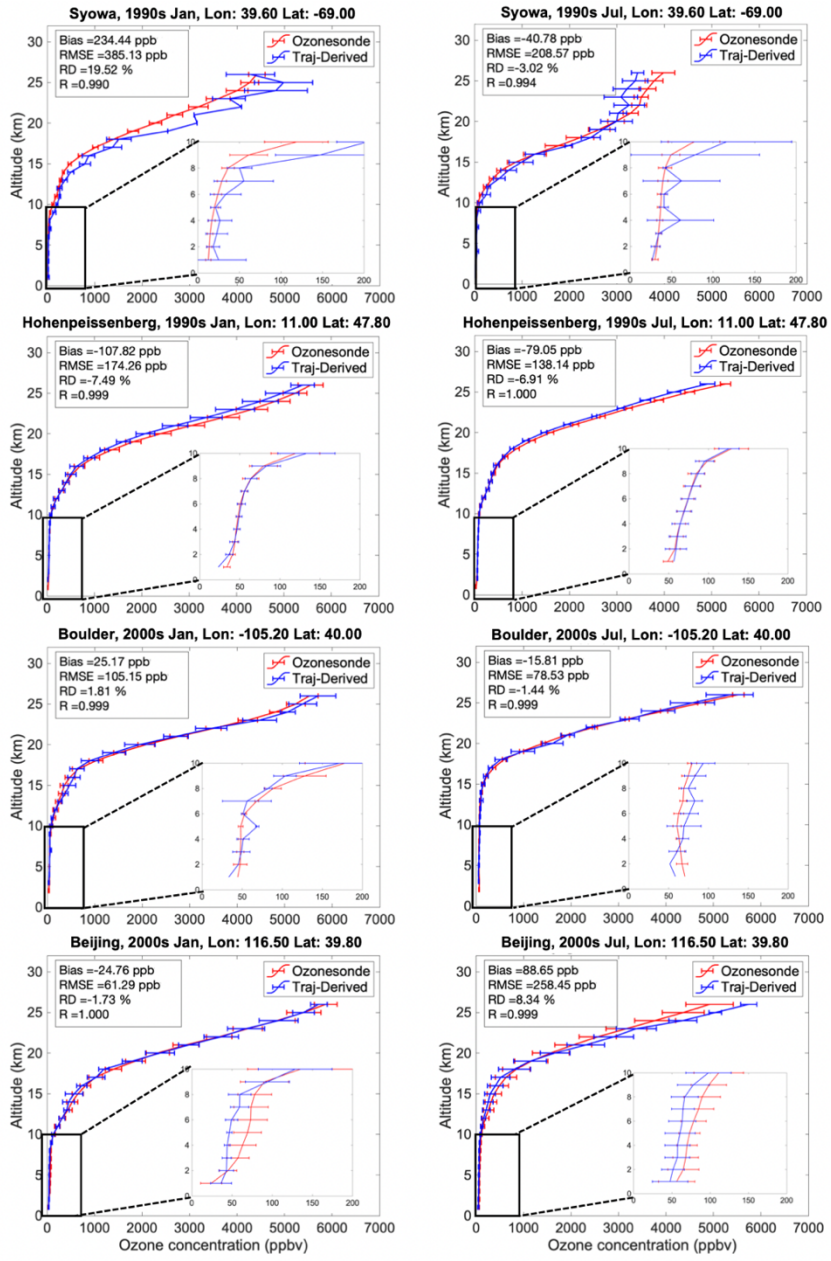
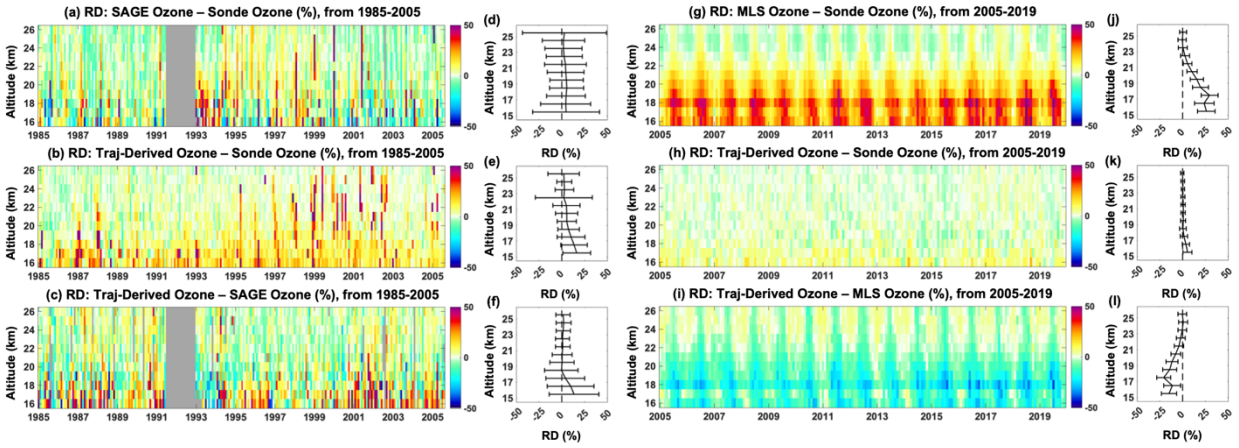


Figure 4. Decadal monthly mean ozone profiles at Syowa and Hohenpeissenberg in January and July of the 1990s, and at Boulder and Beijing in January and July of the 2000s. The red line denotes ozonesonde ozone and the blue line denotes trajectory-derived ozone without the input from the station itself. The error bar is ±2 times the standard error of the mean (equivalent to 95 % confidence limits on the averages). To better compare the difference of ozone profiles in the troposphere, a zoom-in window from 0-10 km is provided in each sub-figure.



1105

1106 Figure 5. (a). The relative difference (RD) between ozonesonde and SAGE ozone data in each
1107 month and at each altitude during 1985-2005 over 16-26 km [RD = 100×(SAGE ozone –
1108 ozonesonde ozone)/ozonesonde ozone, in %]. The mean RD over 1985-2005 at each level is
1109 shown on the right (d), where the error bars represent the standard deviation of the monthly RD
1110 over 1985-2005. Note that the Pinatubo-affected SAGE profiles are excluded during July 1991-
1111 December 1992 (filled with gray color). (b) the same as (a), but for the RD between ozonesondes
1112 and Traj-derived data, [RD = 100×(Traj-derived ozone – ozonesonde ozone)/ozonesonde ozone,
1113 in %]. (c) the same as (a), but for the RD between Traj-derived and SAGE ozone data [RD = 100
1114 × (Traj-derived ozone – SAGE ozone)/(0.5 × Traj-derived ozone + 0.5 × SAGE ozone), in %].
1115 (d-f) the averaged RD by altitude corresponding to (a-c). (g-l) same as (a-f), but for the period of
1116 2005-2019 and the satellite measurements are from MLS ozone.

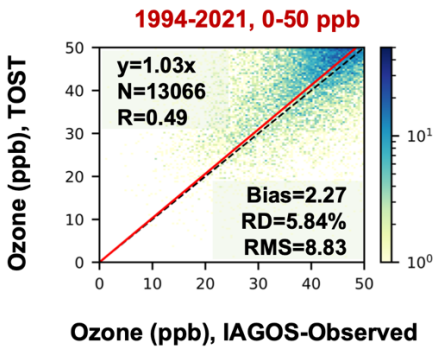


Figure 6. The comparison of monthly ozone mixing ratios between IAGOS-observed (x-axis labeled: IAGOS-Observed) and TOST data (y-axis labeled: TOST) from 1994-2021 for ozone concentrations below 50 ppb. Solid red lines represent the linear fitting line (with the intercept set to 0) and dashed black lines denote the 1:1 axis. N is the total number of data points, R is the correlation coefficient (unitless), Bias is the difference in monthly mean values [TOST ozone – IAGOS ozone, unit: ppb], RD is the relative difference [100 × (TOST ozone - IAGOS ozone)/(0.5 × TOST ozone + 0.5 × IAGOS ozone)], and RMS the root mean square difference (unit: ppb).

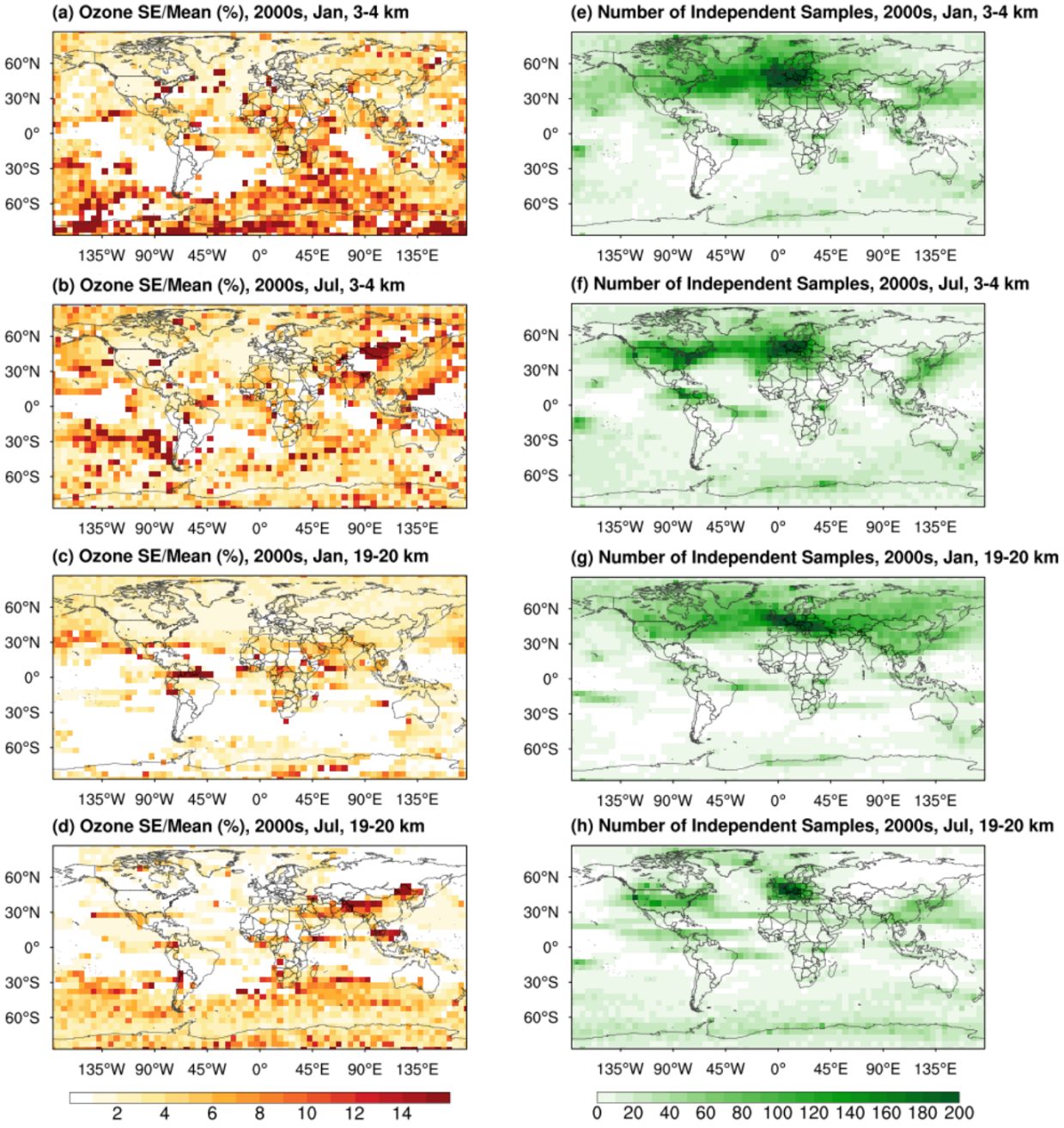


Figure 7. (a-d) Global distribution of the SE/Mean (left panels, in %) for the decadal monthly mean ozone in January and July 2000s at 3-4 km (a and b) and 19-20 km (c and d). (e-h) the same as (a-d), but for the number of independent samples in each 5 × 5° bin.

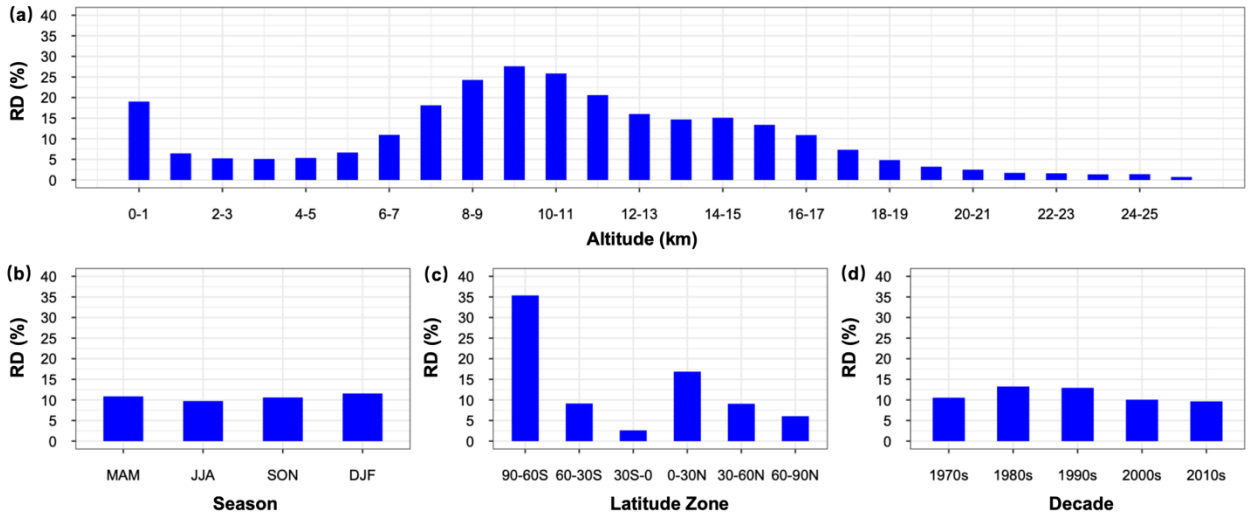


Figure 8. The RDs (in %) of TOST over 1990-2021 by altitude (a), and the average RDs over all altitudes by season [MAM (March-April-May), JJA (June-July-August), SON (September-October-November) and DJF (December-January-February)] (b), latitudinal zone (c), and decade (d). RD is calculated by the mean ozone mixing ratio from Traj-Derived ozone and ozonesondes measurements [100 × (Traj-Derived ozone - Sonde-Observed ozone)/ Sonde-Observed ozone].

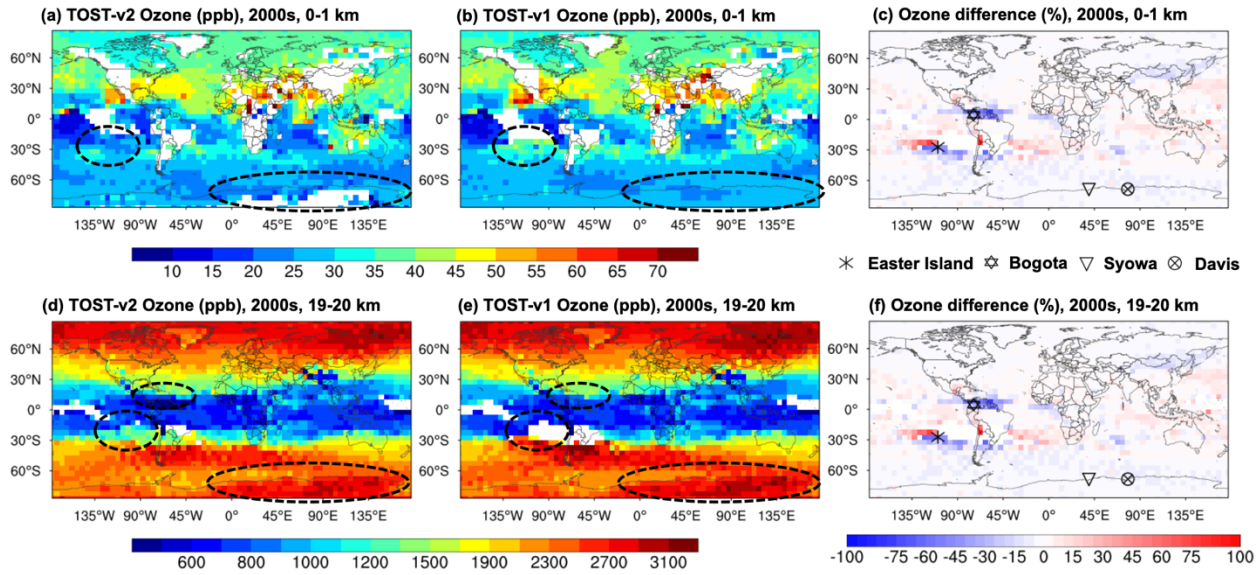


Figure 9. (a, b) The global distributions of ozone in TOST-v2 (a) and TOST-v1 (b) over 0-1 km in the 2000s. (d, e) The same as for (a, b), but over 19-20 km. The dashed circles indicate regions with large differences between the two versions. (c) The global distributions of RD between TOST-v2 and TOST-v1 [RD = 100 × (TOST-v2 – TOST-v1)/(0.5 × TOST-v2 + 0.5 × TOST-v1), in %] over 0-1 km in the 2000s. (f) The same as for (c), but over 19-20 km. The markers indicate the positions of Davis, Easter Island, Bogota, and Syowa stations.

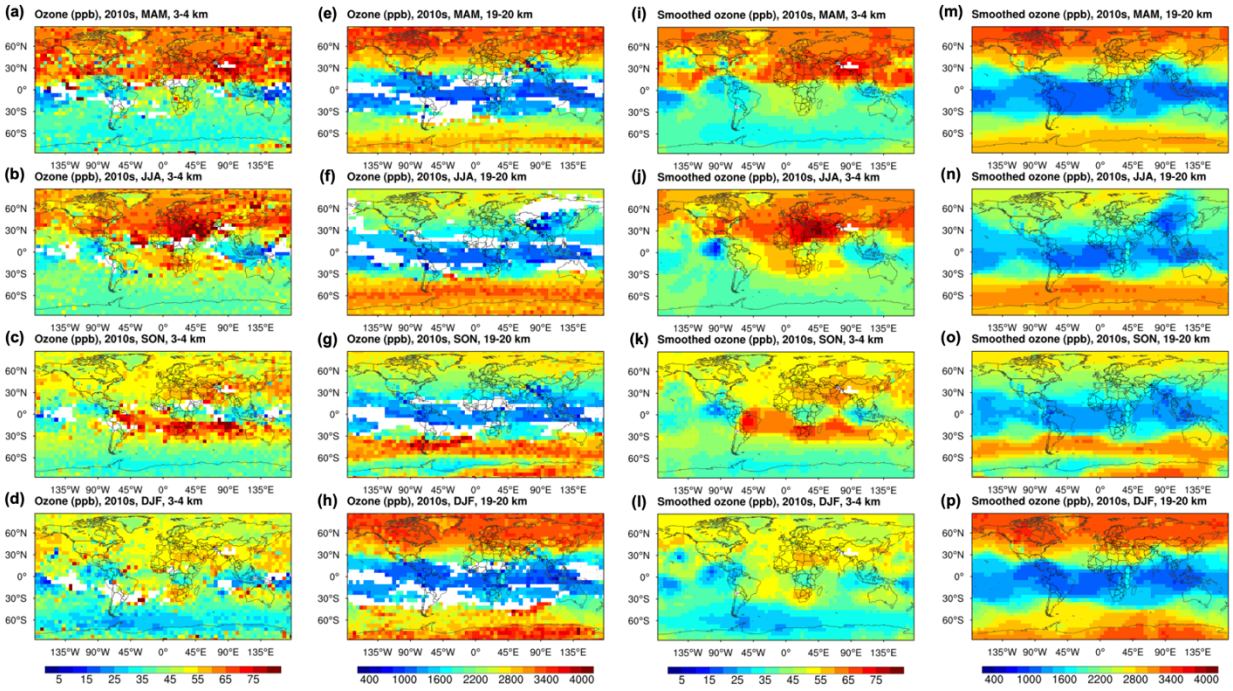


Figure 10. Global distribution of decadal mean TOST ozone at 3-4 km and 19-20 km in MAM (March-April-May), JJA (June-July-August), SON (September-October-November) and DJF (December-January-February) in the 2010s (a-h), and the corresponding smoothed TOST ozone (i-p).

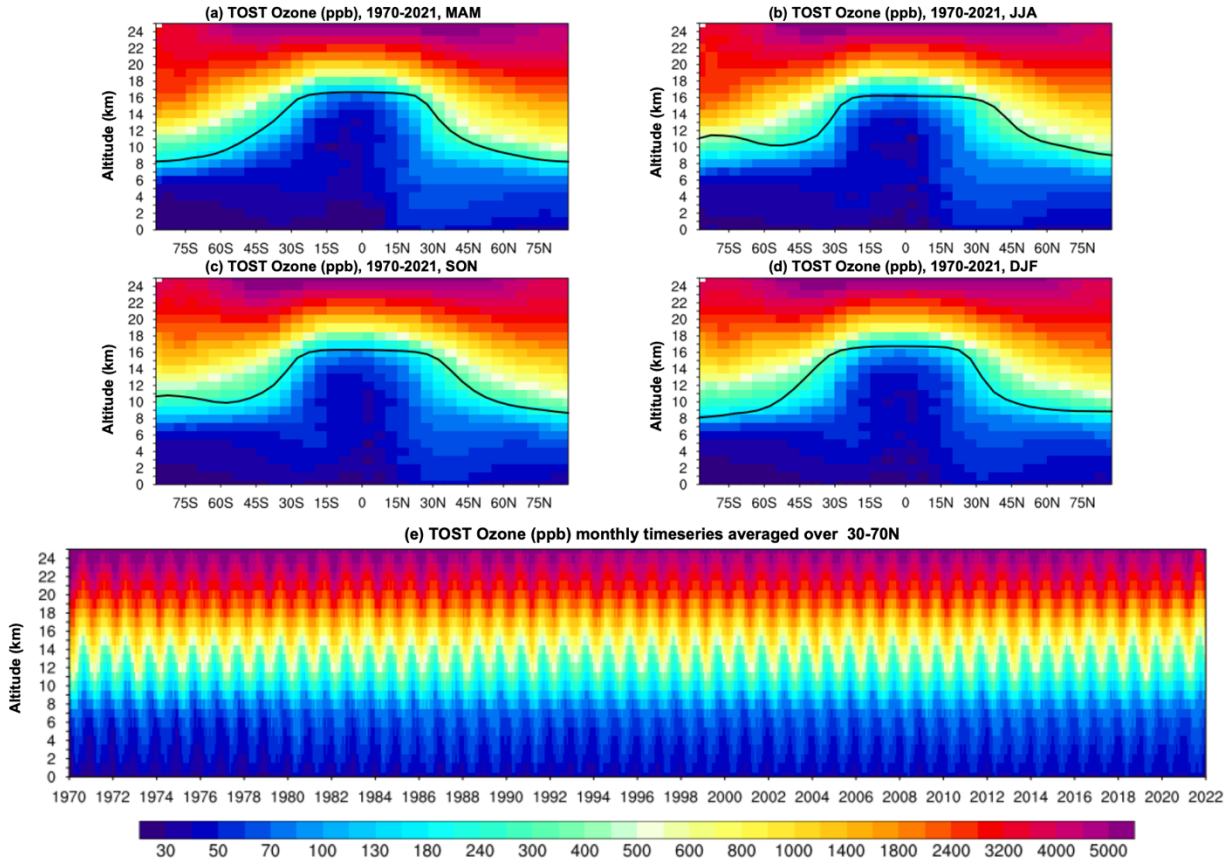
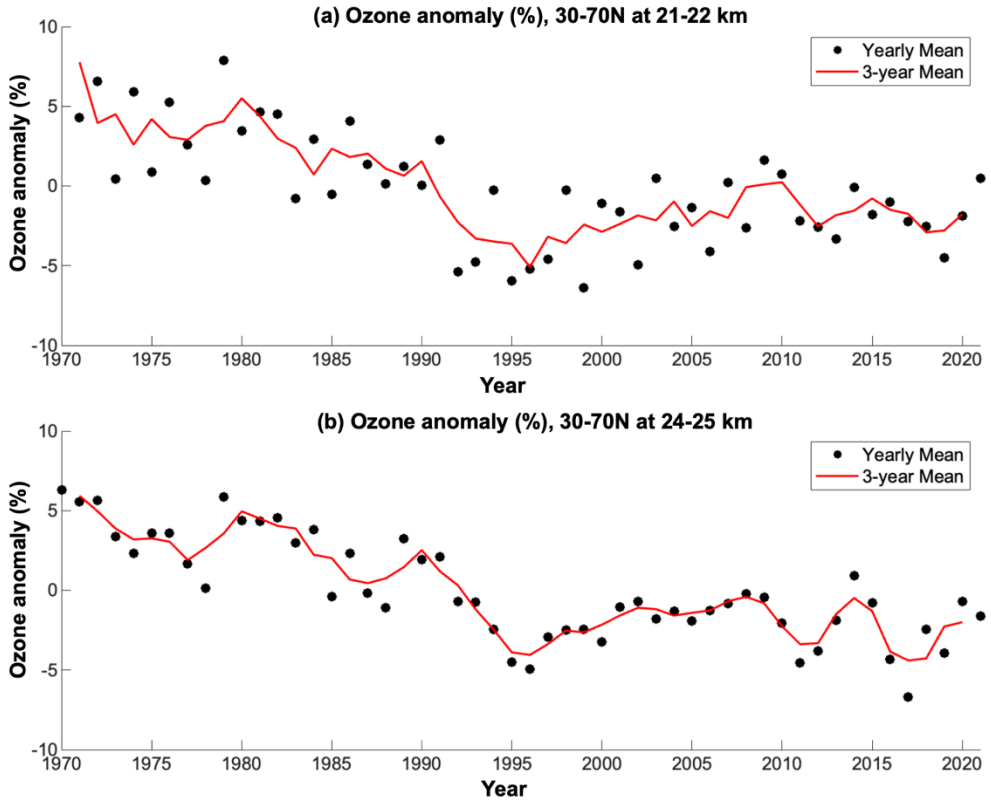


Figure 11. (a-d) The latitude-altitude distribution of TOST ozone averaged over 1970-2021 in each season. The solid black lines represent the mean tropopause height over 1970-2021 in each season. (e) time series of the monthly mean TOST ozone over 30-70°N at each altitude level from 1970 to 2021.



1163

1164 Figure 12. TOST time series of the annual mean ozone mixing ratios anomaly (in %) averaged

1165 over 30°-70°N over 21-22 km altitude (a) and 24-25 km altitude (b). The black dots represent the

1166 annual mean ozone concentrations from the area-weighted average of the grid cells over 30-70°N

1167 with ozone data throughout 1970-2021. The red line is the 3-year running mean.

1168

1169 Table 1. The description of the classifiers and the corresponding types for the TOST data.

Classifier	Type	Description
1. Vertical coordinate	Geometric	Altitude coordinates are 1, 2, …, 25, and 26 km at 1-km vertical resolution.
	Pressure	Altitude coordinates are 950, 850, 750, 650, 550, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 90, 80, 70, 60, 50, 40, 35, 30, 25, 20 hPa.
2. Starting level	Sea-level	Data start at the altitude of the sea surface. Ozone value for levels beneath the topography of the Earth's surface is set to null.
	Ground-level	Data start at the altitude of the ground surface.
3. Temporal resolution	Seasonal	Data are the mean for each season of the year (1970–2021).
	Annual	Data are the annual mean for each year (from 1970-2021). Each grid requires at least one value per season to be included in the annual data.
	Decadal-monthly	Data are the monthly mean for each month of a decade (from the 1970s to 2010s).
4. Ozonesonde data used	Trop-stra	Data are based on ozonesonde profiles in both the troposphere and stratosphere. This is the main dataset of TOST.
	Troposphere-only	Data are based on ozonesonde profiles only in the troposphere.
	Stratosphere-only	Data are based on ozonesonde profiles only in the stratosphere.
5. Ozone variation	Mean	The mean ozone concentrations for each grid cell over a period (a month, a year, or a month of a decade).
	25th, 50th and 75th percentiles	The 25th, 50th and 75th percentiles of ozone concentrations for each grid cell over a period (a month, a year, or a month of a decade). Only available for trop-stra data.
6. Supplement data	Smoothed data	Smoothed ozone fields by fitting the maps at each level with a linear combination of spherical functions.
	N	The total number of samples in each grid cell.
	N of independent samples	The total number of trajectories passed in each grid cell. A trajectory is counted only once in a grid cell when the trajectory passes that cell regardless of how long the trajectory stays in that cell.
	STD	The standard deviation for each grid cell.
	CV(STD/mean)	The CV for each grid cell.
	SE	The standard error for each grid cell.
	SE/Mean	The ratio of the standard error to the mean in each grid cell.

1170