Università degli Studi della Basilicata

Dottorato di Ricerca in
"Ingegneria per l'innovazione e lo sviluppo sostenibile"

TITOLO DELLA TESI
"Study of ozone, temperature, and water vapour in the UT/LS using upper-air measurements"

Settore Scientifico-Disciplinare
"Metodi e Tecnologie per il monitoraggio e la tutela ambientale"

Coordinatore del Dottorato   Dottorando

Prof.ssa Aurelia Sole   Dott. Fabrizio Marra

Relatori

Prof. Valerio Tramutoli
Prof. Fabio Madonna

Ciclo XXXVI

Index

Abstract

This work focuses on estimating trends in ozone partial pressure, temperature, and water vapour in the upper troposphere/lower stratosphere (UT/LS) region using a new unified ozonesounding profile database and a novel homogenised dataset named RHARM (Radiosounding HARMonization).

Studying temperature, water vapour, and ozone trends is key for studying climate change and climate variability. Temperature changes in the UT/LS are related to both internal processes, for example, changes in sea surface temperature (SST) and external forcing, such as greenhouse gases (GHGs) and ozone-depleting substances (ODS) (Randel et al., 2009)[1]. Various studies have been conducted to estimate the trends for different climate variables at a regional and global scale. The radiative effects of rising GHGs and changes in stratospheric ozone as a response to human emissions of ODS have led to net warming of the troposphere and cooling of the stratosphere (Hartmann et al., 2013)[2]. The impact of ODS on tropical upwelling, revealed by the absence of lower-stratospheric cooling, has been reported since 1998 (Polvani et al., 2017)[3]. Moreover, an increased tropopause temperature in the period 2001–2011 associated with a weaker tropopause inversion layer, due to the weakened upwelling in the Tropics, was found using Global Positioning System Radio Occultation (GNSS-RO) data and simulations with the National Center for Atmospheric Research's Whole Atmosphere Community Climate Model (WACCM). Such changes in the thermal structure of the UT/LS may have important implications for climate, such as a possible rise in water vapour in the lower stratosphere (Wang et al., 2013)[4]. More recently, balloon-borne radiation measurements proved that the stratosphere is warming after years of cooling (Philipona et al., 2016)[5]. Whether a slowdown or change in temperature sign in the UT/LS will persist in the future is an open question.

---

[1] W. J. Randel, K. P. Shine, J. Austin, J. Barnett, C. Claud, N. P. Gillett, P. Keckhut, U. Langematz, R. Lin, C. Long, C. Mears, A. Miller, J. Nash, D. J. Seidel, D. W. J. Thompson, F. Wu, and S. Yoden, « An update of observed stratospheric temperature trends,» *JGR,* vol. 114, D02107, 2009.

[2] J. Hartmann, A. West, P. Renforth, P. Köhler, C. L. De La Rocha, D. A. Wolf-Gladrow, H. H. Dürr, and J. Scheffran, «Enhanced chemical weathering as a geoengineering strategy to reduce atmospheric carbon dioxide, supply nutrients, and mitigate ocean acidification», Reviews of Geophysics, vol. 51, p. 113-149, 2013.

[3] L. M. Polvani, L. Wang, V. Aquila, and D. W. Waugh, «The Impact of Ozone-Depleting Substances on Tropical Upwelling, as Revealed by the Absence of Lower-Stratospheric Cooling since the Late 1990s,», Journal of Climate, vol. 30, 2523–2534, 2017.

[4] Wang, R., «GOZCARDS merged data for ozone monthly zonal means on a geodetic latitude and pressure grid v1.01», NASA Goddard Earth Sciences Data and Information Services Center, 2013.

[5] R. Philipona, and A. Kräuchi «Return glider radiosonde for in situ upper-air research measurements,», AMT, vol. 9, 2535–2544, 2016.

This work estimates ozone trends using balloon-borne measurements from three existing datasets: the Southern Hemisphere Additional OZonesondes (SHADOZ), Network for the Detection of Atmospheric Composition Change (NDACC), and World Ozone and Ultraviolet Radiation Data Centre (WOUDC) networks, developed respectively by NASA, NOAA, with the collaboration of many different institutes and PIs around the world, and WMO. These datasets are merged to provide appropriate data coverage at different latitudes and increase sampling for improving the calculation of anomalies and trends in the ozone concentration at the global scale. The resulting unified dataset removes duplicated profiles. Duplication for ozonesounding profiles often occurs when measurements from the same station are submitted to several networks, which in theory should be identical but are often provided for different periods, using different data formats, and providing a different amount of individual data points. Metadata may also differ. This also means that the different networks do not always report the same number of ozone levels for the same profile. A range of selection criteria has been applied to overcome this issue and harmonise the existing ozonesounding datasets to refine the quality and ensure the identification of outliers by applying a series of quality checks (QC) listed below:

- Plausibility checks: reported values should be within plausible physical range;
- Completeness check: on a monthly basis to verify that all variables are complete;
- Outliers check: using the Inter-Quartile Range method as follows:

  -
$$median - 3 \cdot IQR \leq observation \leq median + 3 \cdot IQR,$$
  -

- Vertical coverage checks: on a monthly basis to verify if ozone profiles reach 10 hPa;
- Vertical completeness checks: to ensure a minimum number of reports are available for each vertical region covered by the ozonesoundings;
- Statistics of missing values: to check the coherency with the source datasets.

The unified dataset is then grouped according to their monthly coverage to quantify sampling uncertainties in the trend calculation. The 155 available stations are separated into three different clusters:

1. Long Coverage (LC): 26 stations (with a data time series of at least 20 years).
2. Medium coverage (MC): 23 stations (with a data time series between 10 and 20 years).
3. Short coverage (SC): 106 stations (with data time series of less than 10 years).

The first two clusters are the only ones with sufficient data coverage for estimating anomalies and trends. The latter is estimated from monthly mean anomalies using LC, MC, and their combination. Different regression methods are used for estimating trends to provide for quantification of structural uncertainties in the trend calculation, including:

- Least-square linear regression (Reinsel et al. 2002)[6];
- LOTUS regression (Petropavlovskikh et al., 2019[7]; Godin-Beekmann et al., 2022[8]);
- Least Absolute Deviation (LAD) regression (Rice and White, 1964[9]; Barrodale, 1968[10]; Wong and Schneider Jr, 1989[11]; Calitz and Rüther, 1996[12]; Santer et al., 2000[13]);
- Theil-Sen regression (Theil, 1950[14]; Siegel and Benson, 1982[15]; Helsel and Hirsch, 1992[16]);
- The Mann-Kendal (MK) test (Kendall, 1975[17]; Mann, 1945[18]) is also used to statistically assess if there is a significant trend of the variable of interest over time.

The comparison shows that comparing trends estimated from LC data and a combination of LC and MC provides very similar percentage trends. For the 50-1 hPa layer, for example, the differences, comparing the different regressors, range from 0.6%/decade to 1.2%/decade. These represent an estimate of sampling uncertainty in cases where the trends are significant. Therefore, the LC data, representing the highest quality data according to the above criteria, is used to estimate trends. The estimates on this cluster show, for the Northern Hemisphere mid-latitudes (NH):

- a negative trend of 5% for the period pre-2000 at 50-1 hPa layer, reducing to 1% for the period post-2000 at 50-1 hPa,
- a negative trend of 10% for the period pre-2000 at 100-50 hPa, in contrast to a positive trend of 4% for the period post-2000 at 100-50 hPa.

[6] G. C. Reinsel, E. Weatherhead, G. C. Tiao, A. J. Miller, R. M. Nagatani, D. J. Wuebbles, and L. E. Flynn, «On detection of turnaround and recovery in trend for ozone,», JGR, vol. 107, ACH 1-1-ACH 1-12, 2002.

[7] I. Petropavlovskikh, S. Godin-Beekmann, D. Hubert, R. Damadeo, B. Hassler, and V. Sofieva, «SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere», SPARC Report No. 9, GAW Report No. 241, WCRP-17/2018, 2019.

[8] S. Godin-Beekmann, N. Azouz, V. F. Sofieva, D. Hubert, I. Petropavlovskikh, P. Effertz, G. Ancellet, D. A. Degenstein, D. Zawada, L. Froidevaux, S. Frith, J. Wild, S. Davis, W. Steinbrecht, T. Leblanc, R. Querel, K. Tourpali, R. Damadeo, E. Maillard Barras, R. Stübi, C. Vigouroux, C. Arosio, G. Nedoluha, I. Boyd, R. Van Malderen, E. Mahieu, D. Smale, and R. Sussmann, «Updated trends of the stratospheric ozone vertical distribution in the 60° S–60° N latitude range based on the LOTUS regression model», ACP, vol. 22, 11657–11673, 2022.

[9] J. R. Rice and J. S. White, «Norms for smoothing and estimation», SIAM Review, vol. 6, No. 3, 1964.

[10] I. Barrodale, «L1 Approximation and the Analysis of Data», Royal Statistical Society, vol 17, 51-57, 1968.

[11] R. K.W. Wong, C. Schneider, and P. W. Mielke Jr, «Geometric consistency for regression model estimation and testing in climatology and meteorology», Atmosphere-Ocean, vol. 27, 508-520, 1989.

[12] M.F. Calitz, H. Rüther, «Least absolute deviation (LAD) image matching», ISPRS Journal of Photogrammetry and Remote Sensing, vol. 51, 223-229, 1996.

[13] B. D. Santer, T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, K. E. Taylor, «Statistical significance of trends and trend differences in layer-average atmospheric temperature time series», JGR, vol. 105, 7337-7356, 2000.

[14] H. Theil, «A rank-invariant method of linear and polynomial regression analysis», Indagationes Mathematicae, vol. 12, 1950.

[15] A. F. Siegel, R. H. Benson, «A Robust Comparison of Biological Shapes», Biometrics, Vol. 38, No. 2, 341-3, 1982.

[16] D. R. Helsel, R. M. Hirsch, «Statistical methods in water resources», Elsevier, 1992.

[17] M. G. Kendall, «Rank Correlation Methods», 4th edition, Charles Griffin, London, 1975.

[18] H. B. Mann, «Non-parametric tests against trend», Econometrica, vol. 13, 163-171, 1945.

For the Tropics (TR) sector,

- for the period pre-2000, a positive trend of about 8% at 50-1 hPa, in contrast to a positive trend of 2% at 50-1 hPa for post-2000,
- for the period pre-2000, a positive trend reached 10% at 100-50 hPa on the other hand, for the post-2000 period the estimated trend did not pass the MK test so it is not significant.

The estimates for the NH sector are consistent with those presented by Petropavlovskikh et al. (2019): for the pre-2000 time series, in the lower stratosphere, there is a negative trend of 5% per decade that also reaches a negative trend of 10% at 100 hPa; for the post-2000 time series, in the lower stratosphere, there is a small negative trend of 1%, and at 100 hPa, a negative trend of 2%. This value has an uncertainty of ±7%, which makes the result produced in this work (about 4% positive trend) within the uncertainty range. For the TR, the pre-2000 time series, in the lower stratosphere, the trends presented in Petropavlovskikh et al. (2019) show a negative trend of 2% per decade and more, reaching a negative trend of 10% at 100 hPa, in contrast to the estimate shown in this work that presents a positive trend of about 8% in the lower stratosphere, reaching 10% at 100 hPa. This discrepancy is probably due to the small number of stations (only 4) used in this work for trend estimation. However, in the post-2000 time series, there is a positive trend of 2% in the lower stratosphere, as also shown in this work, and, at 100 hPa, a positive trend of 8%. This value, like the NH, has an uncertainty limit of ±7%, however, the trend evaluated in this work for the 100-50 hPa layer was not found to be significant by the MK test.

Finally, for temperature and water vapour trends, this study uses a novel dataset, named Radiosounding HARMonization (RHARM), providing a homogenized data set of temperature, humidity, and wind profiles along with an estimation of the measurement uncertainties for 697 radiosounding stations globally. The RHARM method has been used to adjust twice daily (0000 and 1200 UTC) radiosonde data holdings at 16 pressure levels in the range of 1,000–10 hPa, from 1978 to the present, provided by the Integrated Global Radiosonde Archive. Relative humidity data are limited to 250 hPa. The applied adjustments are interpolated to all reported levels. RHARM is the first data set to provide a homogenized time series with an estimation of the observational uncertainty at each sounding pressure level. By construction, RHARM-adjusted fields are not affected by cross-contamination of biases across stations and are fully independent of reanalysis data. RHARM shows warming trends of 0.39 K/decade at 300 hPa in the NH and 0.25 K/decade in the TR. The RHARM adjustments also reduce differences with the European Centre for Medium-Range Weather Forecast ERA5 reanalysis, with the strongest effect in the NH for temperature and relative humidity.

1. Vertical structure of the atmosphere

1.1 Water vapour and ozone in the atmosphere

Identifying, explaining and quantifying variations of climate-relevant atmospheric trace gases, in particular ozone and water vapour, are among the central tasks of atmospheric research.

Water vapour is the most important atmospheric greenhouse gas. Human activities, even if indirect, are the main cause of the increase in the tropospheric concentration of water vapour. The combustion of fossil fuels is the main cause of the increasing concentrations of CO_2 and CH_4 which lead to warmer tropospheric temperatures, responsible for the increase in atmospheric humidity, leading to a warmer climate (IPCC, 2007). Other processes that can alter stratospheric water vapour are polar dehydration and the transport of air from the troposphere to the stratosphere that bypasses the tropical cold trap (one example being convection that overshoots the tropopause). Water vapour in the upper troposphere-lower stratosphere (UT/LS) region plays an important role in the Earth's radiation budget and climate system. Its presence at these altitudes is predominantly associated with two main sources: transport from the troposphere, which occurs mainly in the Tropics, and in situ oxidation of methane. Changes in temperature and water vapour concentration in the UT/LS cause alterations in radiative forcing (e.g., Riese et al., 2012). Observations have shown that stratospheric water vapour concentration increases with increasing tropospheric temperature, implying the existence of stratospheric water vapour feedback (Dessler et al., 2013). The strength of this feedback has been estimated to be approximately 0.3 W m^{-2} K^{-1} (Dessler et al., 2013). Stratospheric water vapour also plays an important role in forming stratospheric clouds, a key element in stratospheric ozone depletion mechanisms (Di Sarra et al., 1992; Di Girolamo et al., 1994). Furthermore, stratospheric water vapour has primary importance in the processes leading to the formation of hydrogen radicals, and consequently in stratospheric chemistry and ozone depletion mechanisms (Lossow et al., 2019).

Ozone is a gas mostly found in the stratosphere and absorbs UV radiation from the sun, warming the stratosphere. Ozone is also a greenhouse gas, warming the troposphere, and thus is crucial to the climate system. Additionally, it can filter off potentially dangerous UV rays and defend surface life (IPCC/TEAP, 2005). Kirk-Davidoff et al. (1999) noted that climate change could impact stratospheric ozone through changes in water vapour. The scenario envisioned in that study was that tropical tropopause temperature would increase in response to greenhouse gas increases, thereby allowing more water vapour to enter the stratosphere. They

noted that the threshold temperature for halogen activation allowing heterogeneous ozone-depleting reactions on polar stratospheric cloud (PSC) particles was a function of water vapour concentration; as water increases, that threshold temperature increases and ozone-depleting reactions can continue later into the spring season. An increase in water vapour could then lead to enhanced ozone depletion during polar winter/spring in conjunction with enhanced frequency of polar stratospheric clouds (PSCs), in particular in the Arctic.

Before the year 2000, several ozone-depleting chemicals (ODSs), the majority of which are also greenhouse gases, were released into the atmosphere, causing stratospheric ozone levels to decrease. The Paris Convention and its Amendments and Adaptations currently regulate chlorofluorocarbons (CFCs) and halons, which are the principal causes of the anthropogenic ozone hole. The findings provide strong indications of falling ozone concentrations in the upper stratosphere (at altitudes above the 10-5 hPa level) from the mid-1980s through the late 1990s. In the recent decades post-2000, however, positive trends were found throughout the upper stratosphere and part of the mid-latitudes from both satellite and ground-based observations, although results vary for terrestrial data depending on the observation technique (Petropavlovskikh et al., 2019).

In this chapter, we aim to provide the essentials of physical and chemical processes occurring in the stratosphere that determine the content of ozone and water vapour as well as their effect on radiation budget and climate.

1.2 Thermal structure of the atmosphere

The atmosphere is classified based on its thermal structure (Figure 1), which determines the dynamic properties of individual regions. The simplest picture of the thermal structure of the atmosphere is provided by the vertical profile of global mean temperature in Figure 4 (panel b). From the surface to about 10 km, the temperature decreases with altitude at a nearly constant rate. This layer immediately above the Earth's surface is known as the troposphere. Having a global mean lapse rate of approximately 6.5 K/km, the troposphere contains the majority of what is known as weather and is ultimately driven by surface warming. The upper limit of the troposphere or "tropopause" is located at an altitude of approximately 10 km (100 hPa) and is characterized by an abrupt change in time-lapse rate. The region from the tropopause to an altitude of about 85 km is known as the middle atmosphere. Above the tropopause, the temperature first remains almost constant and then increases into the stratosphere. The increase in temperature with altitude (negative lapse rate) in the stratosphere reflects the warming of ozone, which results from the absorption of solar UV rays. In contrast to the troposphere, the stratosphere involves only weak vertical motions and is dominated by radiative processes. The upper limit of the stratosphere or "stratopause" is located at an altitude of about 50 km (1 hPa), where the temperature reaches its maximum. Above the stratopause, the temperature decreases again with altitude in the mesosphere, where ozone warming decreases. Convective motions and radiative processes are both important in the mesosphere. Meteor trails form in this region of the atmosphere, as do the lower layers of the ionosphere during daylight hours. The "mesopause" is located at an altitude of approximately 85 km (or 0.01 hPa), where a second temperature minimum is reached. Above the mesopause, the temperature increases steadily in the thermosphere. Unlike the lower regions, the thermosphere cannot be treated as an electrically neutral continuum. The ionization of molecules by energetic solar radiation produces a plasma of free electrons and ions, each of which interacts differently with the Earth's electric and magnetic fields (Salby, 1995).
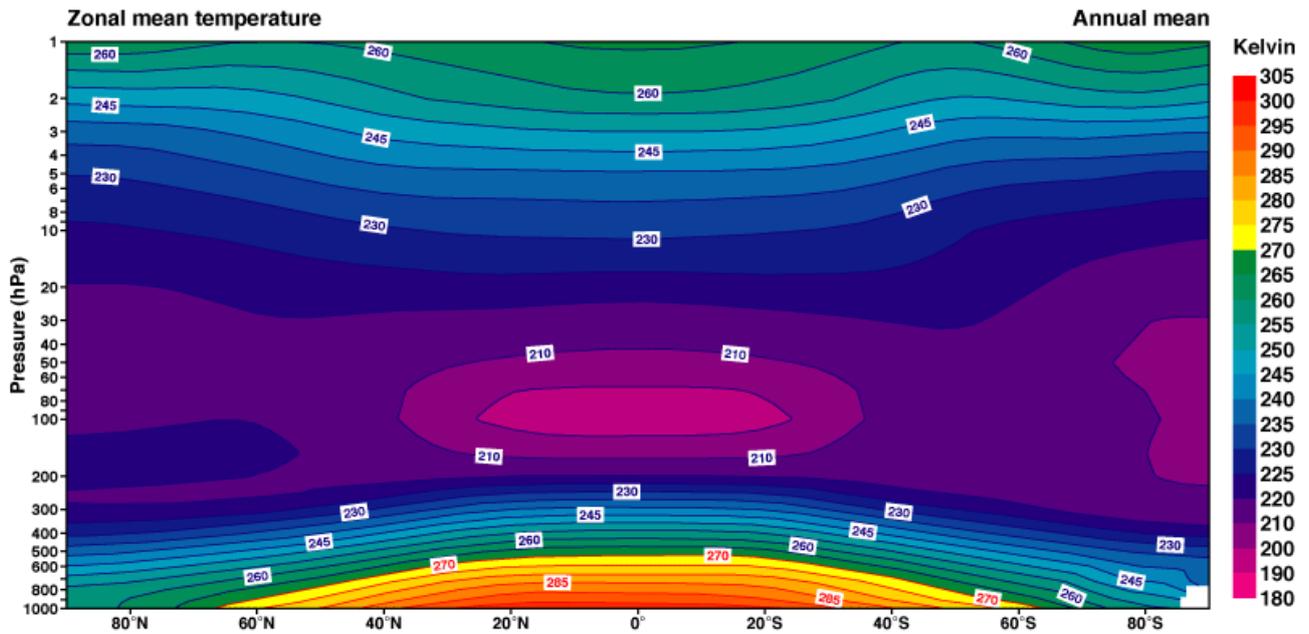
Figure 1. Zonal-mean temperature (annual mean) as a function of latitude and altitude.

Temperatures are warmest on the summer pole and steadily decrease to the coldest values on the winter pole. In the mesosphere, where the temperature decreases again with altitude, the horizontal thermal gradient is reversed. Temperatures are coldest on the summer pole, which is in perpetual daylight, and steadily increase to warmer values on the winter pole, which is in perpetual darkness. This peculiarity of the temperature distribution, contrary to radiative considerations, illustrates the importance of dynamics in establishing the observed thermal structure. The thermal structure in Figure 1 is closely related to the mean zonal circulation $\bar{u}$, which is shown in Figure 2 at the same time of year. In the troposphere, the circulation is characterized by subtropical jet streams, which strengthen with altitude up to the tropopause. These jets describe the circumpolar motion that occurs westward in each hemisphere (Salby, 1995).
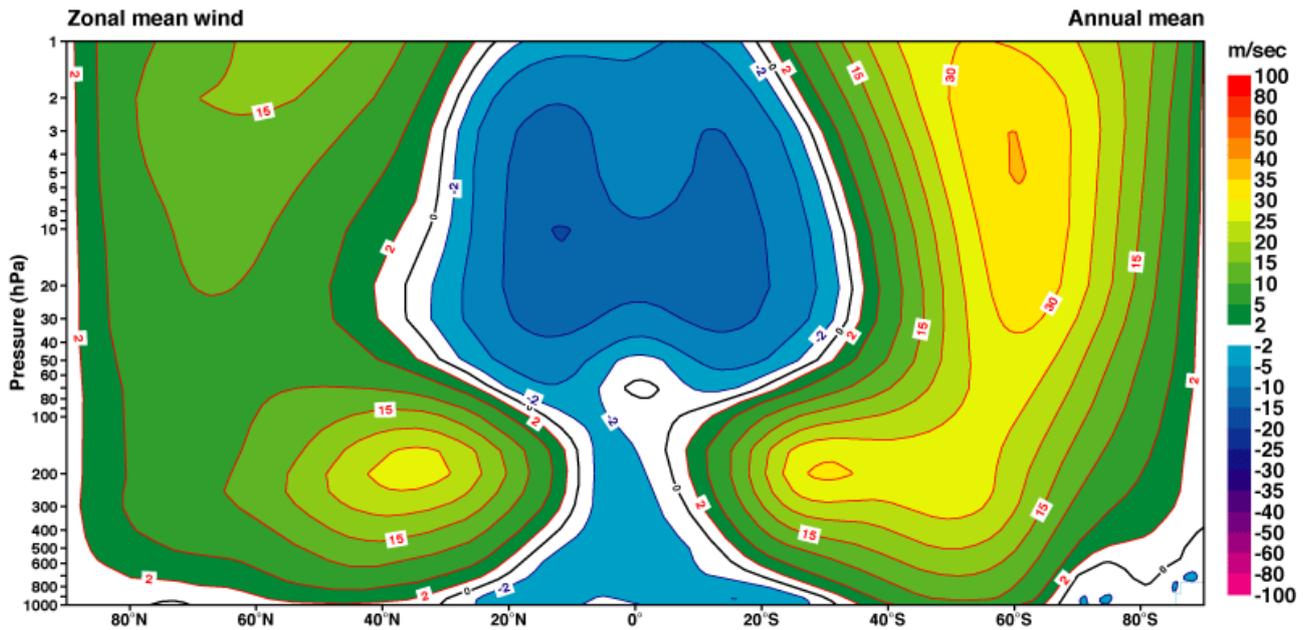
Figure 2. Zonal-mean wind (annual mean) as a function of latitude and altitude.

Above the subtropical jets, the zonal mean flow weakens with altitude and intensifies with opposite signs in the two hemispheres. In the winter hemisphere, westerly winds intensify above the tropopause in the polar night jet, reaching speeds of 60 m/s in the lower mesosphere. In the summer hemisphere, the westerly flow weakens above the tropopause and is then replaced by the easterly flow, intensifying up to the mesosphere. Reaching speeds somewhat higher than the mean zonal flow in the winter hemisphere, this easterly circulation merges with weak easterly winds in the tropical troposphere. On individual days, the circulation is more complex and involves much greater variability than represented in the zonal and time-averaged distributions in Figures 1 and 2 (Salby, 1995).

1.3 Water vapour in the atmosphere

Because it is continuously created in some areas and destroyed in others rather than being merely transferred by atmospheric motions, which would homogenize it, water vapour is a highly variable element. Circulation has a significant impact on this species and makes its distribution dynamic by moving it from its origin to its accumulation zones. Water vapour is the most significant trace species in the atmosphere because of its role in radiative processes, cloud formation, and energy exchanges with the oceans (Salby, 1995). According to latitude and altitude, Figure 3 depicts the zonal-mean distribution of water vapour (Oort and Pexito, 1983).
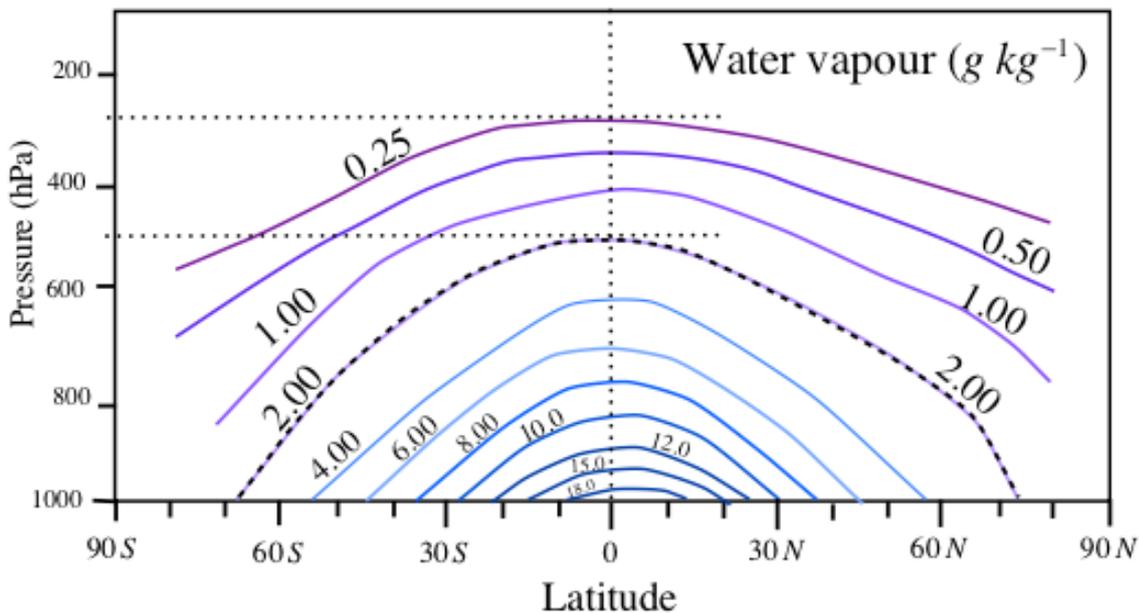
Figure 3. Variation of zonal mean water vapour mixing ratio with latitude and pressure. The shaded area represents 60% of the maximum value. Mixing ratios superior to 18 g/kg are at the surface and 8 g/kg at about 800 hPa (~ 2 km) in the Tropics as found by Oort and Pexito (1983).

Water vapour is confined almost exclusively to the troposphere. Its zonal mean mixing ratio $\bar{r}_{H_2O}$ decreases steadily with altitude, from a maximum of about 20 g/kg at the surface in the Tropics to a minimum of a few parts per million at the tropopause. The absolute concentration of water vapour absolute humidity, $\bar{\rho}_{H_2O}$ decreases with altitude even more rapidly. The density of the *i*-th constituent is simply its mixing ratio (r_i) multiplied by the density of dry air (ρ_d):

$$\rho_i = r_i \rho_d \qquad\qquad [1.1]$$

14

Since ρ_d decreases exponentially with altitude, water vapour tends to concentrate in the lowest 2 km of the atmosphere (Salby, 1995). The zonal-mean mixing ratio also decreases with latitude, falling below 5 g/kg poleward of 60°. These characteristics of water vapour reflect its production on the Earth's surface, redistribution by atmospheric circulation and destruction at altitude and middle and high latitudes through condensation and precipitation. Because of these production and destruction mechanisms and the rapid transport of air between source and sink regions, tropospheric water vapour is short-lived. A characteristic duration, which can be defined as the time required for r_{H_2O} within a single parcel to change significantly, is on the order of days. Every few days, a parcel of air encounters a warm ocean surface, where it absorbs moisture through evaporation, or a cloudy region, where it loses water vapour through condensation and precipitation. Most of the water vapour in Figure 3 originates near the equator, on warm ocean surfaces. Consequently, transport through the circulation plays a key role in determining the mean distribution $\bar{r}_{H_2O}$. Vertical and horizontal transport, called convection and advection respectively, each contribute to the redistribution of r_{H_2O} (Salby, 1995). Introduced to the surface of the tropical atmosphere, water vapour is transported upwards by deep convective cells and horizontally by large-scale vortices that disperse r_{H_2O} across the globe in a complex manner. Unlike the average distribution in Figure 3, which is quite uniform, the global distribution of water vapour on a single day is quite variable. The moisture pattern is granular in the Tropics, where water vapour has been moved vertically by deep convective cells that range in size from tens to a few hundred kilometres. At medium and high latitudes, the pattern is more fluid but still complex. Local abundance reflects the history of the air parcel residing at that location, i.e., where that parcel has been and what processes affecting water vapour have acted on it (Salby, 1995).

1.4 Ozone in the atmosphere

Ozone, consisting of three oxygen atoms, is a minor constituent (0.000004%) in the atmosphere and was first identified in the laboratory by Christian Fredrich Schönbein in 1840 and was first detected in the air by André Houzeau in 1858 (Brasseur, 2008). Figure 4 illustrates several important concepts concerning the ozone layer (Figure 4a) and its role in the stratosphere (the region between approximately 15 and 50 km altitude in which temperature rises with altitude, Figure 4b), the layer in which most of the ozone is confined. Ozone is a significant and naturally occurring greenhouse gas, similar to water vapour and carbon dioxide. It absorbs harmful ultraviolet radiation (UV), especially UV-B (280-315 nm) and UV-C (100-280 nm), acting as a heat source for the positive temperature gradient in the stratosphere. Ozone also emits thermal infrared radiation, which traps heat to warm the Earth's surface (Figure 4c). Two characteristics set stratospheric ozone apart from the so-called well-mixed greenhouse gases (WMGHGs):

1. Its relatively short chemical lifetime means that it is not uniformly mixed throughout the atmosphere and therefore its distribution is controlled by both dynamical and chemical processes. In fact, unlike the WMGHGs, ozone is produced entirely within the atmosphere rather than being emitted into it.

2. Second (Figure 4c), it is a very strong absorber of short-wavelength UV radiation (it is also a weak absorber of visible radiation). The ozone layer's absorption of this UV radiation leads to the characteristic increase of temperature with altitude in the stratosphere and, in consequence, to a strong resistance to vertical motion.

As well as ozone's role in climate, it also has more direct links to humans: this is the only gas in the atmosphere that absorbs UV-B radiation and protects much of Earth's biota from this potentially damaging short-wavelength radiation, helping maintain life on Earth. In contrast to the benefits of stratospheric ozone, high surface ozone values are detrimental to human health. The distribution of ozone in the atmosphere is maintained by a balance between photochemical production and loss, and by transport between regions of net production and net loss (IPCC/TEAP, 2005).
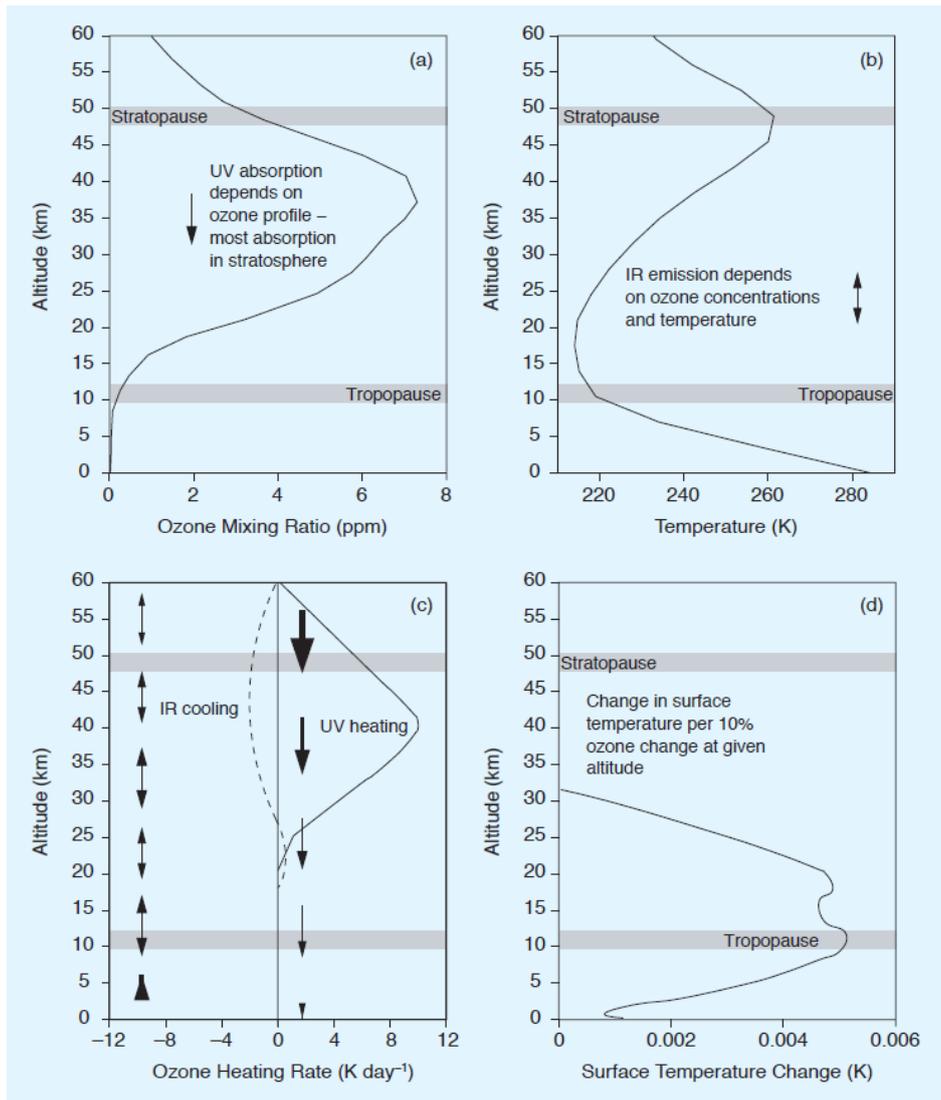
Figure 4. Vertical profiles of ozone-related quantities. (a) Typical mid-latitude ozone mixing ratio profile, based on an update of Fortuin and Langematz (1994); (b) atmospheric temperature profile, based on Fleming et al. (1990), showing the stratosphere bounded by the tropopause below and the stratopause above; (c) schematic showing the ultraviolet (UV) radiative flux through the atmosphere (single-headed arrows) and infrared (IR) emission at around 9.6 μm (the ozone absorption band, double-headed arrows), and the heating in the ultraviolet (solid curve) and infrared (dashed curve) associated with these fluxes; (d) schematic of the change in surface temperature due to a 10% change in ozone concentration at different altitudes (based on Figure 6.1 of IPCC, 2001) (IPCC/TEAP, 2005).

17

1.4.1 Stratospheric chemistry

Several chemical regimes can be identified for ozone. In the upper stratosphere, ozone distribution results from a balance between production through the photolysis of molecular oxygen and destruction through a series of catalytic cycles involving radical species of hydrogen, nitrogen, and halogens (IPCC/TEAP, 2005). Photochemical reactions dominate these processes. A photochemical model for the vertical distribution of ozone in the stratosphere was first formulated by Chapman, known as the Chapman mechanism (Chapman, 1930).

According to Chapman reactions, ozone is produced naturally by photolysis of molecular oxygen (O_2) at ultraviolet wavelengths below 242 nm:

$$O_2 + h\nu \rightarrow O + O \qquad [1.2]$$

The atomic oxygen produced in this reaction reacts rapidly with O_2 to form ozone (O_3),

$$O + O_2 + M \rightarrow O_3 + M \qquad [1.3]$$

where M is an inert air molecule stabilizing the reaction by removing excess energy. The photochemical production of ozone is balanced by its loss through the photolytic process.

$$O_3 + h\nu \rightarrow O + O_2 \qquad [1.4]$$

O_3 and O establish a rapid photochemical equilibrium through Reactions [1.3] and [1.4], and together are called "odd oxygen". Finally, in this sequence of reactions, ozone is removed by:

$$O + O_3 \rightarrow 2O_2 \qquad [1.5]$$

Destruction by reaction [1.5] alone cannot explain observed ozone abundances in the stratosphere and it is now known that, away from polar latitudes, the ozone production through the reaction [1.2] is largely balanced by destruction in catalytic cycles of the form [Cycle 1]:

$$XO + O \rightarrow X + O_2 \qquad [1.6]$$

$$X + O_3 \rightarrow XO + O_2 \qquad [1.7]$$

$$Net: O + O_3 \rightarrow 2O_2$$
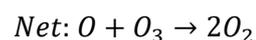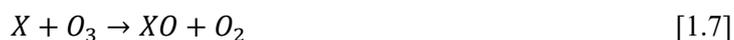
The net reaction is equivalent to the reaction [1.5]. Note that the cycle eliminates two molecules of "odd oxygen" since O and O_3 are in quick photochemical equilibrium and the loss of an oxygen atom effectively results in the loss of an ozone molecule as well. Furthermore, X, the catalyst, is not depleted during the reaction cycle. In the stratosphere, these cycles are particularly significant when reactive nitrogen (X = NO), halogen (X = Cl), and hydrogen (X = H, OH) radicals are present. Br-catalysed processes in the lower stratosphere also

contribute to ozone depletion. The frequency of these cycles, as well as the rate of ozone formation, significantly rises between 25 and 40 kilometres due to the large increase in O with altitude (IPCC/TEAP, 2005).

In the polar regions, the abundance of ClO increases during the winter, as a result of reactions on the surface of polar stratospheric cloud particles (IPCC/TEAP, 2005). However, the atomic oxygen, O, has very low concentrations, which limits the efficiency of Cycle 1. In this case, two other catalytic cycles become the dominant reaction mechanisms for polar ozone loss (Stolarski and Cicerone, 1974; Molina and Rowland, 1974). The first, the so-called ClO dimer cycle, is initiated by the reaction of ClO with another ClO [Cycle 2]:

$$ClO + ClO + M \rightarrow (ClO)_2 + M \qquad [1.8]$$

$$(ClO)_2 + hv \rightarrow ClOO + Cl \qquad [1.9]$$

$$ClOO + M \rightarrow Cl + O_2 + M \qquad [1.10]$$

$$2(Cl + O_3 \rightarrow ClO + O_2) \qquad [1.11]$$

$$Net: 2O_3 \rightarrow 3O_2 \qquad [1.12]$$

and the second, the ClO-BrO cycle, is initiated by the reaction of ClO with BrO [Cycle 3]:

$$ClO + BrO \rightarrow Cl + Br + O_2 \qquad [1.13]$$

$$Cl + O_3 \rightarrow ClO + O_2 \qquad [1.14]$$

$$Br + O_3 \rightarrow BrO + O_2 \qquad [1.15]$$

$$Net: 2O_3 \rightarrow 3O_2 \qquad [1.12]$$

The net result of cycle 2 and cycle 3 is the destruction of two ozone molecules and the production of three oxygen molecules. Both cycles are catalytic, as chlorine (Cl) and bromine (Br) are not lost during the cycles. Sunlight is needed to complete the cycles and to help maintain the high abundance of ClO. Cycles 2 and 3 are responsible for most of the ozone loss observed at the end of the winter-spring season in the Arctic and Antarctic stratosphere. At high ClO abundances, the rate of ozone destruction can reach 2-3% per day in late winter-spring. Outside the polar regions, the ClO-BrO cycle is of minor importance due to much lower ClO concentrations and the effect of the ClO dimer cycle is negligible, as the cycle is only effective at low polar temperatures.

1.4.2 Dynamical processes

The large-scale circulation of the stratosphere, known as the Brewer-Dobson (BD) circulation, systematically transports ozone poleward and downward. Using this mean meridional circulation, Brewer explained the low water vapour mixing ratios observed in the stratosphere (Brewer, 1949) and Dobson pointed out the high ozone concentration observed in the polar lower stratosphere (Dobson, 1956). The motion of BD circulation consists of an upward motion at the Tropics and a downward motion at the extratropic, together with an associated mass flux towards the pole. Photochemical ozone reactions occur rapidly in the sunlit upper stratosphere. Therefore, this transport has little effect on the distribution of ozone in that area, since ozone removal due to transport is rapidly replenished by photochemical production. However, this transport leads to significant changes in ozone in the extra-tropical lower stratosphere, where the photochemical relaxation time is very long (several months or more) and ozone can accumulate on seasonal time scales (Haynes, 2005). Due to the filtering by large-scale stratospheric winds, the vertical propagation of planetary waves in the stratosphere occurs mainly during winter, and this seasonality in wave forcing explains the winter maximum of the BD circulation. Furthermore, planetary waves are stronger (and more variable) in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH), due to the asymmetric distribution of surface features (topography and land-sea thermal contrasts) that, in combination with surface winds, force the waves. In the case of ozone, the BD circulation (together with the associated horizontal mixing) transports ozone poleward and downward and leads to a spring maximum in extratropical ozone abundance, as explained in Figure 5 (Andrews et al., 1987).
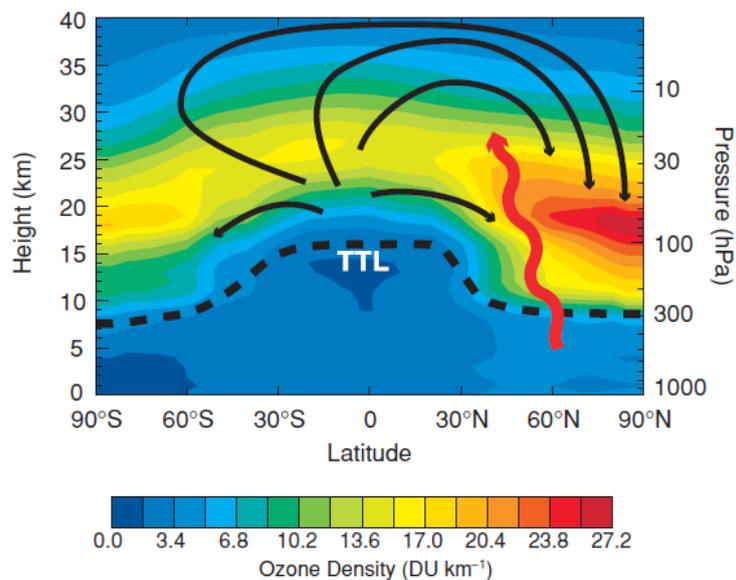
Figure 5. Meridian cross-section of the atmosphere showing ozone density (coloured contours; in Dobson units (DU) per km) during NH winter (January to March), from the climatology of Fortuin and Kelder (1998). The dashed line indicates the tropopause and TTL stands for the tropical tropopause layer. The black arrows indicate the Brewer-Dobson circulation during the NH winter and the wavy circulation. The red wavy arrow represents the planetary waves propagating from the troposphere to the winter stratosphere.

The distribution of the total ozone column (TOC), measured in Dobson Units (DU) is dominated by the distribution in the lower stratosphere and reflects its seasonality. Figure 6 shows the well-defined seasonal cycle, which increases with latitude and has the highest values in the NH at high latitudes, where TOC has values around 400 DU. At mid-latitudes (between 30° and 60° in both hemispheres), O_3 concentration reaches maximum values in spring and minimum values in autumn. In the equatorial region, the seasonal maximum is in September-October and the minimum in December-January. The mean diabatic circulation allows the transport of stratospheric O_3, which explains the seasonal variation of TOC. The seasonal variations of TOC are significant, reaching around 100 DU in the Northern Hemisphere (Bojilova et al., 2022).
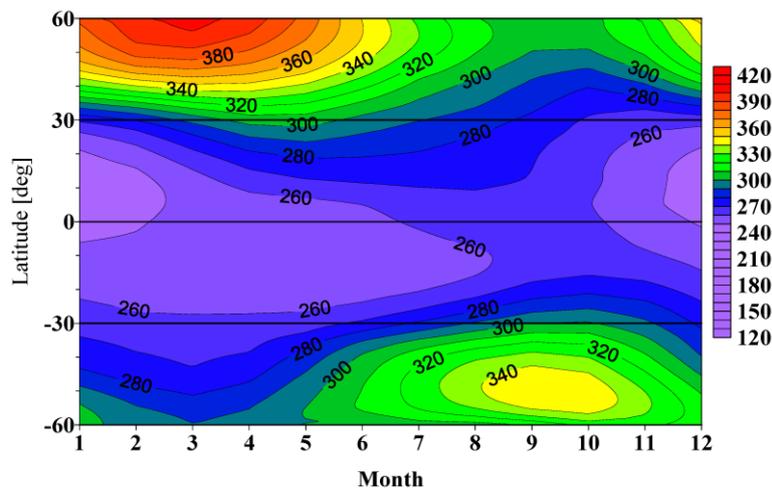
Figure 6. The climatological seasonal cycle of total ozone column (TOC) as a function of latitude and month for the period 2005–2020.

Changes in BD circulation also affect polar temperatures in the lower stratosphere (through vertical motions); increased wave forcing coincides with increased circulation and higher polar temperatures (and increased ozone transport). Since temperature influences ozone chemistry, the dynamic and chemical effects on ozone act in concert and are coupled.

1.4.3 Ozone depletion issue

The depletion of the stratospheric ozone layer has been one of the most important environmental issues of the last 40 years. Serious concern about ozone depletion began in the 1970s when it was realised that the breakdown of man-made compounds, such as CFCs, in the mid-stratosphere releases chlorine atoms capable of catalytically destroying the ozone layer (Molina & Rowland, 1974; Stolarski & Cicerone, 1974). Research activity increased dramatically after the discovery in 1985 of a large and unexpected ozone depletion in the Antarctic lower stratosphere during the spring, the so-called Antarctic ozone hole (Farman et al., 1985). This depletion was caused by increased levels of chlorine and bromine in the atmosphere but, more importantly, by the conversion of stable chlorine species into active forms that destroy ozone on the surface of polar stratospheric clouds that form in winter and spring (WMO, 2014). Atmospheric scientists were unable to predict the ozone hole in advance because the models used to predict the evolution of the ozone layer did not include these processes (Solomon et al., 1986). At the time of the discovery of the Antarctic ozone hole, the process of international protection of the ozone layer had already been initiated and the framework for its implementation was established with the signing of the Vienna Convention in 1985. Furthermore, the 1987 Montreal Protocol on Substances that Deplete the Ozone Layer was a major achievement in terms of global environmental protection, although it initially placed only modest limits on the production and consumption of key ODS, such as CFCs and halons containing bromine. This protocol was further strengthened through subsequent amendments and adjustments, which over time led to an almost complete ban on the main classes of ODS, including CFCs, substituted hydrochlorofluorocarbons (HCFCs) and related compounds such as methyl chloroform and carbon tetrachloride. These compounds have atmospheric lifetimes on the order of 10-100 years (Ko et al., 2013) and thus the response of atmospheric chlorine and bromine loading to changes in emissions is slow. However, observations show (WMO, 2014) that the abundance of these gases in the lower atmosphere is largely meeting the Montreal Protocol limits, as expected, and most of the major ODS are decreasing, as shown in Figure 7.
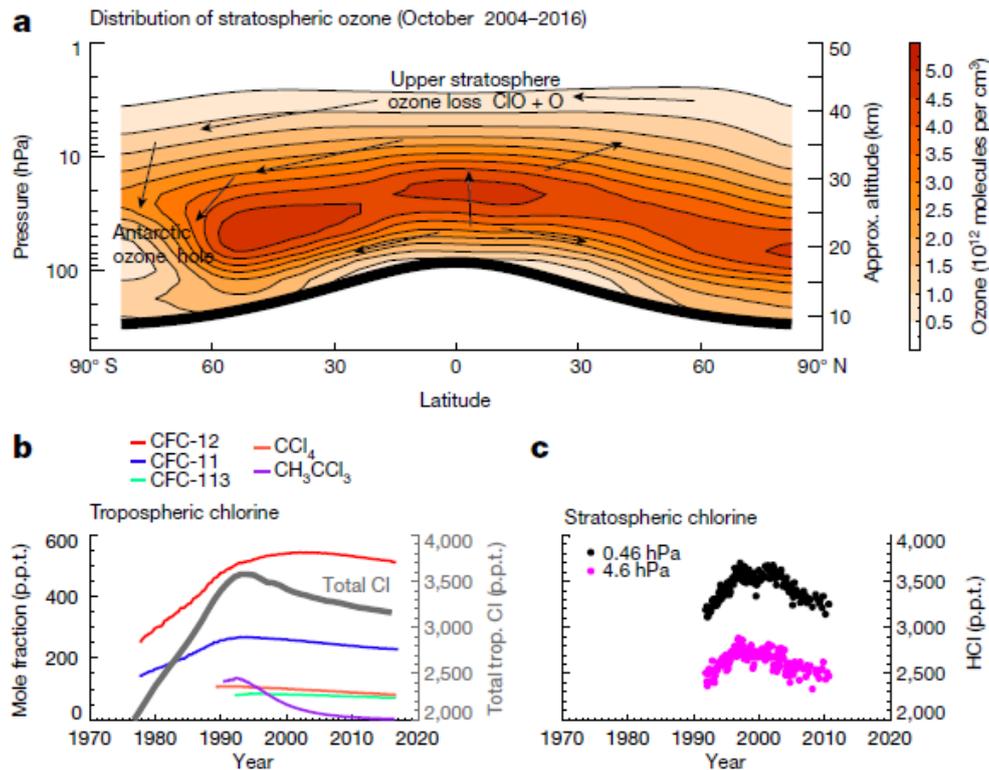
Figure 7. Latitude–height cross section of stratospheric ozone and time series of chlorine in the troposphere and stratosphere. Panel A: October average (2004-2016) ozone concentration (1012 molecules per cm3) detected by the Microwave Limb Sounder (MLS) instrument on board the Aura satellite. The thick black lines indicate the position of the climatological tropopause. The annotated text shows the main regions where ozone is most severely destroyed by halogens. The black arrows indicate the Brewer-Dobson circulation, which transports air upwards in the Tropics, towards the poles and downwards at high latitudes, with more intense transport towards the winter pole. Panel B: Monthly average observed surface mole fraction (parts per trillion, ppt) of selected ozone-depleting substances (left axis; see legend for colour coding) from the National Oceanic and Atmospheric Administration (NOAA), a long-term monitoring programme. The thick grey line shows the evolution of total tropospheric chlorine (includes contributions from other halocarbons, e.g., HCFCs) from the World Meteorological Organisation (WMO) A1 scenario (right axis). Panel C: Time series of monthly mean HCl (p.p.t.) in the tropical upper stratosphere (4.6 hPa, approx. 40 km, pink) and lower mesosphere (0.46 hPa, approx. 55 km, black) from GOZCARDS satellite measurements. HCl is a breakdown product of chlorine-containing ODS and increases with altitude.

The sum of tropospheric chlorine peaked in 1993 (O'Doherty et al., 2004; Montzka et al., 1999), while the sum of tropospheric bromine peaked a few years later in 1997 (Montzka et al., 2003). The stratospheric abundance of chlorine (found largely in the form of HCl in the upper stratosphere, Michelsen et al., 1996) and bromine, which are derived from these ODS, has followed these tropospheric variations, but with a lag of about 3-7 years (depending on the region) due to the slow transport and degradation of ODS through the stratosphere (Newman et al., 2007).

1.4.4 Recovery of the ozone layer and trends

The variability of other factors influencing ozone, such as stratospheric dynamics (wind and temperature), aerosol loading and solar irradiance mask the small sign of ozone recovery due to ODS reduction, which is expected to be about a couple of percentage points per decade globally. An important question is therefore to what extent and where ozone recovery can be detected. Another important question is the ultimate extent to which the ozone layer will recover, given the increasing impact of climate change on the structure and composition of the atmosphere. The continuous increase in CO_2 is altering the physical structure of the atmosphere: the tropopause is rising and the stratosphere is thinning; its thermal structure is changing and the Brewer-Dobson circulation in which air is transported into and through the stratosphere (see Figure 5) may accelerate in the future (Butchart et al., 2010). As a result, stronger upwelling in the Tropics and faster downwelling in the mid-latitudes and polar regions might occur. As ozone abundance increases with altitude in the lower stratosphere (where most of the ozone is found), these changes in circulation would lead to a decrease in ozone in the Tropics and an increase at higher latitudes. Cooling in the upper stratosphere is already increasing ozone in that region, slowing down the gas phase ozone destruction cycles (Haigh & Pyle, 1982; Jonsson et al., 2009). All these changes are driven fundamentally by the increase in GHG, particularly CO_2. At the same time, N_2O and CH_4 levels are also increasing, with CH_4 being particularly sensitive to changes in emissions from natural and anthropogenic sources (Nisbet et al., 2016). The balance between the various catalytic cycles that result from the degradation of ODS, CH_4 and N_2O and lead to ozone loss is also set to change. What is clear is that the chemistry and dynamics of the stratosphere will have changed sufficiently to make the recovery of pre-depletion ozone levels meaningless. The picture becomes even more complicated if we consider how recovery occurs in different parts of the atmosphere or different seasons. However, possible changes in non-stratospheric factors such as tropospheric ozone, clouds, aerosols and terrestrial albedo, along with their great variability, make it difficult to draw meaningful conclusions about surface UV-B recovery (McKenzie et al., 2007). This is an even more complex picture than that of ozone. From a regulatory point of view, since the control of production and consumption is the tool that policymakers can use, the success of the Montreal Protocol is mainly judged by changes in atmospheric ODS concentrations. From this point of view, the Montreal Protocol is already undoubtedly a success; ODS levels are decreasing (Montzka et al., 1999) with expected benefits for ozone and UV-B radiation as well as climate. However, the impact of these ODS decreases on ozone levels has proven much more difficult to detect. Definitions of ozone recovery tend to be based on the concept of the state or phase reached. Since recovery is often defined concerning the effect of ODSs (the key driver of the Montreal Protocol), each stage requires a clear attribution of ozone changes to the decline and ultimately the return of ODS to pre-industrial levels (WMO, 2007). The following stages (or fingerprints) of recovery have been defined (WMO, 2007):

1. a significant slowing of stratospheric ozone decline;
2. then the onset of a significant increase;

3. and finally, the full recovery of ozone from ODSs, when ozone is no longer significantly affected by them.

However, it is more beneficial to think of recovery as the direction of travel rather than the destination. Indeed, full recovery does not necessarily imply a return of stratospheric ozone to pre-1980 levels because the influence of other factors, in particular rising levels of GHGs, is growing. ODS levels in the atmosphere are decreasing (see Figure 7) and the first phase (or "fingerprint") of the ozone response, the end of ozone decline, has been observed (WMO, 2014; Newchurch et al., 2003). However, it has been difficult to establish the occurrence of the next phase, i.e., a general upward trend in ozone due to decreasing ODS. This may be surprising since ODS levels have been declining for 15-20 years now. However, due to the long atmospheric lifetime of ODS (typically many decades, Ko et al, 2013), this decline is approximately three times slower than their rapid increase before the entry into force of the Montreal Protocol. Then it took 10-15 years to detect the significant decrease in global ozone. All things being equal, we might expect 30-40 years to pass before it becomes possible to detect a significant upward trend in global ozone due to declining ODS levels (Weatherhead et al., 2000). Figure 8 shows an overview of ozone profile trends from past and recent assessments: WMO (2014), Harris et al. (2015), Steinbrecht et al. (2017), and LOTUS (Petropavlovskikh et al., 2019).
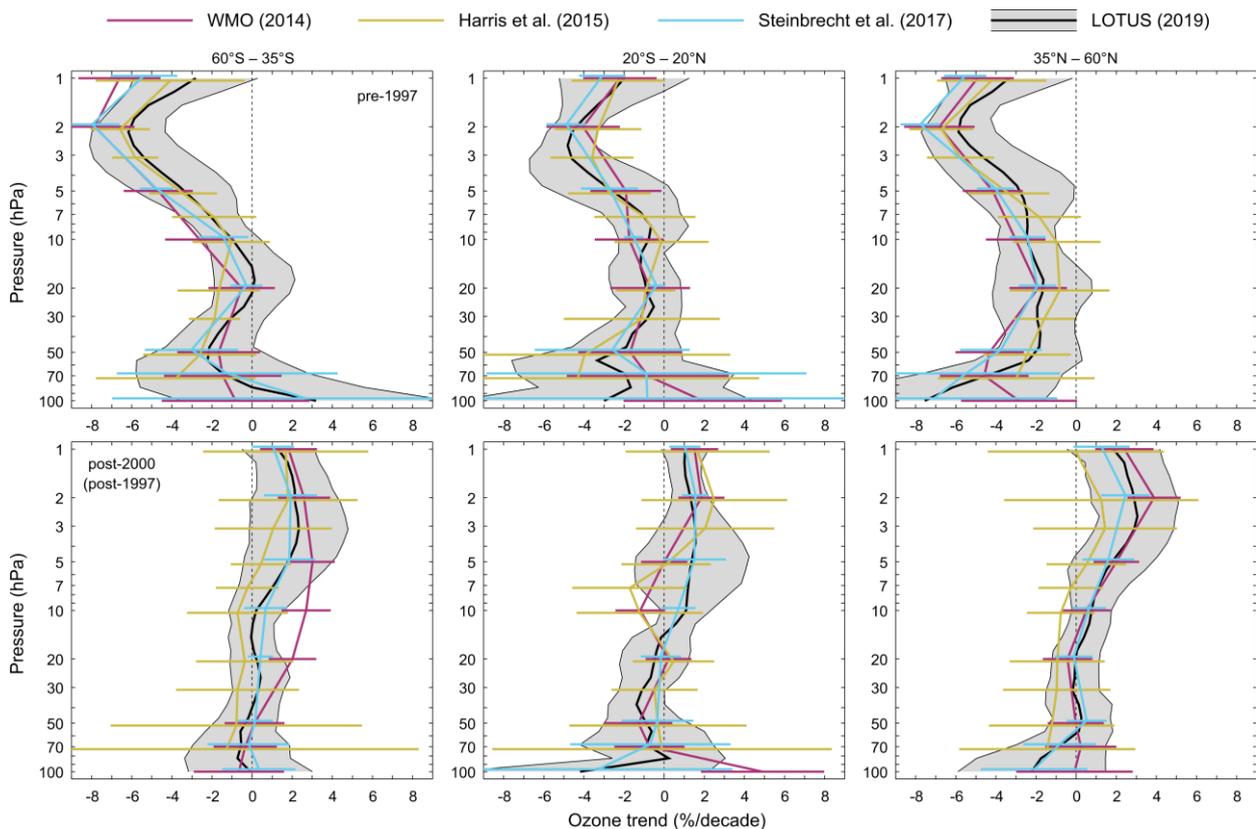
Figure 8. Overview of ozone profile trends from past and recent assessments: WMO (2014), Harris et al. (2015), Steinbrecht et al. (2017), and LOTUS (Petropavlovskikh et al., 2019) are shown in red, orange, blue, and black respectively. The top row shows trends before the turnaround of ODSs and the bottom row since the turnaround (the analysis period differs by assessment). The shaded area and error bars represent the 95 % confidence interval for the combined trend. Coloured profiles are slightly offset on the vertical axis for display purposes.

1.4.4.1 pre-2000 trends

Negative trends are found throughout almost the entire stratosphere in the pre-2000 period for almost all satellite and ground data. Individual and combined satellite data show highly statistically significant evidence of declining ozone concentrations in the upper stratosphere (at altitudes above the 10-5 hPa level) from the mid-1980s through the 1990s. Depletion reaches a maximum rate near 2 hPa (~42 km) of 5.9–6.2% per decade in the mid-latitudes and 4.8% per decade in the Tropics (TR) (Petropavlovskikh et al., 2019). Rates of ozone decline in the middle stratosphere (30–15 hPa) are considerably lower, with statistically insignificant values at most 1–2% per decade. Negative trends are found in the lower stratosphere (up to 50 hPa), while in the lowermost stratosphere (up to 100 hPa) trends differ depending on latitude, with large negative trends significant at around 5% per decade in the NH (Petropavlovskikh et al., 2019). However, confidence in trend results is reduced in the lower stratosphere due to large natural variability, low ozone values and reduced sensitivity of satellite observations. Trends derived from ground-based measurements generally confirm the results of satellite trends. However, due to their greater dispersion in space and time, especially during this first period, the significance of the trends is not so high and the trend values differ. The results agree well with those of model simulations (within 1% per decade) across the middle and upper stratosphere at all latitudes, giving confidence that these ozone losses were the result of the chemical forcing of ODSs as predicted by the model. However, larger differences exist between satellite and model results in the lower stratosphere, with disagreements outside of large uncertainties in the SH (Petropavlovskikh et al., 2019).

1.4.4.2 post-2000 trends

Positive trends are found throughout the upper stratosphere and parts of the middle stratosphere in the post-2000 period for both satellite and ground-based trends, although results vary for ground-based data depending on the observation technique. Satellite results show statistically significant positive trends in the mid-latitude NH of 2–3% per decade in the upper stratosphere (between ~5–1 hPa) and 1–1.5% per decade in the TR (between ~3–1 hPa). Positive trends of approximately 2% per decade are also found in the SH near 2 hPa in mid-latitudes, although statistical confidence is lower (Petropavlovskikh et al., 2019). In the mid-latitudes of NH, trends in the upper stratosphere are significantly lower than 4 hPa. At altitudes below 4 hPa, mid-latitude trends are no longer statistically significant, going from a positive value of 1.8% per decade at 5 hPa to almost zero between 50 and 20 hPa (Petropavlovskikh et al., 2019). In the TR, trends become negative below 15 hPa although even estimates of 0.5–1.5% per decade are statistically insignificant (Petropavlovskikh et al., 2019). Generally, these satellite-based results are in agreement with ground-based observations and model simulations. Persistent negative trends in the middle and lower stratosphere over the TR are likely the consequence of radiative forcing and greenhouse gas dynamics as predicted by the model (WMO, 2014). The

derived trends differ markedly in the lower stratosphere, below 50 hPa, depending on the dataset and latitude. For example, satellite-based results show statistically insignificant negative trends (or close to zero in the SH) and ground-based trends agree in sign, except in the TR where there are significant positive trends. Model simulations, however, predict positive mid-latitude trends in both hemispheres over this altitude range.

2. Temperature, water vapour, and ozone in situ measurements

This section provides a description of the temperature, water vapour, and ozone measurement techniques used to investigate the UT/LS.

2.1 Temperature, water vapour and wind radiosondes

Measurements of temperature, relative humidity, wind speed, direction, and pressure, belong to the basic input quantities to analyse and describe actual weather patterns and serve as input for the initialization of short- and medium-range operational numerical weather forecast models, as well as to study climate trends and variability. These variables are measurable simultaneously by radiosondes.

The basic concept of radiosounding measurements is to use sensors designed for in-situ measurements in direct contact with the air and immediately transmit the measured values to a ground station by radio. Radiosondes are attached to a flying balloon, carrying it through the atmosphere. The data are then disseminated through the GTS (Global Telecommunications System), nowadays renamed WIS (WMO Information System). The sonde and sensor technology has developed continuously since the first working radiosondes were developed in the late 20s of the last centuries. For many decades, radiosondes have been launched worldwide by more than 1200 stations and internationally coordinated in daily routine, primarily by national weather services. Operational soundings reach heights of 20 km to 35 km, depending on balloon size and local atmospheric conditions. Different sonde types from manufacturers in about 10 countries are nowadays in use, and provide data with appropriate quality, primarily for weather applications. The near real-time data availability, high vertical resolution, high potential for automatization, as well as comparatively simple applicability and moderate costs, make radiosondes an essential data source for that purpose.

Beyond the weather-related applications which primarily lean on tropospheric observations, radiosondes supply valuable data from the higher altitude range of the upper troposphere through the tropopause into the lower stratosphere (UT/LS). Upper-air radiosounding measurements are still considered one of the most relevant anchor points for atmospheric reanalysis (Hersbach et al., 2020) and have been largely used to investigate climate trends (e.g., Haimberger et al., 2012; Madonna et al., 2022). However, the extremely dry and solar radiation-influenced conditions and the low pressure at altitudes above the tropopause

are still a challenge for radiosonde sensor technologies. This issue is overcome by reference measurements, such as those made by GRUAN, the GCOS Reference Upper-Air Network, as may adjusted using statistical methods or post-processing the data when metadata are available.

To describe in detail a radiosonde sensor, the next paragraph is focused on the Vaisala RS41 radiosonde, one of the most modern in-situ upper-air sensors, which has also been operated by the author of this thesis during routine measurements.

The RS41 radiosonde consists of a sensor boom connected to an electronic board, which is powered by an attached Li-battery pack, and mechanically and thermally protected by an EPS foam housing. The antenna, connected to the opposite end of the board, consists of a trailing thin wire (see Figure 9). The radiosonde is suspended to the unwinder string in such a way that the sensor boom points sideways up from the body when the radiosonde is ascending. The standard unwinder is equipped with a 55 m long thin non-UV treated polypropylene string.

Figure 9. Vaisala Radiosonde RS41-SG.

Except for the humidity sensor chip, the entire sensor boom including the temperature sensing element is covered with a highly reflective aluminium coating to minimise undesired thermal effects, especially warming by short-wave solar radiation or IR radiative cooling. The coating also has hydrophobic properties. Especially in terms of the temperature sensor element, this lowers the risk of wetting during the passage of water clouds and subsequent temperature 'contamination' due to evaporative cooling. During soundings, the boom is bent at an angle of about 45° away from the sonde body, with the sensitive side of the humidity chip pointing downwards against the direction of the airflow. The angle is intended to optimise the measurement performance of two opposing effects:

1. On the one hand, the surface of the sonde housing as seen from the sensors should be as low as possible to minimise radiative effects that are connected with the sonde housing.
2. On the other hand, effective ventilation of the sensors should be ensured especially for humidity.

The housing of the radiosonde is approximately 15 cm long, 6 cm wide, and 4.5 cm thick. It contains the main integrated board of the radiosonde, as well as the battery pack and the optional pressure sensor board. The data logger, radio transmitter and antenna, and the GPS antenna and receiver are directly soldered to the main integrated board of the radiosonde. The sensor boom is attached to the main board through a detachable connector.

The Vaisala RS41 temperature sensor is designed and manufactured in-house and uses resistive platinum technology (Pt1000). The measurement signal is the electrical resistance which increases linearly with temperature. The sensor element has dimensions of about 5 mm in length and 0.25 mm in thickness. Uncertainty related to non-linearity of 0.05 K (k = 2) is stated, with no systematic bias in sensor calibration. During these tests, the uncertainty of the reference temperature measurement was 0.04 K (k = 2). The results are valid for the temperature range of −98 ∘C to 39 ∘C and indicate a strong linearity.

The humidity sensor is based on the widely used capacitive polymer technology. It is developed by the manufacturer and produced in its clean-room facilities. Capacitive humidity sensors essentially consist of a hydrophilic porous polymer material that is arranged as a thin film (thickness d) between two flat electrodes (surface area A) which form a plate capacitor. These layers are arranged on a glass or ceramic substrate. The polymer exchanges water molecules with the surrounding moist air by diffusion. Depending on the amount of adsorbed water, the effective dielectric constant ε_r of the polymer and therefore the capacitance changes sensitively due to the high relative permittivity (or dielectric constant) of water. This capacity change can be measured as frequency change which serves as a measure for the water vapour content in the air. The capacitance of the polymer measures the relative humidity rather than the absolute water content because the adsorption-desorption equilibrium changes with relative humidity:

$$C = \varepsilon_r \frac{A}{d}.$$
[2.1]

The humidity sensor of the RS41 radiosonde is permanently heated so that its temperature T_i is kept above the ambient air temperature T_a at a constant offset of about 5 K. T_i is measured by a separate temperature sensor. Both the resistance heating element and the temperature sensor are integrated into the humidity sensor chip next to the polymer. T_a is measured by the Pt1000 sensor near the tip of the sensor boom. The water vapour partial pressure directly above the surface of the humidity sensor, $e_i = U_i \cdot e_s(T_i)$, with $e_s(T_i)$ the saturation vapour pressure, cannot directly be measured. However, by definition, there is saturation at the dew point, i.e., $e_d = e_s(T_d)$, and the dew point T_d of an air parcel does not change with temperature as long as there are no phase transitions. That is, with the definition of relative humidity,

$$C = \varepsilon_r \frac{A}{d}. \tag{2.2}$$

$$e_s(T_d) = U_i \cdot e_s(T_i) \tag{2.3}$$

The same applies to the environment:

$$e_s(T_d) = U_a \cdot e_s(T_a) \tag{2.4}$$

The relative humidity in the ambient air is therefore

$$U_a = U_i \frac{e_s(T_i)}{e_s(T_a)} \tag{2.5}$$

With the measured values for T_i and T_a and the measured relative humidity above the heated polymer U_i can be directly converted into the relative humidity for the ambient air without the need to determine the dew point T_d Vaisala uses the ITS-90 compatible form of Wexler's formula by Hardy (1998) for the calculation of the saturation vapour pressure $e_s(T_i)$ and $e_s(T_a)$. This method ensures that the relative humidity detected by the sensor is always kept below the saturation level at which the polymer is potentially prone to 'contamination' with liquid water, and where the calibration is most difficult. Further advantages are the performance improvement at low ambient temperatures concerning the response time, as well as less susceptibility to direct solar irradiation (no dry bias).

Wind measurements are instead derived from the GPS positioning that is employed by the RS41/SPS311 system, consisting of a base GPS antenna and receiver at the launch site and a rover GPS antenna and receiver onboard the radiosonde. The default GPS setup is designed to provide differential corrections for the positioning of the radiosonde.

Finally, pressure measurements may also be available. The pressure sensor directly measures the atmospheric pressure at the current height in a sounding. The RS41 is equipped with a shock-resistant single-crystal capacitive silicon sensor element. In the case of radiosonde models without the pressure sensor, the pressure value is derived from the GPS altitude. Below, in Figure 10 an example of a vertical profile of temperature (panel a) and relative humidity (panel c) with the related uncertainties (panels b and d respectively).
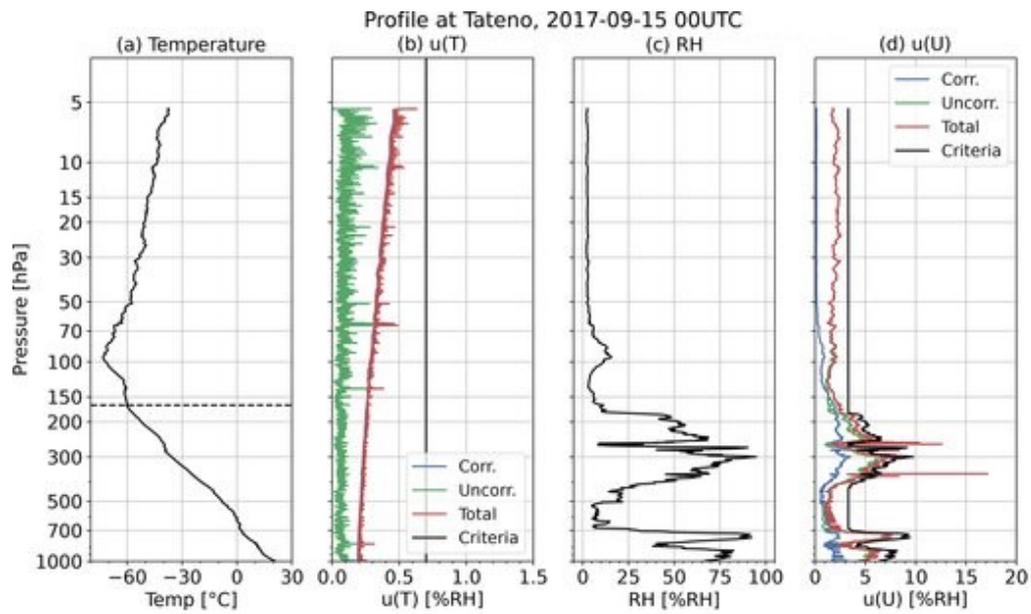
Figure 10. The profile of (a) temperature, (b) temperature uncertainty, (c) RH, and (d) RH uncertainty at 00:00 UTC on 15 September 2017 for Tateno station. The dotted line in panel (a) is the height of the tropopause. The blue, green, and red lines in panels (b) and (d) represent correlated, uncorrelated, and total uncertainty, respectively. The black lines in panels (b) and (d) show the criteria for screening based on the idea that data with uncertainties exceeding the thresholds (shown in Hoshino et al., 2022) are of questionable reliability and need to be verified individually.

An example of monthly temperature anomalies for the period 1980–2018 of the reanalyses (ERA5, ERA-Interim, MERRA-2 and JRA-55), for the RAOBCORE and RICH datasets for the 10° N–10° S region are shown in Figure 11 at pressure levels from 100 hPa to 10 hPa.
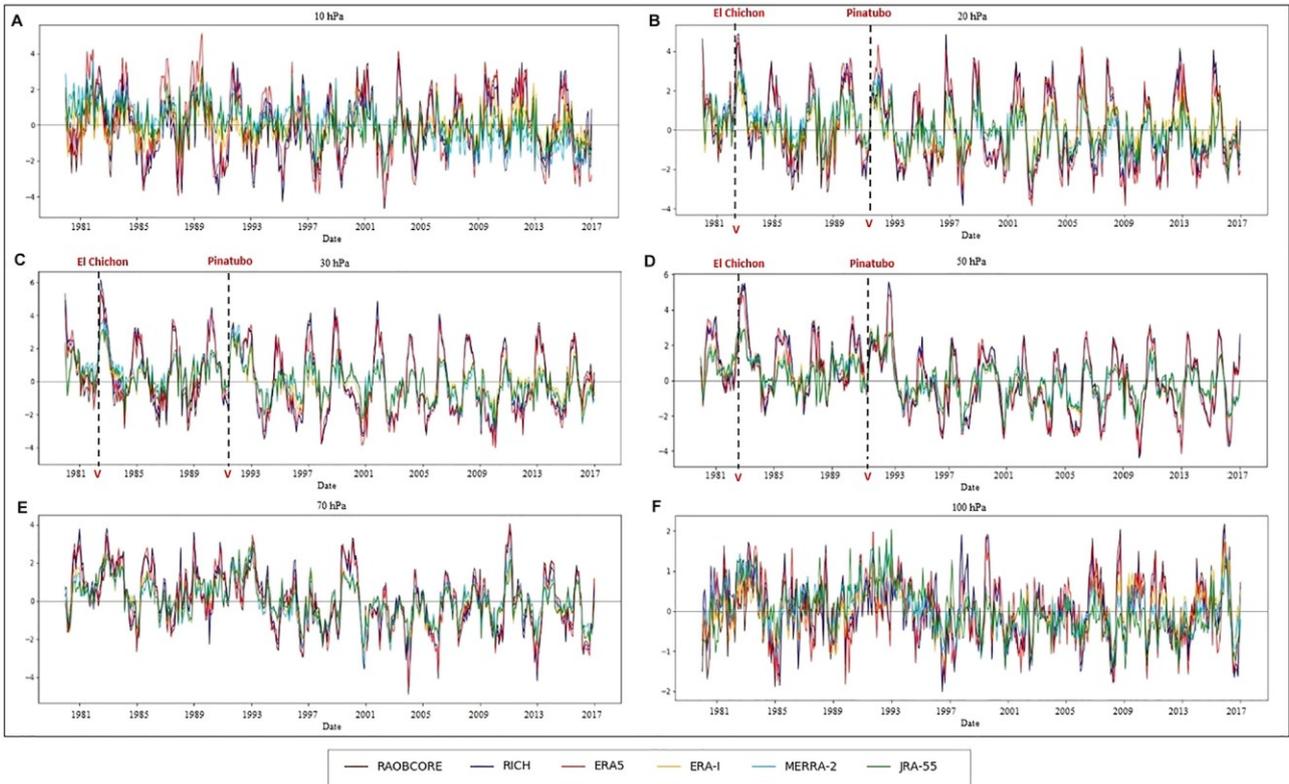
Figure 11. Monthly temperature anomalies at the region 10°S - 10°N of the radiosonde observations (RAOBCORE and RICH) and the four reanalysis datasets (ERA5, ERA-Interim, MERRA-2, and JRA-55), for the vertical levels (A) 10 hPa, (B) 20 hPa, (C) 30 hPa, (D) 50 hPa, (E) 70 hPa, and (F) 100 hPa, over the period 1980–2018.

The well-known variability associated with major volcanic eruptions (El Chichon 1982; Pinatubo 1991) is evident in all data sets, particularly between the 20–30 hPa levels. Among the reanalyses considered, ERA5 shows the lowest differences compared to the radiosonde data, while a notable difference in magnitude, in the order of 4 K, is found for ERA-Interim, MERRA-2 and JRA-55. For 10 hPa and 100 hPa levels, the datasets tend to show slightly larger differences than observations within a few years, within ±4 K and ±2 K, respectively. However, similar patterns are produced in all datasets and this difference decreases after 2000. It is worth mentioning that this comparison is not completely independent because RAOBCORE adjustments were used for radiosonde bias correction in ERA-Interim, MERRA-2 and JRA-55 and RICH is used in ERA5. Differences between reanalysis and RAOBCORE/RICH are likely due to differences in dataset characteristics (e.g., assimilation schemes, the model used, dataset resolution, spatial sampling, etc.). Furthermore, it is especially noteworthy that the resulting anomaly patterns for all considered reanalyses reflect a reliable representation of the atmospheric circulation in the tropical stratosphere, where the temperature is dominated by the QBO signature in this atmospheric region; positive temperature anomalies correspond to westerlies and negative temperature anomalies correspond to easterlies (Pascoe et al., 2005; Yamazaki et al., 2020).

Regarding the uncertainty of the radiosondes, in general, this information is not provided, however, the GRUAN network, presented in this thesis in Chapter 5, provides radiosounding profiles with uncertainties. Furthermore, in Chapter 5, the RHARM network will also be presented, and how the uncertainty is estimated will be discussed (Dirksen et al., 2020).

2.2 Ozonesondes

Ozonesondes are small, lightweight, and compact balloon-borne instruments, developed for measuring the vertical distribution of atmospheric ozone in situ from the ground up to an altitude of about 30-35 km, and ozonesonde records at several measurement stations provide the longest ozone profile time series available, with some starting in the 1960s. During normal flight operation, an ozonesonde is coupled to a standard meteorological radiosonde for transmission to the ground of the measured quantity plus additional meteorological parameters of pressure, temperature, relative humidity, and, since the adoption of radiosondes with GPS, drift-based wind direction and wind speed, and GPS-derived altitude. Using the radiosonde telemetry, the current measured by the ozonesonde is transmitted to the ground station for data processing (Tarasick et al., 2019). The ozone profiles are obtained with an effective resolution ranging from approximately 100m to 150m, while the SHADOZ profiles have been reprocessed at 50m (Thompson et al., 2019).

2.2.1 System

Although other types of ozonesonde instruments have been used in the past (Tarasick et al., 2019b), almost all ozonesounding stations worldwide now use the ECC ozonesonde type, developed by Komhyr (1969) shown in Figure 12, archiving 2,000–3,000 profiles annually (Tarasick et al., 2019b).
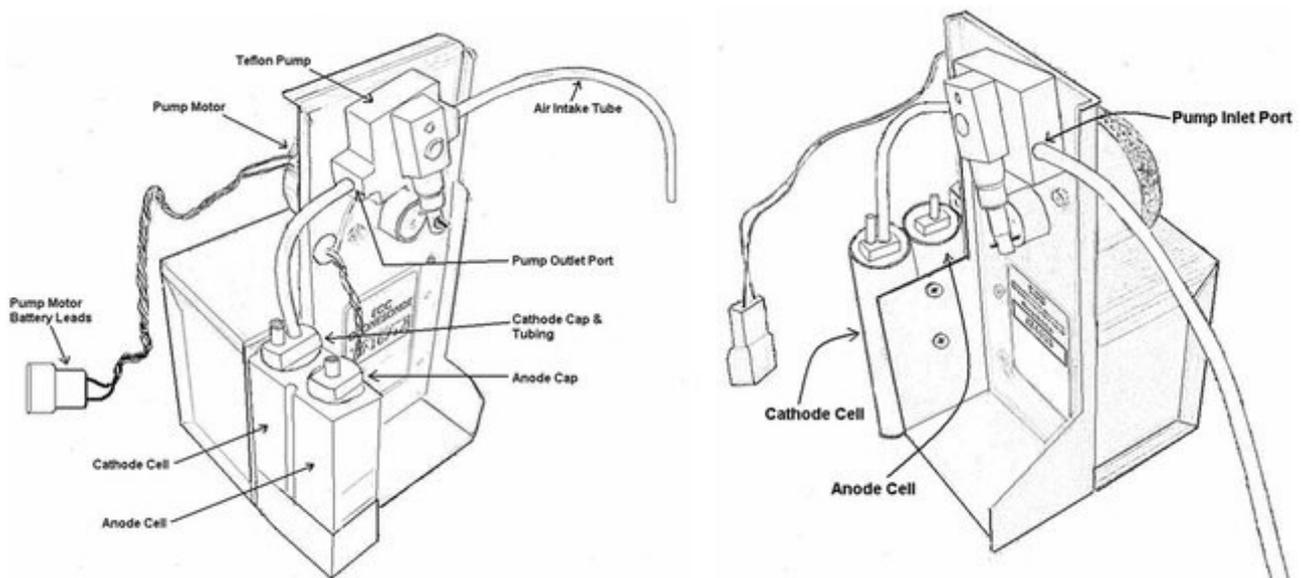
*Figure 12. Two ECC ozonesonde instruments, made by different manufacturers. Left: SPC-6A type made by Science Pump Corporation; Right: ENSCI-Z type made by ENSCI Corporation. Differences are minor: the cathode and anode caps, thermal contact between the cells and the metal housing, and the plastic used to construct the cells and pump (Teflon and moulded plastic, respectively) (Tarasick et al., 2019).*

This sensor measures ozone using iodine/iodide electrode reactions. Two platinum electrodes are immersed in separate cathode and anode chambers of differing concentrations of potassium iodide (KI) solution. The anode cell contains a solution saturated with KI. Both cells contain an equal concentration of potassium bromide (KBr) and a phosphate buffer to maintain a neutral pH. An ion bridge (made typically of densely packed cotton fibres) connecting the two chambers allows ions to flow between the two cells but prevents mixing, thereby preserving their respective concentrations. Ambient air containing ozone is pumped into the cathode cell and reacts with iodide (I^-) in solution to form iodine (I_2). To maintain electrochemical equilibrium, I_2 is converted back to I^- on the platinum electrode resulting in the release of two electrons. Thus, each ozone molecule entering the sensor causes two electrons to flow through the ECC's external circuit, which it measures as a current. The resulting electrical current is proportional to the amount of ozone in the sampled air. The electrochemical technique assumes no secondary reactions take place and a 1:1 stoichiometric relationship of the O_3:I_2 is maintained (Sterling et al., 2018).

2.2.2 Retrieval

The ozonesonde equation for calculating the ozone partial pressure (P_{O3}) is determined by Faraday's first law of electrolysis and the ideal gas law and shown below:

$$P_{o3} = \frac{R}{2F}(I_M - I_{BG})\frac{1}{\Phi_P}T_P\frac{1}{\eta_{OS}}$$

[2.6]

The first term is an empirical constant where R is the universal gas constant and F is the Faraday's constant. The two in the denominator represent the two electrons being delivered to the electrical circuit of the sensing cell for every ozone molecule reacted, assuming a 1:1 stoichiometry. The remaining variables in the equation are the measured cell current (I_M) and the background cell current (I_{BG}) in micro amps, the pump flowrate (Φ_P) in cm^3/s, the pump temperature (T_P) in degrees Kelvin, and the ozone sensor efficiency (η_{OS}). The cell currents, the pump flow rate, and the pump temperature can be measured directly and independently. The ozone sensor efficiency (η_{OS}) is a measure of how efficiently gaseous ozone molecules bubbled through the ozone sensor are converted to electrons and cannot be measured directly (Tarasick et al., 2019). Instead, it is measured by comparison to the reference ozone photometer at the World Calibration Center for Ozone Sondes (WCCOS).

2.2.3 Reprocessing

Before a launch, each ECC sensor is conditioned at least twice, following a repeatable set of operating procedures that are taken either from the manufacturer, the WMO/GAW recommended Standard Operating Procedure (SOP), or station-derived procedures. Because each ozonesonde is a unique instrument and responds to conditioning procedures slightly differently, the singular characteristics of each sensor must be recorded, in particular for the background current and flow rate metadata are parameters in equation 2.6, and these values must be documented and archived in perpetuity (Witte et al., 2017). Because there are a variety of operational procedures, instrument types, and sensing solutions, with data reporting and formats dependent on the processing software, a customized approach to reprocessing is required at each site, using the following procedure:

1. To document ozonesonde profile metadata from available checklist sheets, metadata archived in the original data files, and historical knowledge of the operating procedures and processing software from the station Principal Investigators (PIs) and/or operators;
2. To assess the extent to which data can be fully reprocessed. In cases where metadata and checklists are missing or incomplete, average values are used for parameters such as background current or pump flow rates.
3. To provide a summary, itemize what corrections can be applied per profile.

2.2.4 Uncertainty Budget Analysis

The uncertainty in the calculation of the ozone partial pressure (P_{O3}) is a composite of the individual uncertainties associated with each of the different variables, which are the measured cell current (I_M), the background current (I_{BG}), the volumetric flow rate of the pump (Φ_P), the temperature of the pump (T_P), and the ozone sensor efficiency (η_{OS}) (Sterling et al., 2018; Witte et al., 2018). As part of the data homogenization, all systematic bias effects are removed and hence it is assumed that the uncertainties are random and follow a random normal distribution. The uncertainty calculation also accounts for the increased uncertainty incurred by homogenizing the data record and this is included as well. The Gaussian law of error propagation represents the overall relative uncertainty of P_{O3} (see also H. G. J. Smit and the O3S-DQA Panel, 2012), plus any additional uncertainties (Tarasick et al., 2019):

$$\frac{\Delta P_{O3}}{P_{O3}} = \sqrt{\left(\frac{(\Delta I_M)^2 - (\Delta I_{BG})^2}{(I_M - I_{BG})^2}\right) + \left(\frac{\Delta \Phi_P}{\Phi_P}\right)^2 + \left(\frac{\Delta T_P}{T_P}\right)^2 + \left(\frac{\Delta \eta_{OS}}{\eta_{OS}}\right)^2 + \sum \varepsilon_i^2}$$

[2.7]

Here, the additional term in ε_i represents additional random uncertainties due to other causes: uncertainties associated with any bias corrections applied to the other parameters, and uncertainties in the pressure coordinate or time registration of the ozone signal, which in practice expressed as uncertainties in ozone partial pressure. For example, a pressure error, by assigning an ozone value to the wrong pressure altitude, will translate into an effective error in ozone, of magnitude proportional to the ozone gradient (Tarasick et al., 2019).

Figure 13 shows the uncertainties of each variable as well as the total uncertainty for an example ozone profile from Boulder, CO (Sterling et al., 2018). The relative uncertainties of the measured cell current and background current are the largest contributors to the overall uncertainty in the troposphere, when the difference in the measured and background cell current is the smallest, while the pump temperature uncertainty is the smallest contributor to the total uncertainty through the entire record.
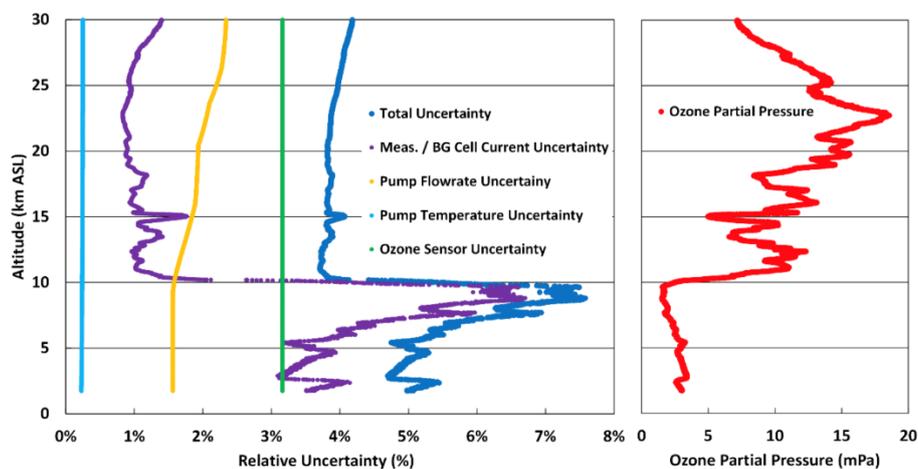
Figure 13. Ozone partial pressure and the relative uncertainty with the relative uncertainty of each variable versus altitude for ozone sounding in Boulder, CO (Figure 6 in Sterling et al., 2018).

2.2.5 Error Component Analysis

Overall, in recent decades, the random error component of sonde measurements is generally within ±5–10% between the tropopause and altitudes less than 26 km for all types of sondes. Systematic biases between all types of ozonesondes or compared to other ozone sensing techniques are smaller than ±5–10%. Above about 26 km altitude the results are not conclusive and the measurement behaviour of the sonde types differs. The uncertainty at the top of the measured profile depends on the type of ozonesonde and sensor solution. For example, intercomparison studies (e.g., Smit et al., 2007; Smit and ASOPOS panel, 2014) indicate that the response of ECC sondes between 28 km and 35 km depends on the type of ECC sonde and sensing solution applied (i.e., 10–20% differences at altitudes near 35 km). However, laboratory studies (Johnson et al., 2002) and international intercomparisons like the Jülich Ozone Sonde Intercomparison Experiment

(JOSIE) (Smit et al., 2007) and the Balloon Experiment on Standards for Ozone Sondes (BESOS) (Deshler et al., 2008) have also clearly demonstrated that even small differences in sensing techniques, sensor types, or sensing solutions can introduce significant inhomogeneities in the long-term sounding records between different sounding stations or within each station individually.

Therefore, existing artefacts in long-term sounding records have to be resolved by homogenization either in space (between different stations) or in time (long-term changes) through the use of generic transfer functions that have been derived from intercomparison experiments (e.g., JOSIE or BESOS) and dual balloon soundings (Deshler et al., 2017). A major goal of the Ozone Sonde Data Quality Assessment (O3S-DQA) is to reduce the uncertainties between long-term sounding records from 10–20% down to 5–10% through the use of generic transfer functions (Smit and O3S-DQA panel, 2012). Currently, a total of about 30 long-term station records have been reevaluated and homogenized by resolving known instrumental bias effects, thereby reducing the uncertainties down to 5–10% (Tarasick et al., 2016; Van Malderen et al., 2016; Deshler et al., 2017; Sterling et al., 2018; Witte et al., 2017; Witte et al., 2018).

2.2.6 Example of ozone partial pressure temporal series

Figure 14 shows an example time series of ozone partial pressure profiles (in mPa) measured with ozonesondes launched at the Boulder station (39.99°N, 105.26°W, 1634 masl).
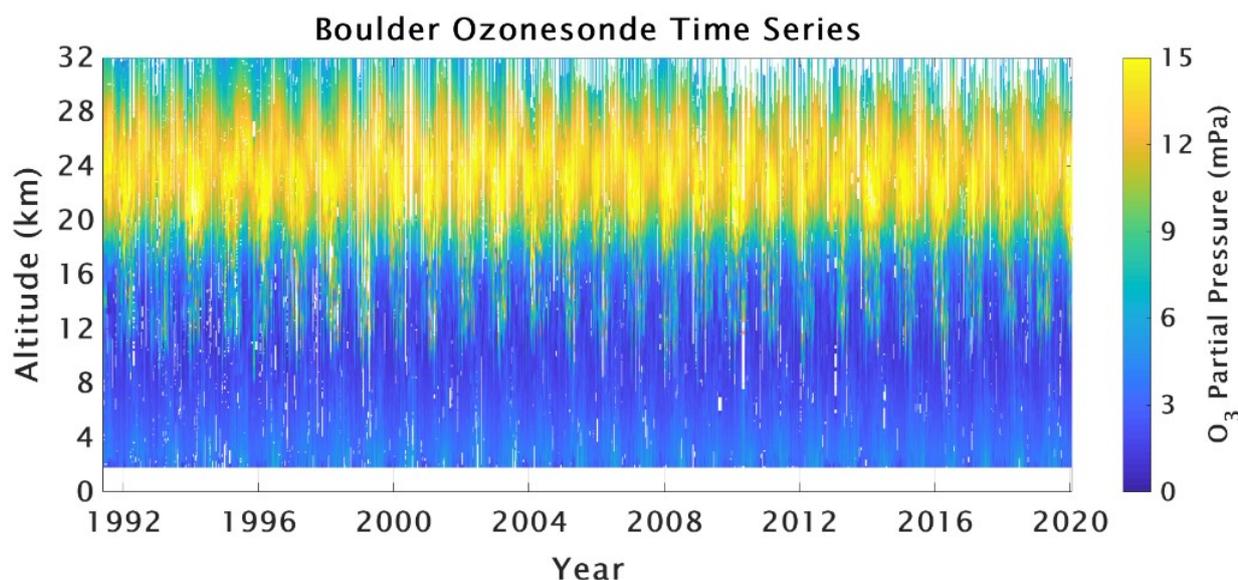
*Figure 14. Time series of ozone partial pressure profiles (in mPa) at the Boulder station (39.99°N, 105.26°W, 1634 masl).*

It can be seen how the highest values of the ozone profile are found in a range from 20 km to 28 km, with the maximum around 24 km, without any obvious trend being visible but with a clear seasonal imprint. Furthermore, having such an accurate estimate is not always possible with other measurement techniques.

2.3 Techniques comparison

Figure 15 shows four ozone profile time series measured in Lauder, New Zealand (45°S, 170°W), using four measurement techniques: ozonesondes, described previously, Light Detection And Ranging (LIDAR, Petropavlovskikh et al., 2019; Mégie and Menzies, 1980; Pelon et al., 1986; Godin-Beekmann et al., 2003; Godin et al., 1999; Leblanc and McDermid, 2000; Leblanc et al., 2016b), Microwave ozone radiometers (MWR, Connor et al., 1995; Hocke et al., 2007; Hassler et al., 2014; Studer et al., 2013; Studer et al., 2014; Nedoluha et al., 2015, Maillard-Barras et al., 2009) and Ground-based Fourier-Transform InfraRed (FTIR, Hase et al., 1999; Hase, 2000; Pougatchev et al., 1995; Rodgers, 2000).
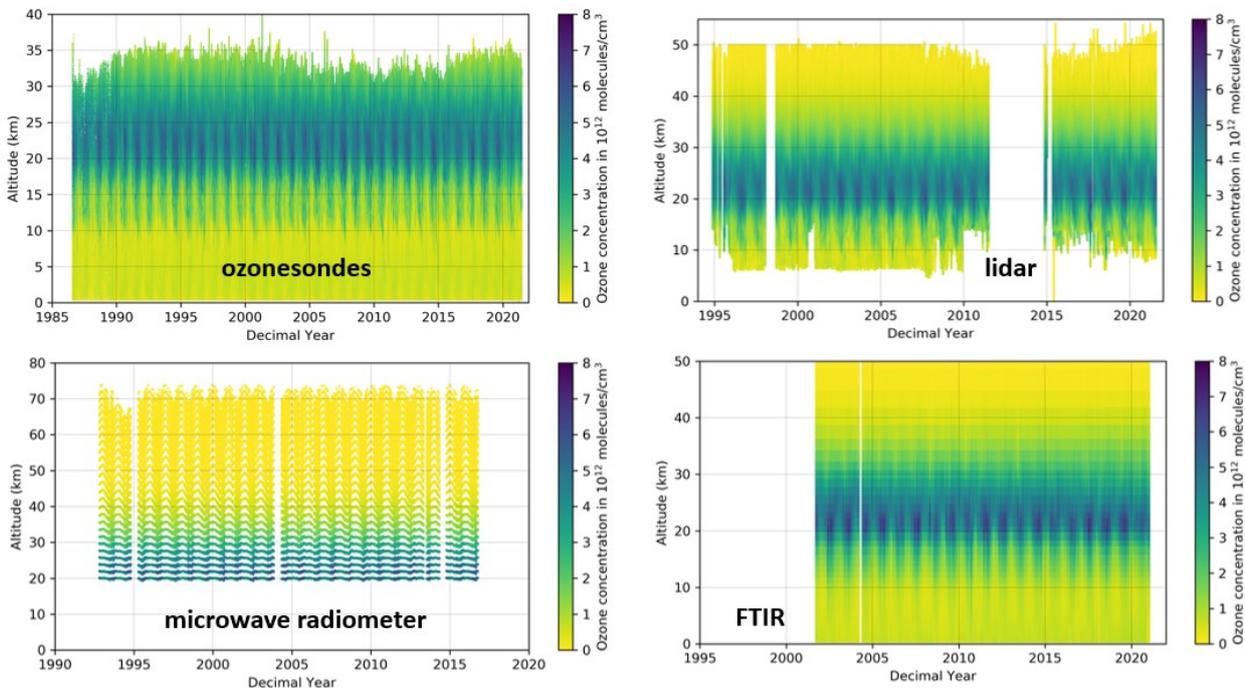
Figure 15. Time series of ozone profiles for the Lauder station, New Zealand (45°S, 170°W) with different measurement techniques: top left ozonesondes, top right LIDAR, bottom left microwave radiometer, and bottom on the right FTIR.

These four long-term, high-quality data sets make it possible to compare the different techniques used for calculating ozone profiles. The different vertical ranges of coverage between the different techniques are immediately apparent. For the ozonesondes, the vertical range of coverage is between the surface and approximately 35 km, for the LIDAR between approximately 7 km and 50 km, for the microwave radiometer between approximately 20 km and 70 km, for the FTIR between the surface and 50 km. Furthermore, these measurement techniques also have different vertical and temporal resolutions: ozonesonde between 100 m and 150 m for 60 seconds, LIDAR between 3.75 m and 15 m for approximately 60 seconds, microwave radiometer

between 8 km and 17 km for approximately one hour, FTIR less than 1 km for 30 minutes, as summarized in Table 1.

Technique	Vertical range	Vertical resolution	Temporal resolution
Ozonesonde	Surface to 35 km	100 m to 150 m	60 s
LIDAR	7 km to 50 km	3.75 m to 15 m	60 s
MWR radiometer	20 km to 70 km	8 km to 17 km	1 h
FTIR	Surface to 50 km	<1 km	30 m

Table 1. Overview of the vertical coverage range, vertical and temporal resolution of ozone sounding, LIDAR, microwave radiometer, and FTIR techniques.

Despite the different sensitivities and different vertical ranges of coverage, all these techniques reproduce the same phenomenon well and, therefore, there is consistency between them. All techniques show higher ozone concentrations from approximately 20 km to 30 km, with a maximum of around 24 km. The ozonesonde data, given the consistency with other measurement techniques, are distributed globally and ensure good coverage for the study of anomalies and trends and it is for this reason that they were used for the analysis presented in this study.

3. Unified Ozonesonde Dataset

To improve the time and spatial coverage of ozonesounding measurements globally and improve the estimation of trends, a unified database of ozonesounding profiles was obtained through the merging of three existing ozonesounding datasets, provided by the Southern Hemisphere Additional OZonesondes (SHADOZ), the Network for the Detection of Atmospheric Composition Change (NDACC), and the World Ozone and Ultraviolet Radiation Data Centre (WOUDC). The unified dataset can provide adequate data coverage at different latitudes and sufficient sampling for the calculation of zonal or global anomalies and trends. In addition, only a few variables of interest were considered to build the unified dataset due to the different formats and amounts of data and metadata provided by each network, despite data often referring to the measurements provided to different initiatives.

For this purpose, an algorithm was implemented able to merge the different datasets by handling their different features and duplicated profiles.

3.1 Unified Dataset Composition

3.1.1 SHADOZ

The SHADOZ network provides ozone profiles measured at 17 sites since 1998. In addition, each of the soundings also yields pressure, temperature, and relative humidity profiles for all stations. The SHADOZ network was designed to reduce data heterogeneity in ozonesonde profiles between several sites measuring in the Tropics and Subtropics, and the network data was recently reprocessed to homogenize the database further. In the unified dataset, the SHADOZ V6.0 for measurements and SHADOZ V1.0 for uncertainty have been considered, providing also uncertainties for 14 of the 17 sites, with the former having all operated for longer than a decade. SHADOZ is the only one of the three datasets providing a detailed estimate of the observational uncertainty for each data record. The original spatial coverage of SHADOZ was determined by two requirements:

1. that the network consists of existing stations;
2. full zonal coverage to resolve an equatorial 'wave-one' feature observed in satellite total column ozone (see Thompson et al., 2017 and references within).

At the initiation of SHADOZ in 1998, nine stations were meeting these criteria, all in the Southern Hemisphere, hence the name of the network. Stations north of the equator joined SHADOZ as follows: Kuala Lumpur (in 1999), Paramaribo (in 1999), Costa Rica (in 2005), Cotonou (operated 2004–2007), and Hanoi (where soundings began in 2004). Hilo, Hawaii, with a record extending back to the 1980s, joined SHADOZ in 2009. The 14 SHADOZ stations which have operated for at least a decade are shown in Figure 16.

Figure 16. Map of the 14 SHADOZ stations (https://tropo.gsfc.nasa.gov/shadoz/) for which data sets have been reprocessed (Thompson et al., 2017).

3.1.2 NDACC

The NDACC network, which officially started operation in 1991, is composed of more than 90 globally distributed, ground-based, remote-sensing research stations with more than 160 currently active instruments.

Despite the high data quality, NDACC does not routinely provide uncertainties for the ozonesounding measurements. Only for a subset of measurements, an uncertainties estimation is provided but not with the same level of detail as SHADOZ. NDACC provides ozone (O_3) vertical profile and column data products from different measurement techniques, which are:

- FTIR spectrometer;
- UV-visible spectrometer (total columns only);
- microwave spectrometer;
- LIDAR;
- Ozonesondes;
- Dobson/Brewer (total columns only).

NDACC datasets are publicly available from the NDACC data portal at ftp://ftp.cpc.ncep.noaa.gov/ndacc.

For this study, only ozonesounding profile data were processed, which were provided by 33 stations, shown in Figure 17.
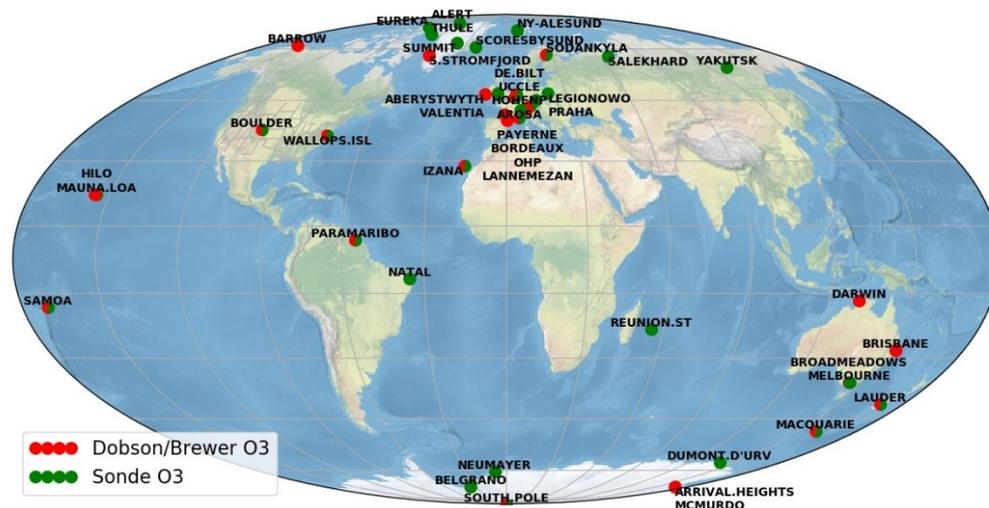
Figure 17. Map of the 33 NDACC ozonesounding stations (ftp://ftp.cpc.ncep.noaa.gov/ndacc) that provide consolidated data to the Copernicus Climate Data Store (CDS).

3.1.3 WOUDC

The WOUDC network, which is part of the Global Atmosphere Watch (GAW) programme of the World Meteorological Organization (WMO), was established in 1960 to collect, quality-control, archive, and provide long-term access to high-quality observation data and metadata from the WMO GAW network of stations measuring ozone column and ozone vertical profiles. Furthermore, the data record provides sustainable high-quality data for all application areas via coordinated activities between the WOUDC and regional and global calibration and data quality centres.

Ozone data from WOUDC are extensively used by hundreds of scientists in a large number of studies and applications including the joint WMO, UNEP sponsored quadrennial Scientific Assessments of Ozone Depletion. Those Assessments are published every four years under the umbrella of the Vienna Convention for the Protection of the Ozone Layer, and they are the most authoritative overview of the state of the atmosphere concerning stratospheric ozone.

To meet the needs of sustainable high-quality data for all application areas, the GAW ozone observation network coordinates activities not only with WOUDC but also with regional and global calibration and data quality centres and is guided by the WMO Ozone Scientific Advisory Group.

WOUDC is the most comprehensive initiative for collecting ozonesounding measurements that, although quality checked for their consistency in the data and metadata, are less thoroughly characterized in terms of traceability and uncertainties than SHADOZ and NDACC. However, data providers are advised to download and use measurements following the guidelines provided by WMO (https://woudc.org/about/data-policy.php).

Detailed information on all 150 ozone observation stations that submitted data to WOUDC, including their GAW and platform IDs, geographic location, altitude, country, contributing organizations, and exact data reporting periods is available on the WOUDC website. The global map of ozone stations which have submitted data to WOUDC is shown in Figure 18.
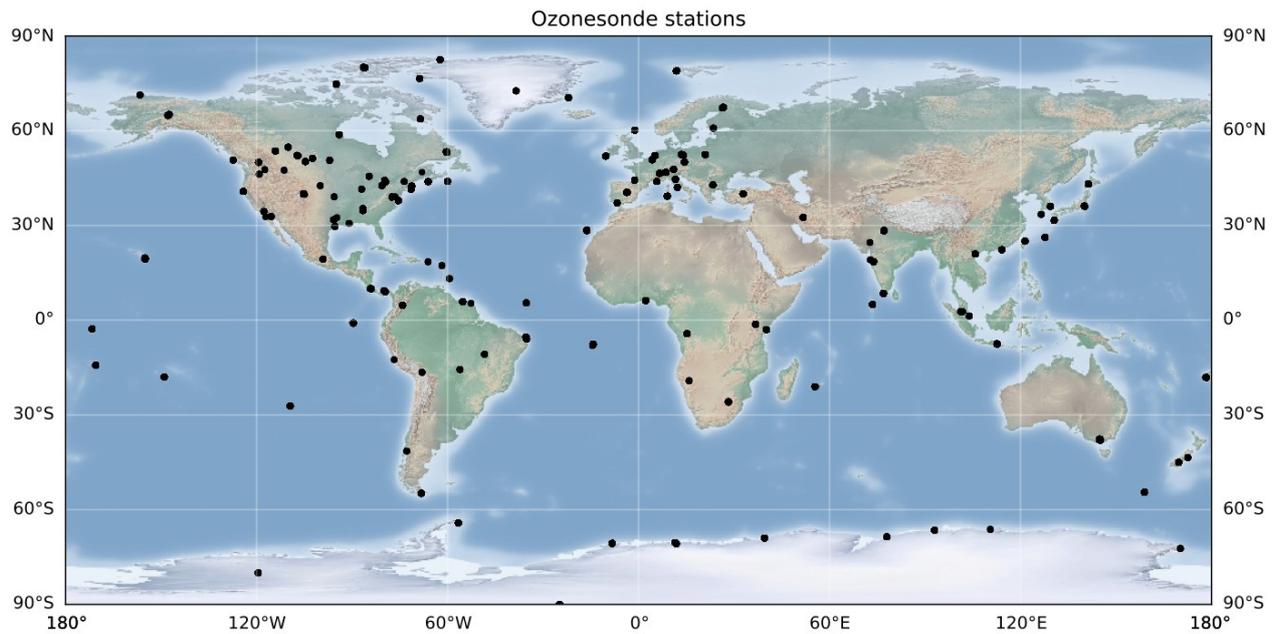
Figure 18. Map of the global distribution of all ozonesonde measurement stations submitting profile ozone data to WOUDC.

3.1.4 Selection of Variables

As a preliminary step for the creation of the unified dataset, metadata and variables shared by the three source network datasets were chosen, and these are shown in Table 2. In addition, whenever available, uncertainty contributions are also included. As mentioned earlier, the three networks have different formats and amounts of data and metadata.

Standard name	Description	Unit
station_name	The name of the station.	N.A.
location_latitude	Latitude of station.	deg
location_longitude	Longitude of station.	deg
report_timestamp	Date when the ozonesonde was launched (in format yyyy-mm-dd hh:mm:ss with time zone).	N.A.
air_pressure	Atmospheric pressure of each level in Pascals.	Pa
air_temperature	Air temperature in degrees Kelvin.	K
relative_humidity	Relative humidity in 1.	1
latitude	Observation latitude (during the flight).	deg
longitude	Observation longitude (during the flight).	deg
observation_height_above_station_surface	Height of sensor above local ground or sea surface. Positive values for above surface (e.g., sondes), negative for below (e.g., xbt). For visual observations, the height of the visual observing platform.	m (a. s. l.)
ozone_partial_pressure	The level partial pressure of ozone in Pascals.	Pa
ozone_concentration	The level mixing ratio of ozone in ppmv.	ppmv
ozone_partial_pressure_total_uncertainty	Total uncertainty in the calculation of the ozone partial pressure as a composite of the individual uncertainty contribution. Uncertainties due to systematic bias are assumed as random and follow a random normal distribution. The uncertainty calculation also accounts for the increased uncertainty incurred by homogenizing the data record.	Pa
profile_source_network	Source network of the profile.	N.A.
type_flag	Station classification flag.	N.A.

Table 2. List of the selected variables for storing in the unified dataset.

3.1.5 Handling of the duplicated profiles

The main problem in merging the three datasets into a single database is the handling of the duplicated profiles. For duplication, it means the at least two profiles for the same station and day exist in two different datasets. This can occur when measurements from the same station are submitted to more than one network, which in theory should be identical but are often provided for different periods, using different data formats, and providing the amount of individual data points. Metadata may also differ. This also means that the different networks do not always report the same number of ozone levels for the same profile. To overcome this issue, a range of selection criteria have been applied to select the profiles according to their quality, coverage, and characterization of the measurement uncertainties. The selection criteria are summarized below:

- In case of a single profile for a day (no duplicates), this profile will be saved in the database;
- In case of duplicates, the profile selection is driven by the dataset maturity and by the availability of measurement uncertainties on the ozone concentration profiles. The uncertainties are provided mainly in the SHADOZ database and for some locations also in the NDACC database. WOUDC, on the other hand, does not provide any uncertainty estimates;
- Discarded profiles are eventually used to densify the vertical sampling if they provide measurements at different pressure levels not available in the selected profiles.

Table 3 provides an overview of the main information of the three datasets mentioned above and of the unified database, including the number of profiles for each of them.

DATABASE	# STATIONS	PERIOD	# PROFILES	UNCERTAINTY DATA
SHADOZ	14	1998-2022	9343	Yes
NDACC	33	1969-2020	45382	Yes (for a minor fraction of data).
WOUDC	150	1962-2022	97252	No
UNIFIED	155	1978-2022	78885	Yes

Table 3. Overview of information on the number of stations, period, number of current profiles, and presence of uncertainty data for the three datasets used.

As shown in Table 3, the number of profiles present in the unified dataset is lower than the number of profiles available for the WOUDC network because it was chosen to unify only the profiles for the period 1978-2022, and WOUDC, since it also includes profiles measured during for the period 1962-1977, appears

to have more profiles than those of the unified dataset. Furthermore, NDACC profiles from the period 1969-1977 were also excluded from the analysis.

Below, the composition of the unified database, in terms of number of records per network, is listed:

- SHADOZ record stored: 20839851 (14.97% of unified total record);
- NDACC record stored: 64363898 (46.21% of unified total record);
- WOUDC record stored: 54069095 (38.82% of unified total record);
- UNIFIED total record: 139272844.

To show the benefits of using the unified ozonesounding database, an example of the comparison of the time series of ozone profiles for the Hilo station (19.72°N; 155.07°W) available from SHADOZ, NDACC, and WOUDC with the corresponding time series obtained in the unified database is shown in Figure 19.
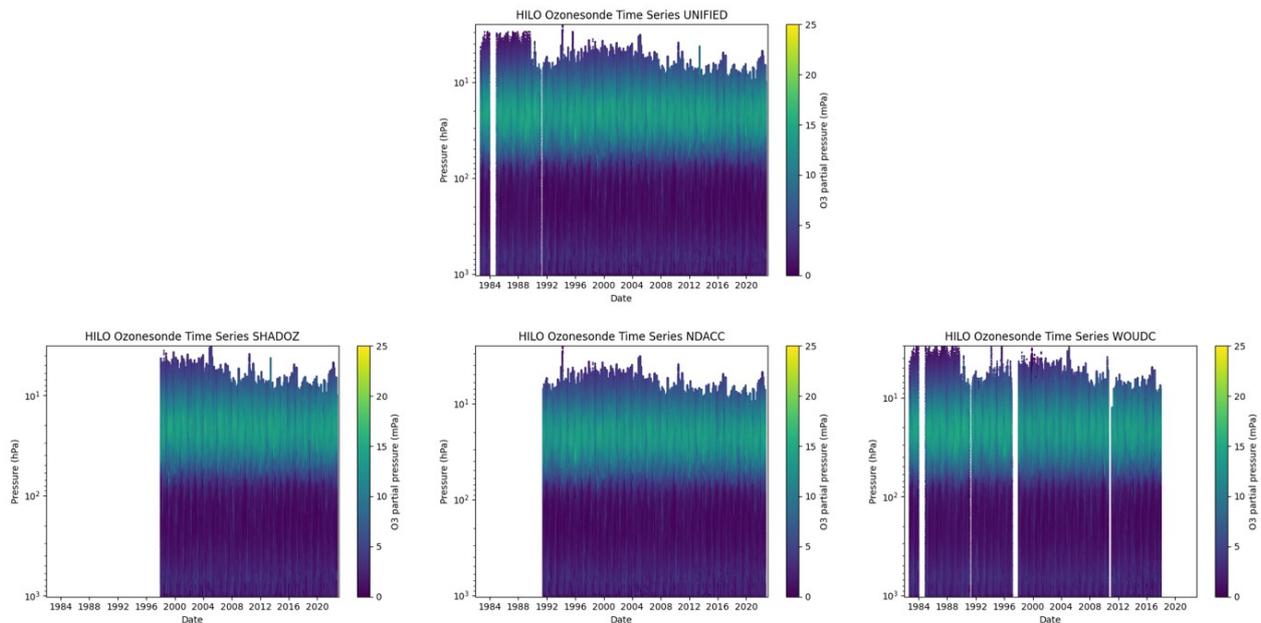
Figure 19. Hilo station (19.72°N; 155.07°W) time series for the unified database (top panel), SHADOZ (bottom left), NDACC (bottom centre), and WOUDC (bottom right) datasets. The time series of the unified database is generated by the unification algorithm with the contribution of the ozonesonde profiles from SHADOZ, NDACC, and WOUDC networks.

It appears that the unified time series merges all the profiles collected: before 1991 from WOUDC, missing in the SHADOZ and NDACC datasets; between 1991 and 1997, when only NDACC and WOUDC are available; between 1998 and 2018, when all the network provides data, and the algorithm can remove duplicates and select the highest quality profiles. It must be noted that since 2018 WOUDC has not provided ozone concentration profiles for Hilo. The time series of the unified database consists of 1202 (63.00%)

ozonesounding profiles from SHADOZ, 327 (17.14%) from NDACC, and 379 (19.86%) from WOUDC, adding up to a total of 1908 profiles.

3.1.6 Dataset data coverage

Figure 20 shows through a heatmap, the number of monthly ozonesounding profiles available in the unified database at a global scale.
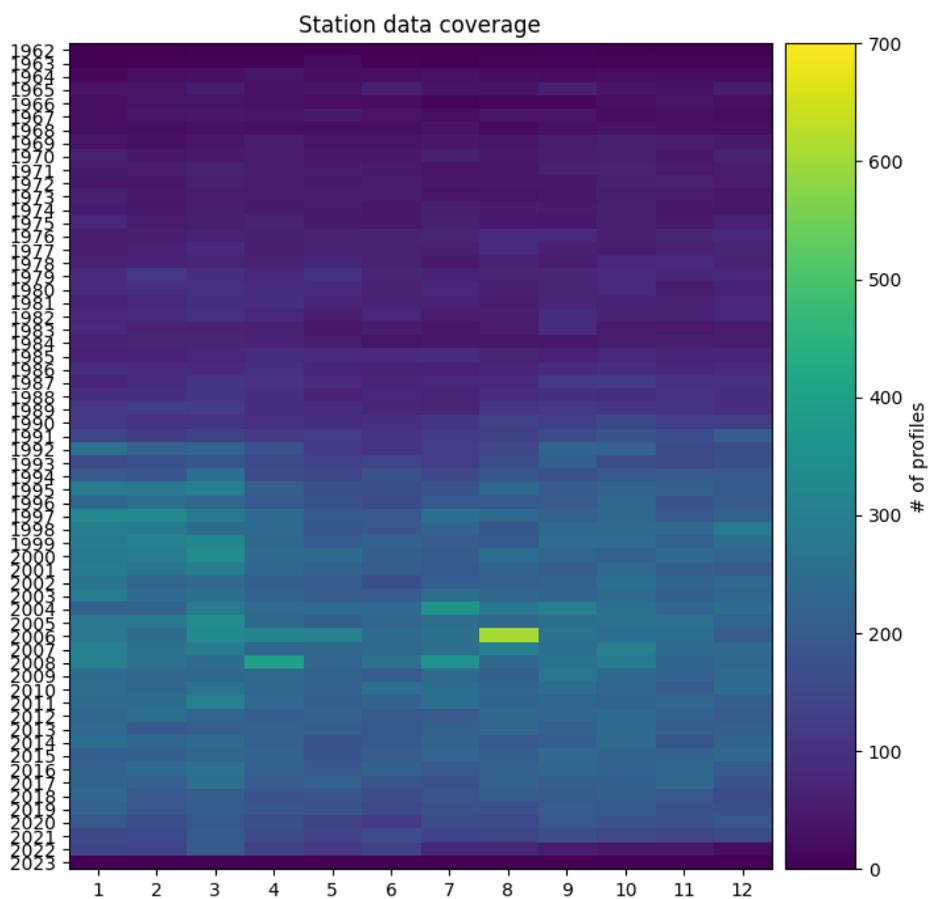
Figure 20. Heatmap of the number of monthly profiles available in the unified ozonesounding database since 1962.

The number of profiles available in the unified database increased after 1992, due to a corresponding increase in the number of stations performing ozonesoundings. Temporal data coverage was also investigated at different latitudes (Northern Hemisphere polar latitudes (NP, 60°N-90°N), Northern Hemisphere mid-

latitudes (NH, 30°N-60°N), Tropics (TR, 30°S-30°N), Southern Hemisphere mid-latitudes (SH, 60°S-30°S), and Southern Hemisphere polar latitudes (SP, 90°S-60°S)) and for four vertical layers (300-200 hPa, 200-100 hPa, 100-50 hPa, and 50-1 hPa). This division into latitudinal bands was considered valid given the good representativeness of the single stations, as discussed later in paragraph 3.4, Figure 21 provides an example of temporal data coverage for the 50-1 hPa vertical altitude range, which is quite similar to the coverage found in the other investigated altitude ranges.
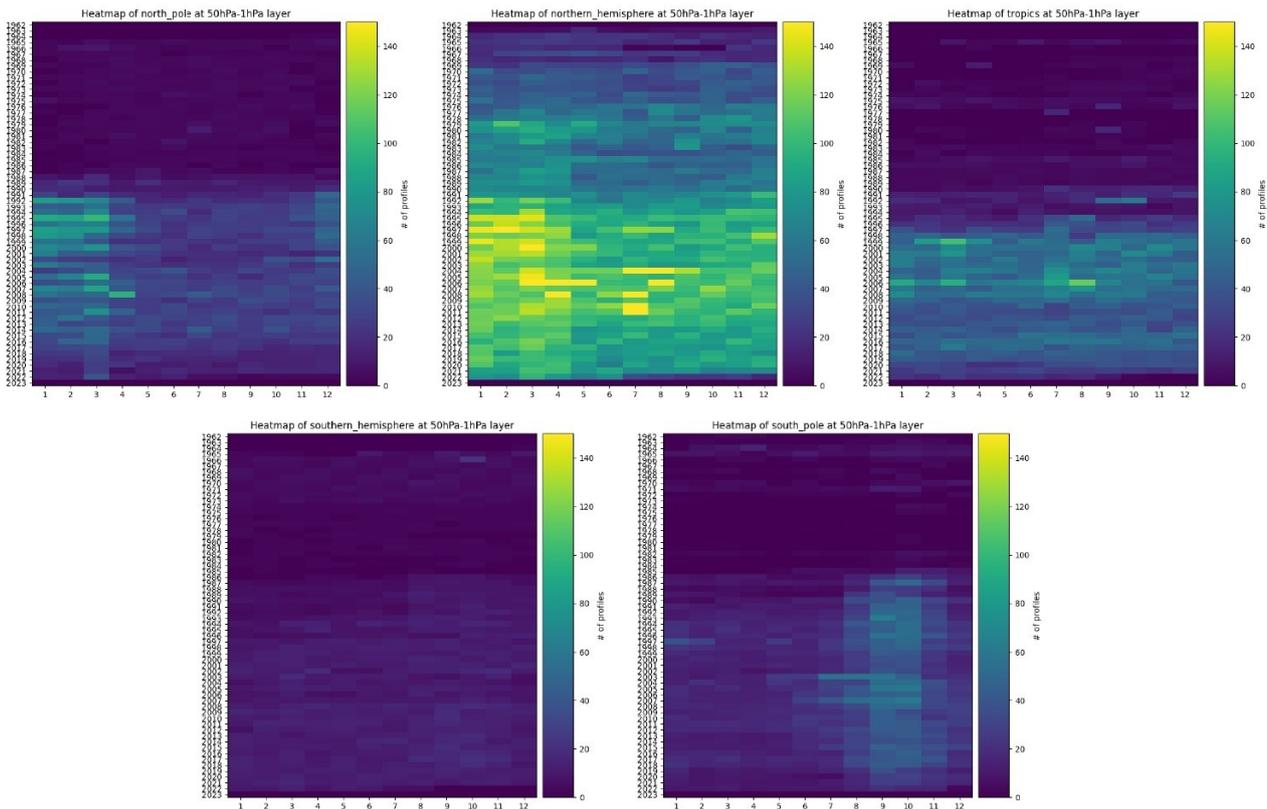
Figure 21. Same as Figure 20 for the profiles covering a vertical range up to 1-50 hPa.

Figure 21 reveals how the vast majority of profiles are taken in the NH and TR, where most of the stations are situated/located. Moreover, for each latitudinal sector, there is a strong difference in the amount of data before and after 1990; this is due to both, the small number of ascents available before 1990 and to the vertical range covered, which gradually improved over the years following 1990. The latter is shown in Figure 22 with a comparison of two profile sets from the same station, the first on the left generated in the summer of 1979 and the second on the right in the spring of 2019.
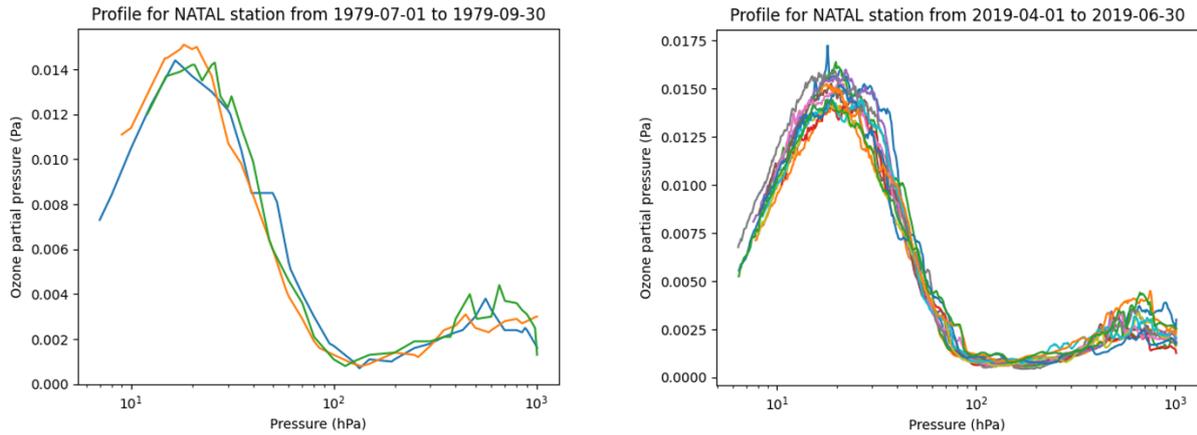
Figure 22. On the left is the plot of all the profiles generated in the summer of 1979 and, on the right, in the spring of 2019 for the station of Natal. Notice the difference in the number of profiles produced: 3 profiles for summer 1979, and 13 profiles for spring 2019.

The number of profiles provided by the station in the summer of 1979 (3 profiles) is significantly lower than the number of profiles in the spring of 2019 (13 profiles). In addition, the profiles measured during the summer of 1979 have fewer points than the ones measured during the spring of 2019, as illustrated in Figure 23. The different atmospheric sampling of these profiles is due to the different sonde types used.
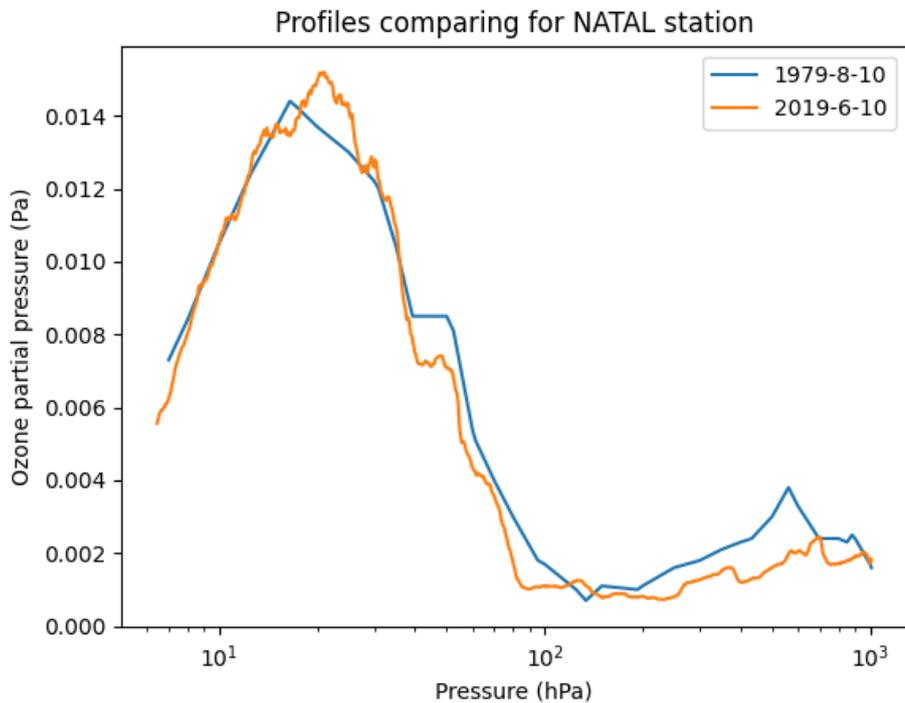
Figure 23. The plot of the profiles was generated in the summer of 1979 and in the spring of 2019 for the station of Natal. It reveals that the difference in the vertical resolution of the ozonesonde profiles is mainly due to the difference in the number of points for a single profile.

Figure 24 shows the stations available in the unified ozonesounding database.

Ozonesonde stations

Total station: 155
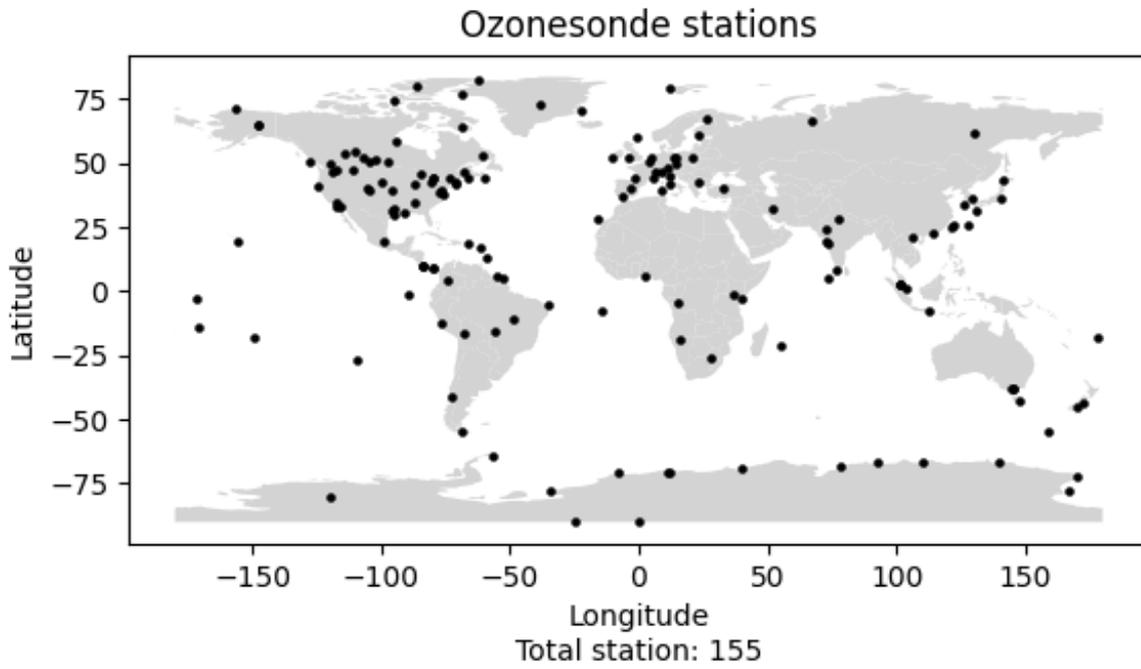
Figure 24. Map of global ozonesounding stations available in the unified database.

3.2 Station Classification

The stations saved within the unified dataset have different time series, which may depend on their activity and the number of ozonesoundings produced monthly. Therefore, for a more accurate analysis of anomalies and trends, constraints were defined to select the highest-quality stations for the following trend analysis. This is because data coverage is key to estimating anomalies and trends with plausible uncertainties. As a consequence, the stations of the unified dataset have been grouped based on their monthly coverage, assuming one month is covered if at least one ozonesounding ascent per month is available. This ensures at least monthly coverage of the time series, considering that not all stations can provide more than one profile per month due to high costs.

Accordingly, as reported in Appendix A, the 155 available stations were grouped into three different clusters:

1. Long coverage (LC): 26 stations (with a continuous data time series of at least 20 years).
2. Medium coverage (MC): 23 stations (with a continuous data time series between 10 and 20 years).
3. Short coverage (SC): 106 stations (with continuous data time series less than 10 years or no data available between 1978 and 2022).

The first two clusters were the only ones able to provide sufficient data coverage for the estimate of anomalies and trends. The difference with the SC cluster is shown in Figure 25, where the coverage of the stations at the South Pole (90°S; 0°W), classified as LC, is compared with Macquarie Island (54.50°S; 158.94°W), classified as MC, and with Praha (50.0°N; 14.44°W), classified as SC.
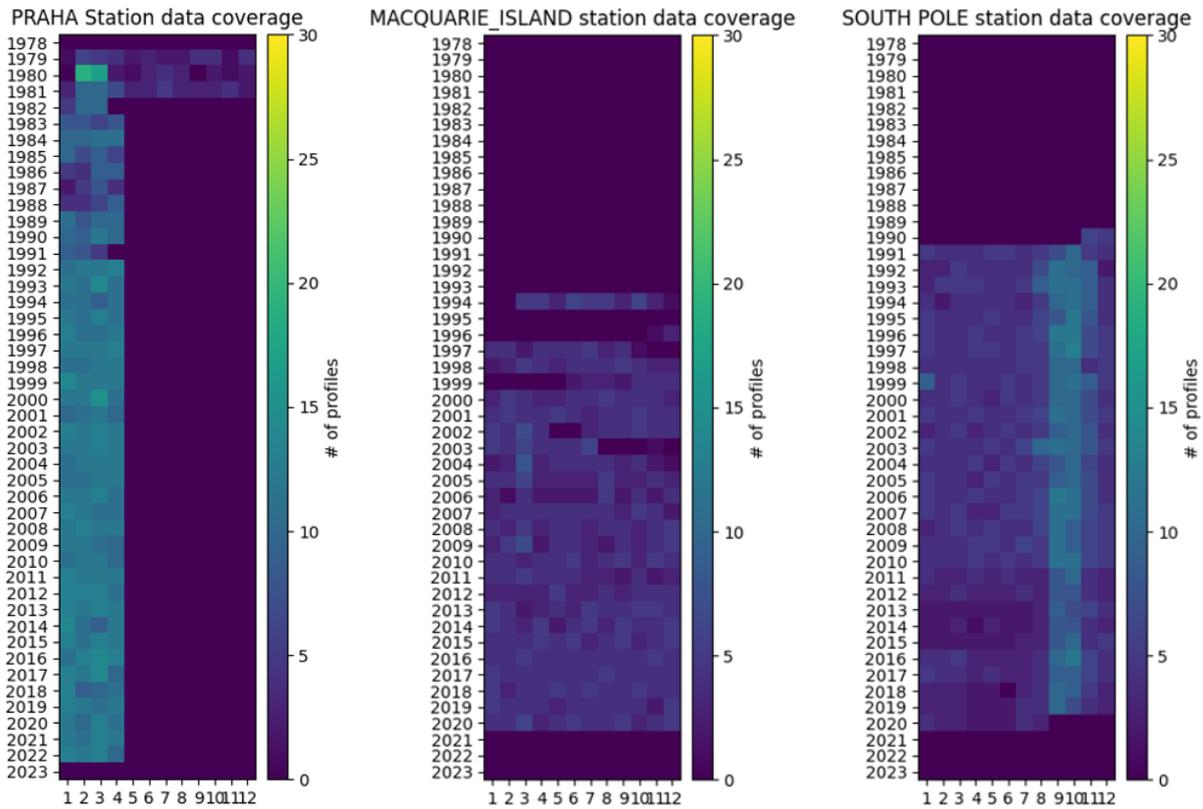
Figure 25. Data coverage of the station of Praha (left), classified as "Short Coverage", Macquarie Island (centre), classified as "Medium Coverage", and South Pole (right), as "Long Coverage". The data coverage of Macquarie Island station would seem to belong to the LC cluster, but in 2003 there was a gap of three months that interrupted the time series; therefore, this station is classified as MC. However, in the future, with the processing of additional new data, the classification of this station may be changed to LC as soon as 20 years of continuous data is reached. In the case of Praha, instead, a large number of profiles is available, but only for the first 4 months of the year, consequently missing the ozone seasonal cycle. As a consequence, the station is classified as SC.

The 49 stations classified as LC or MC, on which the analysis can be carried out, are shown in Figure 26.
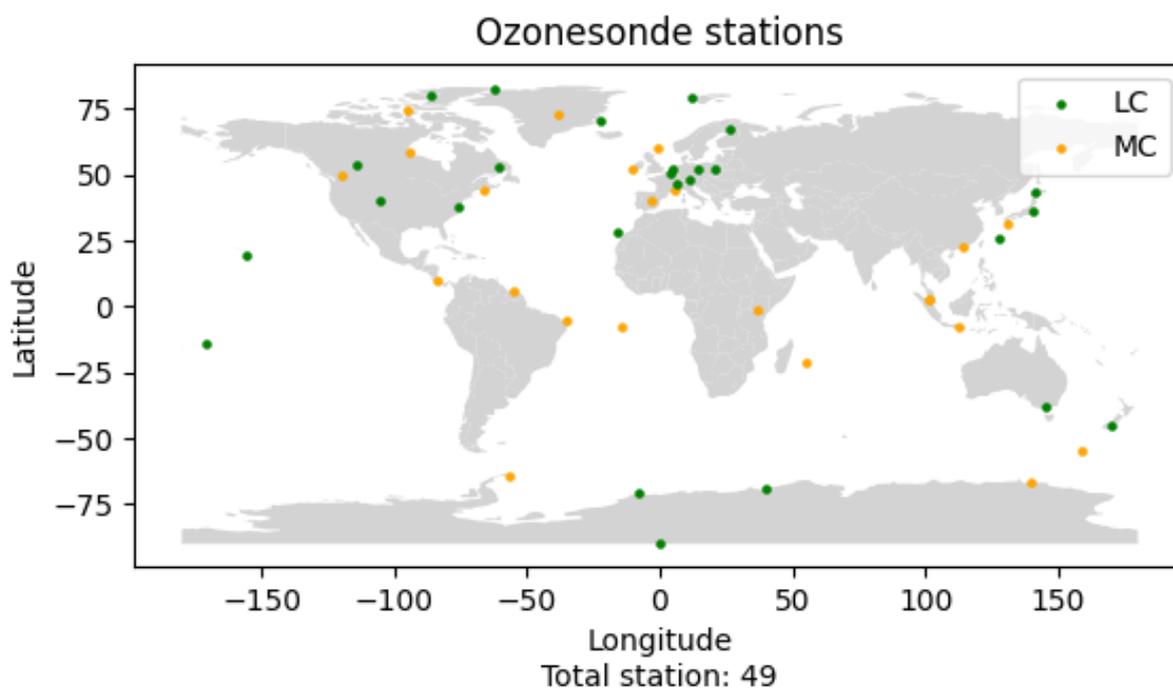
Figure 26. Map of global ozonesounding stations, for long coverage and medium coverage clusters, available in the unified database.

Table 4 shows, for each latitudinal sector, the number of stations available in the unified dataset for each cluster.

	# Long coverage stations	# Medium coverage stations	# Long and medium coverage stations
NP	5	3	8
NH	12	7	19
TR	4	10	14
SH	2	1	3
SP	3	2	5

Table 4. Number of stations for long coverage cluster, medium coverage cluster, and high and medium coverage cluster.

Due to the limited number of high-quality stations for the SP and the SH, the analysis is not carried out on these two sectors. Although trends for these two sectors are significant, the lack of data remarkably increases uncertainties and the calculated trend values cannot be considered reliable estimates.

The classification presented in this section may be subject to change for future versions of the dataset based on other sources that may be considered, such as Harmonization and Evaluation of Ground-based

Instruments for Free Tropospheric Ozone Measurements (HEGIFTOM), and that may provide other records to be added to those already considered.

3.3 Quality checks

Ensuring the quality of products within the unified dataset is of paramount importance to have higher reliability for estimating anomalies and trends. For this reason, several quality checks were applied to the NDACC, SHADOZ, and WOUDC profiles during the data selection (see Table 3) and they are listed below:

- Plausibility checks, to ensure reported values are within plausible physical ranges and comply with the measurement units of the International System of Units (SI). In Table 5, the outcome of these checks for the main variables of the SHADOZ, NDACC, and WOUDC datasets is summarized;
- Outliers check, to verify if there is a suspicious number of extreme values. Candidate outliers are identified using the Inter-Quartile Range method as follows:

-

$$median - 3 \cdot IQR \leq observation \leq median + 3 \cdot IQR,$$

-

If many outliers are found, these are thoroughly investigated and, if verified as outliers, removed from the dataset;
- Completeness checks are undertaken on a monthly basis to verify that all variables are complete (availability of at least 1 profile per month is required);
- Vertical coverage checks are performed on a monthly basis to verify if ozone profiles reach 10 hPa. 59711 profiles exceed 10 hPa (75.69%), 14598 (18.51%) terminate between 20 hPa and 10 hPa, and 4577 (5.8%) between 50 hPa and 20 hPa.

Variable	Plausible physical range	Flagged values SHADOZ	Flagged values NDACC	Flagged values WOUDC
Air pressure	> 0 Pa	0.008%	0.04%	0.01%
Air temperature	150 K ≤ x ≤ 330 K	0.21%	2.78%	0.46%
Relative humidity	0 ≤ x ≤ 1.2	9.23%	17.48%	13.08%
Ozone partial pressure	≥ 0 Pa	2.31%	7.32%	0.85%
Ozone concentration	0 ppmv ≤ x ≤ 30 ppmv	2.31%	86.22%	N.A.
Ozone partial pressure total uncertainty	N.D.	19.75%	68.05%	N.A.

Table 5. Plausibility checks were applied to the main variables of SHADOZ, NDACC, and WOUDC to select data for creating the unified database.

- Vertical completeness checks to ensure a minimum number of reports are available for each vertical region covered by the ozonesoundings. This is quantified in at least one point every 50 meters. In this work, 78885 profiles were checked and only 18281 (23.2%) did not pass this check in some atmospheric regions.
- Statistics of missing values are also investigated to check the coherency with the source datasets. Table 6 shows the percentage of missing values present in the unified database.

Variable	% missing data	Reason
Air temperature	1.68%	Values flagged by plausibility checks.
Relative humidity	14.30%	Values flagged by plausibility checks.
Ozone partial pressure	4.61%	Values flagged by plausibility checks.
Ozone concentration	78.5%	Only available from the SHADOZ and NDACC networks, but not always present in the source files.
Ozone partial pressure total uncertainty	75.49%	Only available from the SHADOZ and NDACC networks, not always present in the SHADOZ files, sometimes present in NDACC files.

Table 6. Missing data percentage of the main variables within the unified ozonesounding database.

All the quality checks described above are used to verify the quality of each profile by ensuring its reliability. Profiles that fail these checks will not be saved within the dataset. An example of rejected profiles is shown in Figure 27.
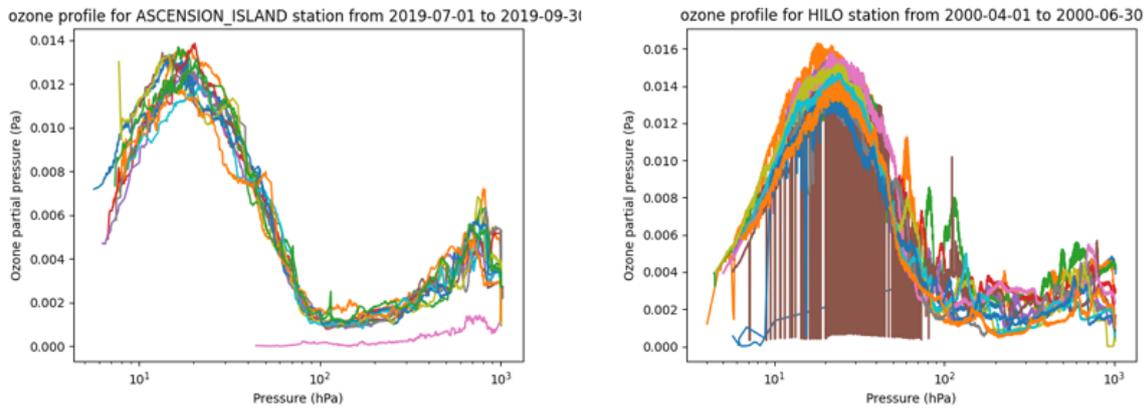
Figure 27. Examples of ozone sonde profiles were rejected after the identification of outliers. The left panel shows incorrect profile values (magenta profile below 400 hPa) compared to the seasonal average profile; the right panel shows a profile with a huge noise level due to a reporting error from the measurement stations, with the profile alternating between plausible and implausible values, the latter being an order of magnitude smaller.

The magenta profile (Figure 27 left panel) presents values completely different from those of the seasonal average and does not reach 400 hPa, not describing the stratosphere in any way; the brown profile (Figure 27 right panel) seems to present a very large noise, but it is due to an alternation between plausible and implausible values. For these reasons, if these profiles were included in the dataset for trend estimation, the estimate would be altered, which is why it was decided to reject them.

3.4 Representativeness of the stations

Representativeness errors are crucial when a climate variable, which is a continuous function, is studied using discrete sampling from an observing network. To study trends in the UT/LS, the network representativeness must be preliminarily quantified to ensure the ozone variability at different latitudes can be accurately captured. In this work, the approach elaborated by Weatherhead and co-authors (2017) is used, which focuses on a measurement site's ability to reproduce monthly variability in seasonally adjusted total ozone column (TOC) at nearby locations, using correlations to describe spatial representativeness, shown in Figure 28. Total Ozone Mapping Spectrometer-Earth Probe (TOMS-EP) ozone satellite data (https://science.nasa.gov/mission/toms-ep/) from July 1996 to December 2005 are used to estimate the correlations of the ozone observations for the stations available in the unified database. Figure 29 shows the representativeness of the LC cluster stations of the unified database only.
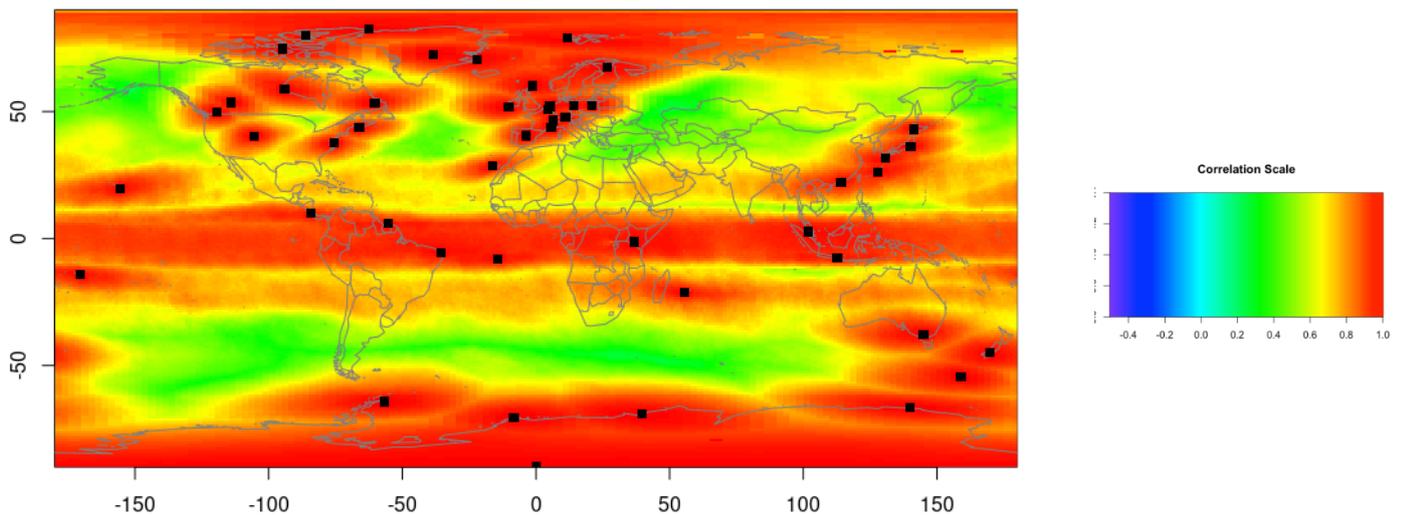
Figure 28. Representativeness of unified *stations using EP TOMS ozone satellite data (July 1996 – December 2005). The map shows the level of correlation between a station and the surrounding area. A value of 1 means that the grid point is fully correlated and 0 is not correlated at all with the other grid points.*
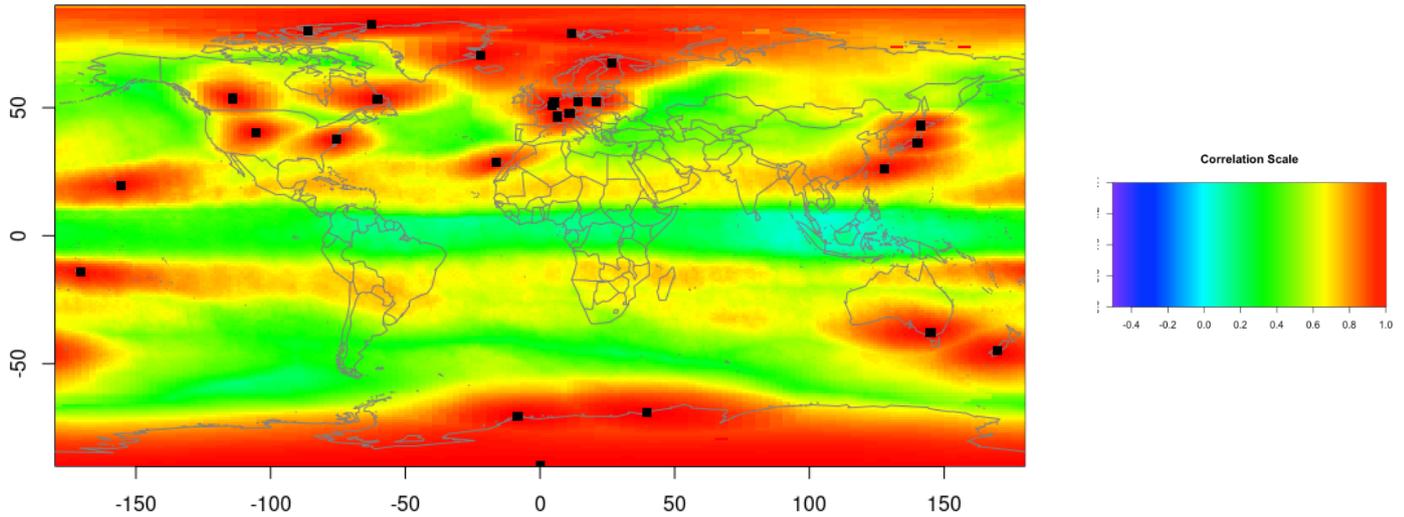
Figure 29. The same as Figure 28 but considering only the stations belonging to the LC cluster.

Figures 28 and 29 reveal a good correlation for the LC cluster for the stations available in the NP, NH, and SP compared to the entire set of stations of the unified database, with the latter logically increasing the area with correlation values larger than 0.8, although not dramatically. For these sectors, the additional stations in the LMC cluster may be considered redundant and we can prioritize the usage of LC station, given the higher quality of the related time series. The most sensible difference is found near the equator, due to the lack of long data records enabling the selection of a station within the LC cluster and some places with a near-zero correlation, while in the remnant of the tropical belt, the difference is much smaller. In this region, we may expect larger differences between the LC and other clusters. Therefore, in the tropics, larger confidence may be given to the trends estimated with LMC rather than the LC. In the SH, the correlation shows no big difference except for Oceania.

4. Ozone concentration trends

4.1 Methodologies

Structural uncertainty is about whether the mathematical structure of a model accurately represents its target (Baldissera Pacchetti, 2021). The following comparison of several linear regression techniques to assess trends aims to quantify structural uncertainty.

4.1.1 Least-square linear regression

The Least-square linear regression method assumes the well-known form (Reinsel et al. 2002):

$$y_t = \mu + S_t + \omega x_t + N_t, \quad t = 1, \dots, T \tag{4.1}$$

where y_t is the monthly anomaly time series, t is the time variable assigned to y_t, μ a constant term or mean level term, S_t is a seasonal component that can often be represented as $S_t = \sum_{j=1}^{4}[\beta_{1j}\sin\left(\frac{2\pi jt}{12}\right) + \beta_{2j}\cos(\frac{2\pi jt}{12})]$, $x_t = \frac{t}{12}$ is the linear trend function, ω is the linear trend and N_t is the residual term that is assumed to be autoregressive of the order of 1 [AR(1)].

The trends estimated in this study concern the average monthly anomalies, which do not take seasonality into account, so equation [4.1] reduces to:

$$y_t = \mu + \omega x_t + N_t, \quad t = 1, \dots, T \tag{4.2}$$

The first-order autoregressive model allows noise to be (auto)correlated between successive measurements, with $Corr(N_t, N_{t-1}) = \phi$, which is typically positive for ozone data. This positive autocorrelation may arise from various natural factors that give rise to slightly variable changes in N_t over time. For example, long-term variations in ozone are influenced by large-scale dynamic circulation processes, which are positively autocorrelated across meteorology. These processes are reflected in climate patterns such as the North Atlantic Oscillation or the Arctic Oscillation, which are correlated with tropopause pressure (Appenzeller et al., 2000; Weiss et al., 2001). Such natural factors may not always be well known or measurable, and the lagged N_{t-1} value can sometimes be considered a proxy to represent these natural factors

that dynamically influence the current N_t value. For ozone data, commonly considered extensions of the basic statistical model in [4.2] include additional explanatory "proxy" variables, when possible, to account for some types of natural dynamic and chemical influences on ozone, such as the Quasi-Biennial Oscillation (QBO), solar cycle variations, and other climate pattern oscillations.

As shown by Tiao et al. (1990) and Weatherhead et al. (1998), autocorrelation is an aspect to consider, but in this work, it was not considered. For this reason, equation [4.2] is further reduced to:

$$y_t = \mu + \omega x_t, \qquad t = 1, \dots, T \qquad [4.3]$$

4.1.2 Least Absolute Deviation Regression

Least Absolute Deviation (LAD) regression is a resistant and nonparametric regression method fitting the paired data to the linear model using a robust and resistant LAD method (Rice and White, 1964; Barrodale, 1968; Wong and Schneider Jr., 1989; Calitz and Rüther, 1996; Santer et al., 2000). This technique derives from an algorithm by Barrodale and Roberts (1974). This algorithm is resistant to outliers and computes a solution l_1 for an overdetermined system of m linear equations in n unknowns, i.e., given equations:

$$\sum_{j=1}^{n} a_{i,j} x_j = b_i \ for \ i = 1, 2, \dots, m, with \ m \geq n \qquad [4.4]$$

the algorithm determines a vector $x = \{x_j\}$ which minimizes the sum of the absolute values of the residuals

$$e(x) = \sum_{i=1}^{m} |b_i - \sum_{j=1}^{n} a_{i,j} x_j| \qquad [4.5]$$

A typical application of the algorithm is to solve the l_1 linear data fit problem. Suppose that data consisting of m points with co-ordinates (t_i, y_j) should be approximated by a linear approximating function $\alpha_1 \phi_1(t) + \alpha_2 \phi_2(t) + \cdots + \alpha_n \phi_n(t)$ in the l_1 norm. This is equivalent to finding an l_1 solution to the system of linear equations:

$$\sum_{j=1}^{n} \phi_j(t_j) \alpha_j = y_i \ for \ i = 1, 2, \dots, m \qquad [4.6]$$

Suppose the data contains some wild points (i.e., values of the dependent variable that are very inaccurate compared to the overall accuracy of the data). In that case, it is advisable to calculate an l_1 approximation rather than an l_2 (least-squares) approximation, or an l_∞ approximation. The algorithm is a modification of the linear programming simplex method applied to the primal formulation of the l_1 problem. A feature of the routine is its ability to pass through several simplex vertices at each iteration.

4.1.3 Theil-Sen regression

Theil-Sen regression involves the computation of the slopes between every possible pair of points in the time series, taking the median value as the trend estimate. It is called the Theil-Sen estimator and can be significantly more accurate than simple linear regression for skewed and heteroskedastic data and shows performances similar to non-robust least squares even for normally distributed data in terms of statistical power (Theil, 1950; Siegel and Benson, 1982; Helsel and Hirsch, 1992). Consider a multiple linear regression model (Dang et al., 2008):

$$Y_i = \alpha + X_i^\top \beta + \epsilon_i, \qquad i = 1, \dots, n \qquad [4.7]$$

where α is the intercept and β is a p-dimension parameter, and $\epsilon_1, \dots, \epsilon_n, \epsilon$ are independent and identically distributed (i.i.d.) random errors.

Starting with a simple linear regression $p = 1$, geometrically to estimate the slope β, only two distinct points $(X_i, Y_i), (X_j, Y_j)$ $(X_i \neq X_j)$, are needed and an estimator of the slope β is $b_{i,j} = \frac{(Y_i - Y_j)}{(X_i - X_j)}$. Alternatively, with any two distinct points, the sum of squares of the residuals is $(Y_i - \alpha - \beta X_i)^2 + (Y_j - \alpha - \beta X_j)^2$, which is minimized when α, β satisfy the equations:

$$Y_i - \alpha - \beta X_i = 0, \quad Y_j - \alpha - \beta X_j = 0 \qquad [4.8]$$

The solutions $a_{i,j} = Y_i - b_{i,j} X_i$ and $b_{i,j} = \frac{(Y_i - Y_j)}{(X_i - X_j)}$ are the least squares estimators (Lanzante, 1996). A robust estimator $\tilde{\beta}_n$ of the slope β is then the median of these least squares estimates:

$$\tilde{\beta}_n = Med\left\{ b_{i,j} = \frac{(Y_i - Y_j)}{(X_i - X_j)} : X_i \neq X_j, 1 \leq i < j \leq n \right\} \qquad [4.9]$$

where $Med\{B_j : j \in J\}$ denotes the median of the numbers $\{B_j : j \in J\}$. This is the well-known Theil-Sen estimator which is robust with high breakdown point. If we are only dealing with the estimate of the slope β, no assumption of identifiability on the error is necessary. In order to estimate the intercept, however, some identifiability conditions on the error distribution are essential. Now suppose that the error has a symmetric distribution around zero, and this is a sufficient condition. Then, similarly, the intercept can be estimated using the median of the least squares estimates:

$$\tilde{\alpha}_n = Med\left\{ a_{i,j} = \frac{(Y_j X_i - Y_i X_j)}{(X_i - X_j)} : X_i \neq X_j, 1 \leq i < j \leq n \right\} \qquad [4.10]$$

These result in a median estimator per component $(\tilde{\alpha}_n, \tilde{\beta}_n)$ of the parameter (α, β). It is known that a median per component estimator may be a very poor estimator. To overcome this flaw, it could use the robust

$\tilde{\beta}_n$ to construct a robust estimator of the intercept α; for example, $Med\{Y_i - \tilde{\beta}_n X_i : 1 \le i,\ j \le n\}$. Alternatively, we can estimate (α, β) simultaneously using the multivariate median:

$$(\tilde{\alpha}_n, \tilde{\beta}_n) = Mmed\{(a_{i,j}, b_{i,j}) : X_i \ne X_j, 1 \le i < j \le n\} \qquad [4.11]$$

where $Mmed\{B_j : j \in J\}$ stands for the multivariate median of the vectors $\{B_j \in \mathbb{R}^d : j \in J\}$.

4.1.4 LOTUS regression

LOTUS regression relies on the classic multiple linear regression method, which estimates time series variability from explanatory variables using the general least squares approach. The explanatory or proxy variables used in the LOTUS model are the Quasi-Biennial Oscillation (QBO), the El Niño-Southern Oscillation (ENSO), the 11-year solar cycle, the Stratospheric Oscillation aerosol optical depth (sAOD) and a long-term trend. Independent linear trend (ILT) terms are used to evaluate long-term changes before and after the peak of ozone-depleting substances (ODS), i.e. before January 1997 and after January 2000. The LOTUS model is applied to weightless ozone records based, for example, on measurement uncertainty. Most datasets are provided as monthly average time series and are seasonally adjusted within the LOTUS model using Fourier components representing annual and semi-annual variations. The fitting of the deseasonalised times series is based on the following equation:

$y(z,t) = \beta_1(z,t) \cdot QBO_1(t) + \beta_2(z,t) \cdot QBO_2(t) + \beta_3(z,t) \cdot ENSO(t) + \beta_4(z,t) \cdot Solar(t) + \beta_5(z,t) \cdot sAOD(t) + (\beta_6(z,t) + \beta_7(z,t)(t-t_1)) \cdot L_{pre}(t) + (\beta_8(z,t) + \beta_9(z,t)(t-t_2)) \cdot L_{post}(t) + \beta_{10}(z,t) \cdot Gap(t) + \varepsilon(z,t)$ [4.12]

$y(z,t)$ is the monthly mean ozone anomaly time series at altitude z, $\beta_{1-10}(z,t)$ are the fitted coefficients and $\varepsilon(z,t)$ represents the residual term. QBO_1 and QBO_2 are two orthogonal components of the QBO calculated with principal component analysis. No lag is applied to the $ENSO$, $sAOD$ and $Solar$ F10.7 proxies. Regarding the trend terms, $L_{pre}(t)$, $L_{post}(t)$ and $Gap(t)$ are written as follows:

$$L_{pre}(t) = \begin{cases} 1 \ if \ t \le t_1 \\ 0 \ if \ t > t_1 \end{cases}$$

$$L_{post}(t) = \begin{cases} 0 \ if \ t \le t_2 \\ 1 \ if \ t > t_2 \end{cases}$$

$$Gap(t) = \begin{cases} 0 \ if \ t \le t_1 \\ 1 \ if \ t > t_1 \ and \ t \le t_2 \\ 0 \ if \ t > t_2 \end{cases}$$

t_1 corresponds to 1 January 1997 and t_2 to 1 January 2000 (Godin-Beekmann et al., 2022; Petropavlovskikh et al., 2019).

4.1.5 Mann-Kendal test

The purpose of the Mann-Kendal (MK) test (Kendall, 1975; Mann, 1945) is to statistically evaluate whether there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward (downward) trend means that the variable increases (decreases) consistently over time, but the trend may or may not be linear. The MK test can be used in place of a parametric linear regression analysis, which can be used to test whether the slope of the estimated linear regression line is different from zero. Regression analysis requires that the residuals of the fitted regression line be normally distributed; an assumption not required by the MK test, i.e., the MK test is a non-parametric (distribution-free) test.

The MK test is based on the following hypotheses:

- When no trend is present, the measurements obtained over time are independent and identically distributed. The independence assumption means that observations are not serially correlated over time;
- Observations obtained over time represent actual conditions at the time of sampling;
- Sample collection, processing and measurement methods provide unbiased and representative observations of underlying populations over time.

There is no requirement that the measurements be normally distributed or that the trend, if present, be linear. The MK test may be calculated if there are missing values and values below one or more limits of detection (LD), but test performance will be adversely affected by such events.

The MK test tests whether to reject the null hypothesis (H_0) and accept the alternative hypothesis (H_a), where:

- H_0: No monotonic trend;
- H_a: Monotonic trend is present.

The initial assumption of the MK test is that the H_0 is true and that the data must be convincing beyond a reasonable doubt before H_0 is rejected and H_a is accepted.

4.1.6 Calculating trends

Decadal ozone trends are estimated using the five distinct linear regression methods described above on the average monthly anomalies determined from the LC categorised stations and their conjunction with the MC cluster. The dataset was separated into four vertical ranges (300-200 hPa, 200-100 hPa, 100-50 hPa, and 50-1 hPa) and five latitudinal sectors (NP (60°N-90°N), NH (30°N-60°N), TR (30°S-30°N), SH (60°S-30°S), and SP (90°S-60°S)) to determine the average monthly anomalies. The average monthly anomalies are calculated using data from all stations within each latitudinal and vertical range, using the following formula:

$$anomaly_{m\_x(y)} = m\_x(y) - mean(m\_x_{1978-2022}) \qquad [4.13]$$

And in percentage:

$$anomaly_{\%\_m\_x(y)} = \frac{m\_x(y) - mean(m\_x_{1978-2022})}{mean(m\_x_{1978-2022})} \cdot 100 \qquad [4.14]$$

Where $m\_x(y)$ is the average of the ozone values of month x of the year y, $mean(m\_x_{1978-2022})$ is the average of the ozone values for month x calculated from the average for that month for all years of the series (from 1978 to 2022). The difference between these two values gives the average anomaly for month x of the year y. A greater or lower amount of ozone than the average for the month under investigation is indicated by a positive or negative number for the average monthly anomaly. The trends are assessed using the various regressors after the series of average monthly anomalies by latitudinal sector and vertical range is determined.

In order to measure the effectiveness of the regression techniques, the median absolute deviation (MAD) of the residuals was also determined. For a univariate data set $X_1, X_2, \ldots, X_n$, the MAD is defined as the median of the absolute deviations from the median of the data:

$$\tilde{X} = median(X):$$

$$MAD = median(|X_i - \tilde{X}|) \qquad [4.15]$$

that is, starting from the residuals (deviations) from the median of the data, the MAD is the median of their absolute values.

4.2 Clusters comparing

The LC and LMC clusters are contrasted to quantify the impact of sampling error on the trend estimate. For each cluster in all latitude areas and all vertical ranges, the numbers of profiles and pressure levels with a valid value of the ozone concentration recorded in the database are shown in Tables 7 and 8, respectively.

	LC				LMC			
	300-200 hPa	200-100 hPa	100-50 hPa	50-1 hPa	300-200 hPa	200-100 hPa	100-50 hPa	50-1 hPa
NP	9403	9403	9403	9404	12504	12503	12502	12508
NH	34645	34647	34644	34702	41528	41536	41533	41592
TR	5392	5394	5389	5394	12564	12568	12561	12573
SH	3171	3171	3171	3171	4327	4327	4326	4327
SP	6005	6005	6004	6007	7883	7883	7883	7886

Table 7. Number of profiles present for long coverage (left) and long and medium coverage (right) clusters, for all latitude regions and four vertical ranges.

	Long coverage				Long and Medium coverage			
	300-200 hPa	200-100 hPa	100-50 hPa	50-1 hPa	300-200 hPa	200-100 hPa	100-50 hPa	50-1 hPa
NP	1534733	2699272	2410944	4254137	2139130	3778380	3372031	5666960
NH	3556187	6392101	6140613	12944258	4904274	8755592	8139961	16470127
TR	1554689	2385231	2233437	4550228	3409766	5229950	4798379	9224337
SH	1069710	1924002	1714171	3229943	1171607	2102759	1873122	3458300
SP	1317119	2253757	2140296	4314751	1709714	2943867	2673790	4859461

Table 8. Number of records present for long coverage (left) and long and medium coverage (right) clusters, for all latitude regions and four vertical ranges.

From the amount of data summarized in Table 8, it is possible to calculate the percentage of the data from the MC cluster added to the LC for obtaining the LMC, as reported in Table 9.

	300-200 hPa	200-100 hPa	100-50 hPa	50-1 hPa
NP	0.28	0.29	0.29	0.25
NH	0.27	0.27	0.25	0.21
TR	0.54	0.54	0.53	0.51
SH	0.09	0.09	0.08	0.07
SP	0.23	0.23	0.20	0.11

Table 9. Percentage of additions to the long coverage cluster to obtain the long and medium coverage cluster, for all latitude regions at the indicated vertical ranges.

Additions for all latitudinal regions do not exceed 30% of the total, except for the TR region, which exceeds 50%. This is due, as highlighted in Table 4 (in Chapter 3), to the greater number of stations belonging to the MC cluster (10 stations) compared to the LC cluster (4 stations).

To quantify the values of using LMC vs LC, the estimates of the ozone concentration trends were calculated and the MK test was applied to check whether the estimated trend was significant or not. Tables 10, 11 and 12 show only the results of the MK test for three different periods 1978-2000, 2000-2022 and 1978-2022, respectively, and include values for all clusters. The analysis presented below aims at quantifying the effect of spatial sampling on the trends and, therefore, on knowledge of the ozone variability over the years.

Comparing the results for all periods, it can be seen that the LC cluster has significant trend estimates in more cases than the MC cluster. The latter rarely improves the significance of the LMC cluster compared to the LC at different pressure ranges. This means that there are no big advantages to using the LMC cluster instead of the LC. This conclusion is likely related to the higher quality of the time series in the LC cluster and the following analysis is carried out using the LC cluster. Figures 30-34 show the comparison between LC and LMC clusters of the trend per decade (in percentage) estimated at all latitudes and vertical ranges using Least-Square Linear (LIN), Least Absolute Deviation (LAD) and Theil-Sen (TS) linear regressors and the LOTUS multiple linear regression, evaluated using LIN, LAD and TS linear regressors.

1978-2000		SP	SH	TR	NH	NP
MC	**50-1 hPa**	I	NT	NT	D	D
	100-50 hPa	NT	NT	NT	D	D
	200-100 hPa	NT	D	D	D	D
	300-200 hPa	I	NT	D	D	D
LC	**50-1 hPa**	D	D	I	D	D
	100-50 hPa	D	NT	I	D	NT
	200-100 hPa	D	NT	I	D	NT
	300-200 hPa	NT	NT	I	D	I
LMC	**50-1 hPa**	D	D	I	D	D
	100-50 hPa	D	NT	I	D	D
	200-100 hPa	D	NT	I	D	D
	300-200 hPa	NT	NT	NT	D	D

Table 10. MK test results for the period 1978-2000 for all clusters, for all latitudinal regions at all vertical ranges. Legend: D (Decreasing trend), I (Increasing trend), NT (No trend).

2000-2022		SP	SH	TR	NH	NP
MC	**50-1 hPa**	NT	I	NT	I	NT
	100-50 hPa	NT	NT	I	NT	I
	200-100 hPa	NT	NT	I	NT	I
	300-200 hPa	NT	NT	I	NT	I
LC	**50-1 hPa**	I	D	I	D	D
	100-50 hPa	I	D	NT	I	NT
	200-100 hPa	NT	D	NT	I	NT
	300-200 hPa	NT	NT	NT	I	NT
LMC	**50-1 hPa**	I	D	I	NT	D
	100-50 hPa	I	D	I	I	NT
	200-100 hPa	NT	NT	I	I	NT
	300-200 hPa	NT	NT	NT	I	NT

Table 11. MK test results for the period 2000-2022 for all clusters, for all latitudinal regions at all vertical ranges. Legend: D (Decreasing trend), I (Increasing trend), NT (No trend).

1978-2022		SP	SH	TR	NH	NP
MC	**50-1 hPa**	NT	I	NT	I	NT
	100-50 hPa	NT	D	I	D	D
	200-100 hPa	NT	NT	D	D	D
	300-200 hPa	I	D	NT	D	D
LC	**50-1 hPa**	D	I	I	D	D
	100-50 hPa	D	D	NT	D	I
	200-100 hPa	D	D	I	D	I
	300-200 hPa	NT	NT	NT	NT	I
LMC	**50-1 hPa**	D	I	I	D	D
	100-50 hPa	D	D	D	D	D
	200-100 hPa	D	D	NT	D	D
	300-200 hPa	NT	NT	NT	D	D

Table 12. MK test results for the period 1978-2022 for all clusters, for all latitudinal regions at all vertical ranges. Legend: D (Decreasing trend), I (Increasing trend), NT (No trend).

Due to the smaller amount of data available, the trends estimated for MC in the period 1978-2000 (Table 10) are significant at fewer latitude/vertical ranges than for LC. Despite the increase in the number of ascents in the period 2000-2022 (Table 11), the LC cluster outperforms the MC in the same period. In the following time series analysis, the year 2000 was chosen as the turning point in the ozone trend according to Petropavlovskikh et al., (2019), and due to the increase in the number of ascents available and the lower signal variability in the time ozone time series.
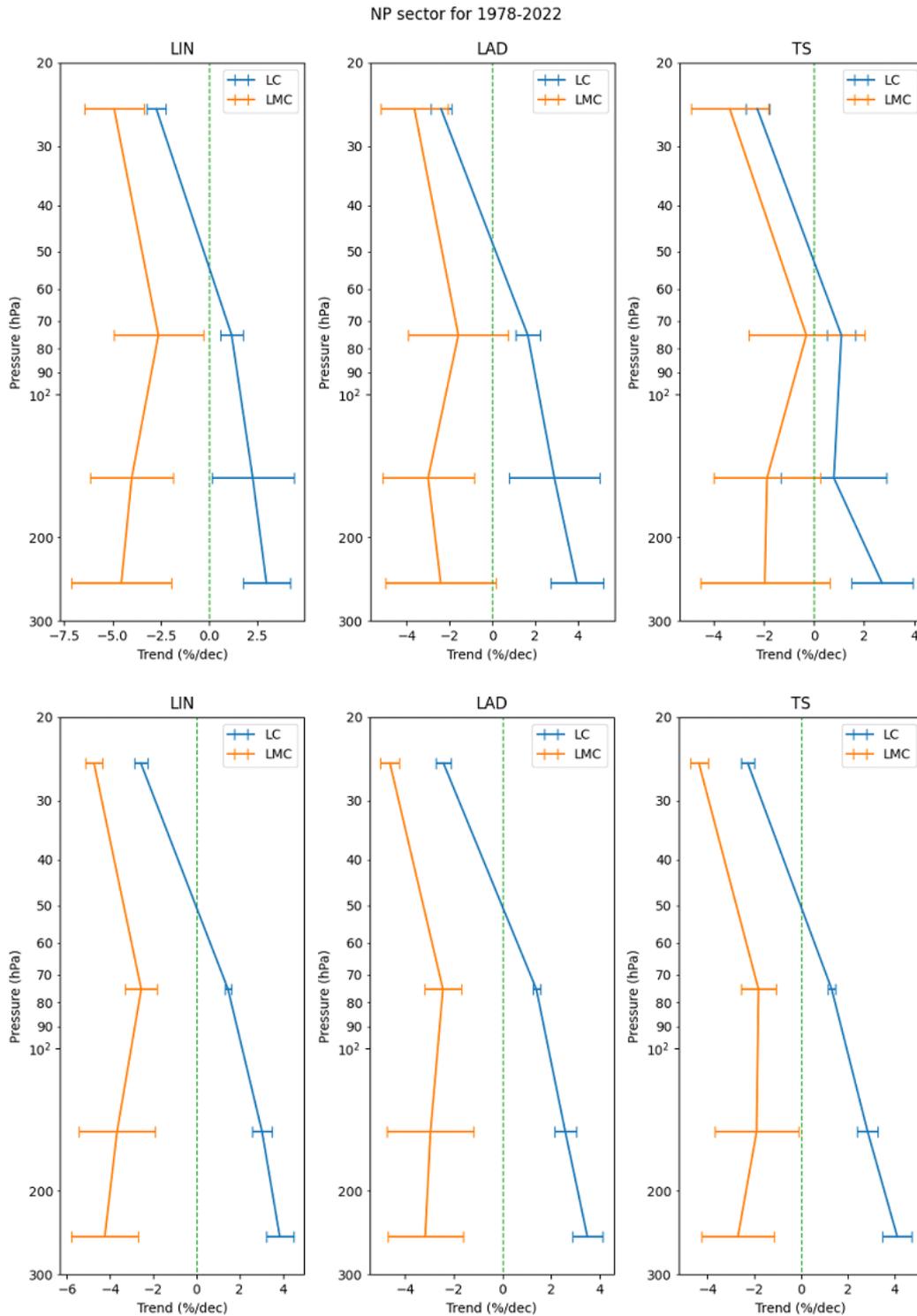
Figure 30. Trend estimates in percentage per decade (%/dec) obtained with the Least-Square Linear (LIN, left panel), Least Absolute Deviation (LAD, middle panel) and Theil-Sen (TS, right panel) regressors (upper row for linear regressor, lower row for LOTUS regression) for the NP latitudinal sector at 250, 150, 75, and 25 hPa for the period 1978-2022 for "Long Coverage" (LC) and "Long and Medium Coverage" (LMC) clusters. Non-significant trend estimates for the MK test are not reported at their respective levels.
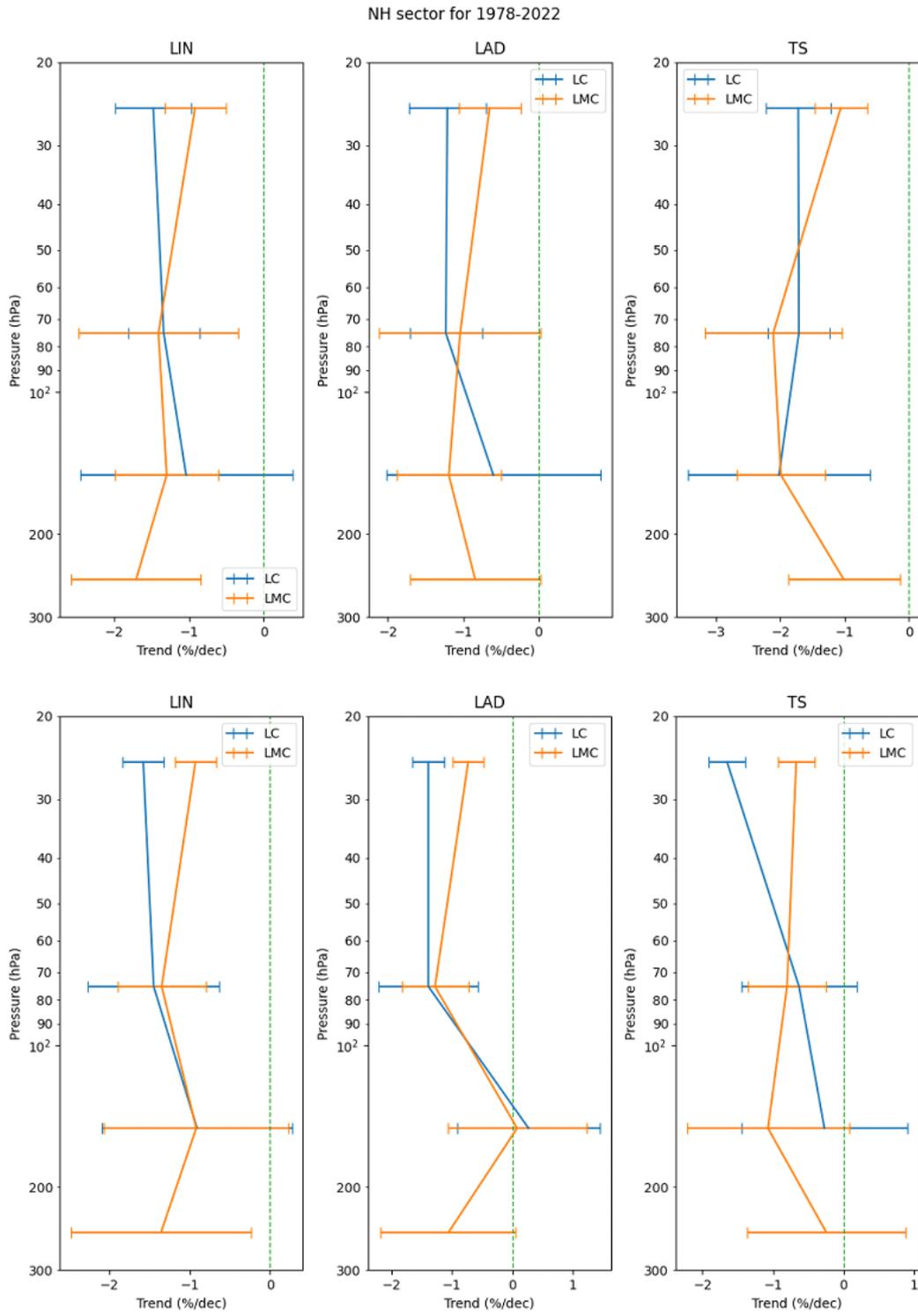
Figure 31. The same as Figure 30 but for the NH latitudinal sector.
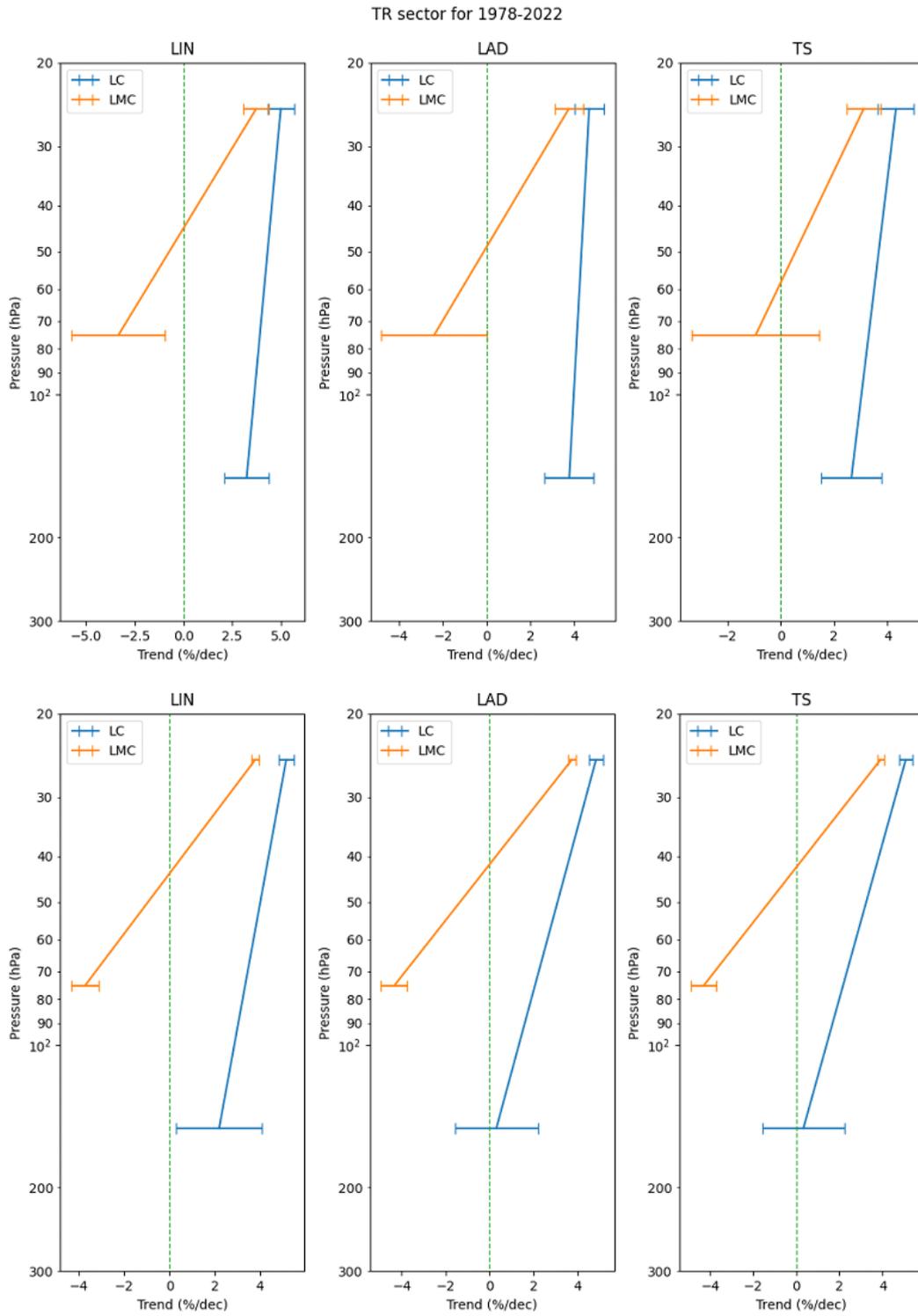
Figure 32. The same as Figure 30 but for the TR latitudinal sector.
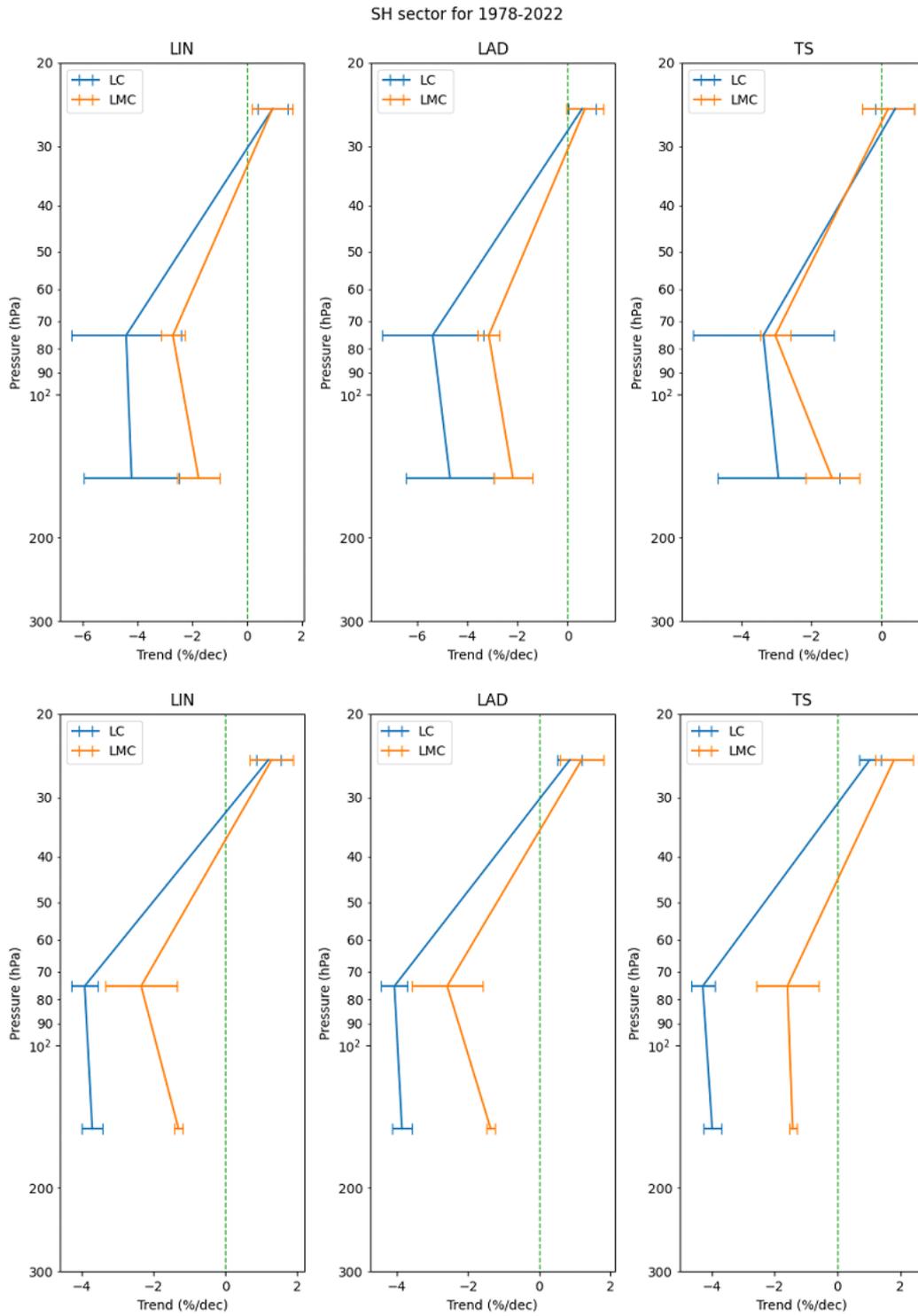
Figure 33. The same as Figure 30 but for the SH latitudinal sector.
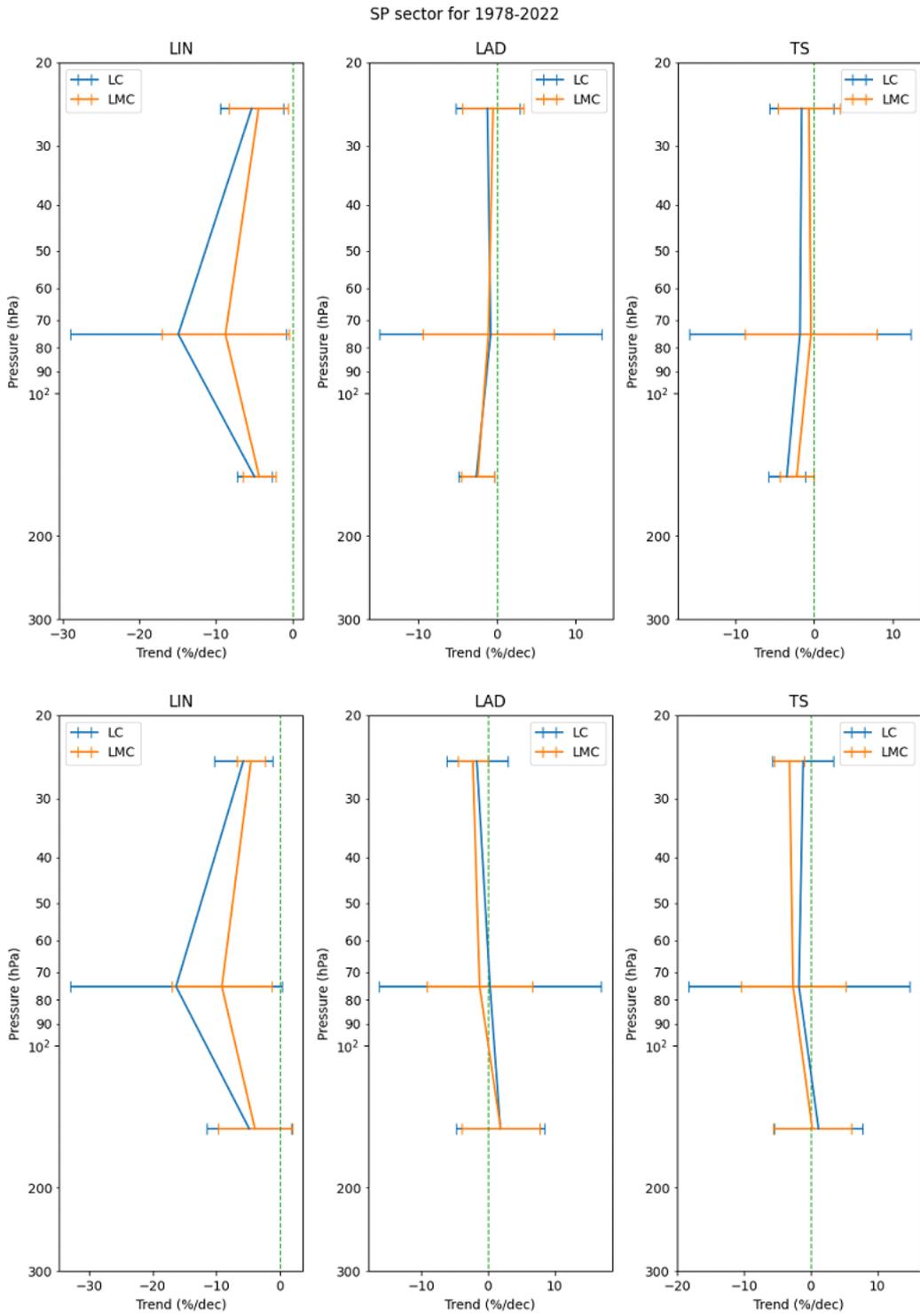
Figure 34. The same as Figure 30 but for the SP latitudinal sector.

For the 1-50 hPa layer, the estimates between the two clusters, for all latitudinal sectors, differ by approximately 1%, except the NP which reaches 2%. This small difference is due to the very large number of data involved in the estimates for both clusters, thus making the differences minimal. For the 50-100 hPa layer, the differences remain small, less than 0.4% for the NH and 2% for the SH; for the SP and the NP, the difference increases, reaching approximately 5% and 4% respectively. For the TR, however, the MK test for the LC cluster reveals a non-significant trend while for the LMC a significant negative trend. Regarding the 100-200 hPa layer, the NH still has a small difference between the cluster estimates (around 0.6%), as does the SP of around 1.5% and the SH of around 2.5%. For the TR, instead, the LMC cluster reveals a non-significant trend, therefore it is not possible to calculate a difference with the trend of the LC cluster that reveals a significant positive trend. Same situation as the previous layer for the NP but the difference this time is around 7%. For the 200-300 hPa layer, most trends are non-significant due to the limited amount of data available in this vertical range, except for the NP, with differences of approximately 8%, and the NH limited to the LMC cluster. The trends estimated with the LOTUS regressor are in agreement with those calculated with the linear regressors, except for NH (for the LAD regressor at 100-200 hPa) and SP (for the LAD and TS regressors at 100-200 hPa) which show positive trends for the LOTUS regressor compared to negative trends for the linear regressors. Furthermore, for LOTUS estimates the resulting structural uncertainty, in terms of maximum slope difference, is lower than for linear estimates, as shown in Table 14.

4.3 Trends comparisons

As mentioned above, the differences between the LC and LMC clusters can be considered minimal and therefore it was decided to use only the subsampled LC cluster for the analysis of the vertical ranges of the UT/LS. Below, Table 13, Table 15, Table 16 and Table 17 report the results obtained for the LC cluster, analysing the percentage per decade (%/dec) for each vertical and latitude range, for the pre-2000, post-2000 and the entire 1978-2022 periods.

50-1 hPa		%/dec	SP	SH	TR	NH	NP
Long coverage cluster	**1978-2000**	**LIN**	-	-10.14(-9.9)	7.20(7.26)	-5.94(-6.4)	-9.58(-9.84)
		LAD	-	-8.34(-11.4)	8.06(7.39)	-5.52(-6.46)	-9.59(-15.0)
		TS	-	-9.01(-10)	7.52(7.56)	-6.45(-6.6)	-7.51(-14.1)
	2000-2022	**LIN**	6.29(6.56)	-2.40(-2.11)	2.09(2.13)	-1.33(-1.47)	-5.83(-5.56)
		LAD	6.53(6.67)	-2.79(-2.12)	1.80(2.05)	-1.09(-1.52)	-5.97(-5.48)
		TS	6.94(6.72)	-2.06(-2.07)	2.02(2.24)	-0.71(-1.5)	-5.42(-5.73)
	1978-2022	**LIN**	-5.30(-5.72)	0.94(1.2)	4.99(5.17)	-1.48(-1.58)	-2.74(-2.58)
		LAD	-1.21(-1.65)	0.57(0.86)	4.69(4.85)	-1.21(-1.39)	-2.39(-2.42)
		TS	-1.57(-1.13)	0.39(1.03)	4.32(5.08)	-1.72(-1.65)	-2.27(-2.28)

Table 13. Percentage per decade (%/dec) of the trend estimates obtained with the Linear (LIN), Least Absolute Deviation (LAD) and Theil-Sen (TS) regressors for all latitudinal sectors at the 50-1 hPa vertical range. In brackets, the respective values obtained from the LOTUS regression are shown. Legend: Positive value (increasing trend); Negative value (decreasing trend); NT (No Trend).

At 50-1 hPa (Table 13) for each latitudinal sector, all trends are significant. Despite this, not all estimated trends can be considered reliable. Indeed, in the SP for the period 1978-2000, as shown in Figure 21 (Chapter 3), there is an imbalance in the coverage of pre-1986 and post-1986 data which causes the trend uncertainty to increase. For this reason, the data will not be used for the comparison. This problem has less weight on the trend estimate for the period 1978-2022 since there is more data available. In the next work, the period will be restricted to 1986-2000. Table 13 reveals that all trends have uncertainty less than 2% except for the SP, which for the period 1978-2022 exceeds 4% for both the linear regressors and for the LOTUS, and for the NP, which for the period 1978-2000 reaches 5% but only for the LOTUS regressor.

Table 14 shows the maximum differences between the slopes of the estimated trends using the considered regression methods, which helps understand the structural uncertainty.

		%/dec	Maximum slope difference			
			50-1 hPa	100-50 hPa	200-100 hPa	300-200 hPa
Long coverage cluster	1978-2000	SP	-	-	-	NT
		SH	1.8(1.47)	NT	NT	NT
		TR	0.8(0.3)	7.0(1.65)	-	-
		NH	0.9(0.2)	0.8(0.75)	0.6(1.16)	2.1(0.88)
		NP	2.0(5.18)	NT	NT	6.0(2.64)
	2000-2022	SP	0.7(0.16)	2.3(0.83)	NT	NT
		SH	0.7(0.05)	1.5(0.5)	1.9(0.61)	NT
		TR	0.3(0.19)	NT	NT	NT
		NH	0.6(0.05)	1.0(0.06)	2.8(0.04)	1.4(0.43)
		NP	0.5(0.25)	NT	NT	NT
	1978-2022	SP	4.1(4.6)	14.0(16.6)	2.3(6.7)	NT
		SH	0.6(0.3)	2.0(0.4)	1.7(0.3)	NT
		TR	0.6(0.3)	NT	1.1(1.9)	NT
		NH	0.5(0.3)	0.5(0.8)	1.4(1.2)	NT
		NP	0.5(0.3)	0.5(0.2)	2.1(0.5)	1.2(0.6)

Table 14. The maximum difference in trend slopes is estimated for all vertical ranges (NT means No Trends). In brackets, the respective values obtained from the LOTUS regression are shown.

The differences for SP are quite large, especially in the period 1978-2000, because of the limited data available and the consequent larger structural uncertainty. Regarding SH, for the same period, trends have been estimated using only two stations: therefore, although the uncertainties are small, caution is recommended in using these estimates. For the other latitudinal sectors, the uncertainties are small, and the data amount and the number of stations is sufficient to characterize the vertical variability of ozone concentration.

100-50 hPa		%/dec	SP	SH	TR	NH	NP
Long coverage cluster	**1978-2000**	LIN	-	NT	8.91(8.16)	-8.32(-8.42)	NT
		LAD	-	NT	10.29(7.08)	-7.58(-7.67)	NT
		TS	-	NT	3.28(6.51)	-8.23(-8.22)	NT
	2000-2022	LIN	4.83(6.96)	-8.53(-7.73)	NT	3.40(3.32)	NT
		LAD	2.52(7.29)	-9.37(-7.69)	NT	4.07(3.33)	NT
		TS	2.78(6.46)	-8.80(-8.19)	NT	4.45(3.27)	NT
	1978-2022	LIN	-14.9(-16.3)	-4.42(-3.91)	NT	-1.34(-1.45)	1.16(1.46)
		LAD	-0.80(0.33)	-5.37(-4.06)	NT	-1.23(-1.39)	1.66(1.4)
		TS	-1.78(-1.75)	-3.37(-4.28)	NT	-1.71(-0.63)	1.09(1.31)

Table 15. The same as Table 13 but for the vertical range 100-50 hPa.

For the 100-50 hPa vertical range, presented in Table 15, the estimates of the SP trend for the period 1978-2000 are considered not reliable in the same way as for the 50-1 hPa vertical range. For the period 1978-2022, the paucity of data before 1986 does not affect the trend estimate if only robust regression methods are used. SH has discrepancies for the three linear regressors used, with an uncertainty of 2%, while it is only 0.4% for the LOTUS regressor; TR has a valid trend only for the period 1978-2000 but has a discrepancy, with an uncertainty of 7%, while it is only 0.4% for the LOTUS regressor. The only sector with all valid trend estimates and no discrepancies is NH, in addition to the NP, but only for the period 1978-2022, both with uncertainties less than 1% for both linear and LOTUS regressors. This is due to the higher data density for NH, in this vertical range where the variability is greater.

200-100 hPa		%/dec	SP	SH	TR	NH	NP
Long coverage cluster	**1978-2000**	LIN	-	NT	-	-9.88(-9.62)	NT
		LAD	-	NT	-	-10.05(-8.46)	NT
		TS	-	NT	-	-9.42(-9.03)	NT
	2000-2022	LIN	NT	-6.89(-5.84)	NT	7.98(9.02)	NT
		LAD	NT	-4.92(-6.21)	NT	7.52(9.05)	NT
		TS	NT	-6.14(-6.45)	NT	10.37(9.01)	NT
	1978-2022	LIN	-4.94(-4.79)	-4.22(-3.7)	3.24(2.2)	-1.04(-0.91)	2.25(3.04)
		LAD	-2.62(1.88)	-4.68(-3.85)	3.78(0.31)	-0.60(0.27)	2.90(2.59)
		TS	-3.46(1.19)	-2.94(-3.98)	2.65(0.33)	-2.02(-0.27)	0.79(2.85)

Table 16. The same as Table 13 but for the vertical range 200-100 hPa.

Regarding the 200-100 hPa layer, reported in Table 16, differences are small for all latitudinal sectors. In the estimates, the small amount of data available is balanced by the less variable ozone content in the UT (Salby, 1995). Again, at these levels, SP is considered not reliable for the period 1978-2000. Trends for TR in the same period are also considered not reliable due to the much smaller data amount available before 1995 than after, making the estimation of the trend over a decadal period not feasible.

300-200 hPa		%/dec	SP	SH	TR	NH	NP
Long coverage cluster	**1978-2000**	LIN	NT	NT	-	-8.75(-9.15)	5.20(4.73)
		LAD	NT	NT	-	-8.18(-8.74)	11.21(7.37)
		TS	NT	NT	-	-6.66(-8.27)	10.02(6.74)
	2000-2022	LIN	NT	NT	NT	7.29(8.46)	NT
		LAD	NT	NT	NT	8.10(8.89)	NT
		TS	NT	NT	NT	6.74(8.55)	NT
	1978-2022	LIN	NT	NT	NT	NT	2.96(3.86)
		LAD	NT	NT	NT	NT	3.95(3.5)
		TS	NT	NT	NT	NT	2.72(4.12)

Table 17. The same as Table 13 but for the vertical range 300-200 hPa.

For the 300-200 hPa layer (Table 17), the situation is the same as at the 200-100 hPa layer, although the differences for NH and NP (for 1978-2022 only) are small. Furthermore, this is the vertical interval with the fewest trends judged significant by the MK test.

The most suitable vertical ranges for this study are 50-1 hPa and 100-50 hPa as they are the ones that have the largest number of data and therefore provide a more suitable data set for trend estimation.

4.4 Trends in UT/LS Ozone Vertical Profiles

The trend estimates presented in the previous paragraphs are influenced by the diversity of the amount of data available for each latitudinal sector. As highlighted in Table 8, the sector with the largest number of data is that of NH followed by TR. The trends of these two latitudinal sectors are compared with the literature at the 50-1 hPa and 100-50 hPa vertical ranges.

4.4.1 Northern Hemisphere mid-latitudes

Figures 35-38 show the trends calculated for the NH sector for the 50-1 hPa and 100-50 hPa vertical ranges, distinguishing the pre-2000 (1978-1999) and post-2000 (2000-2022) time series.
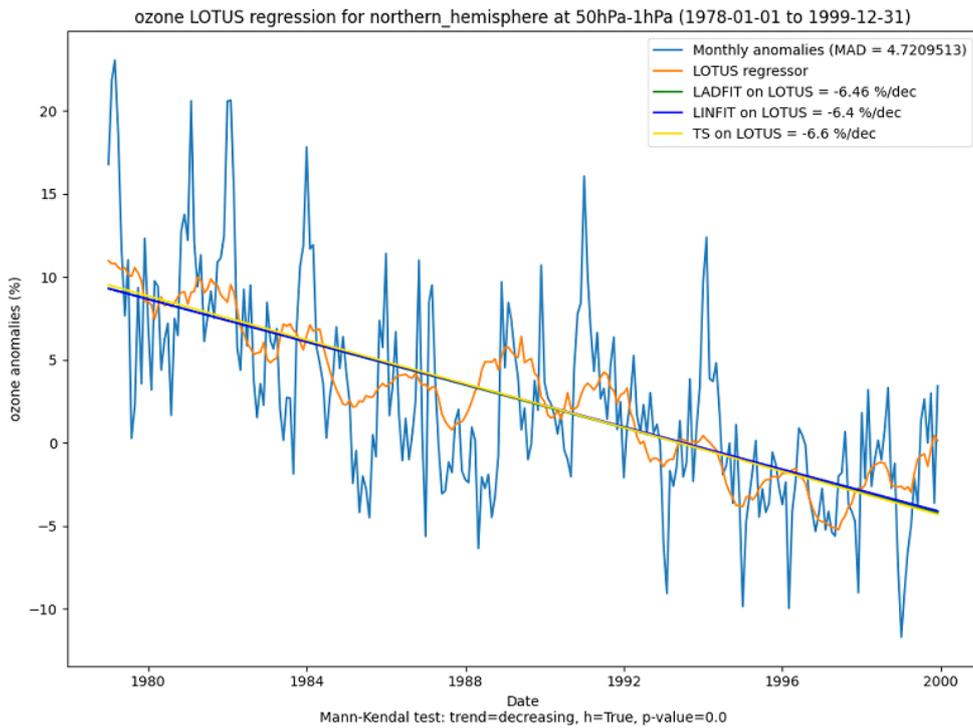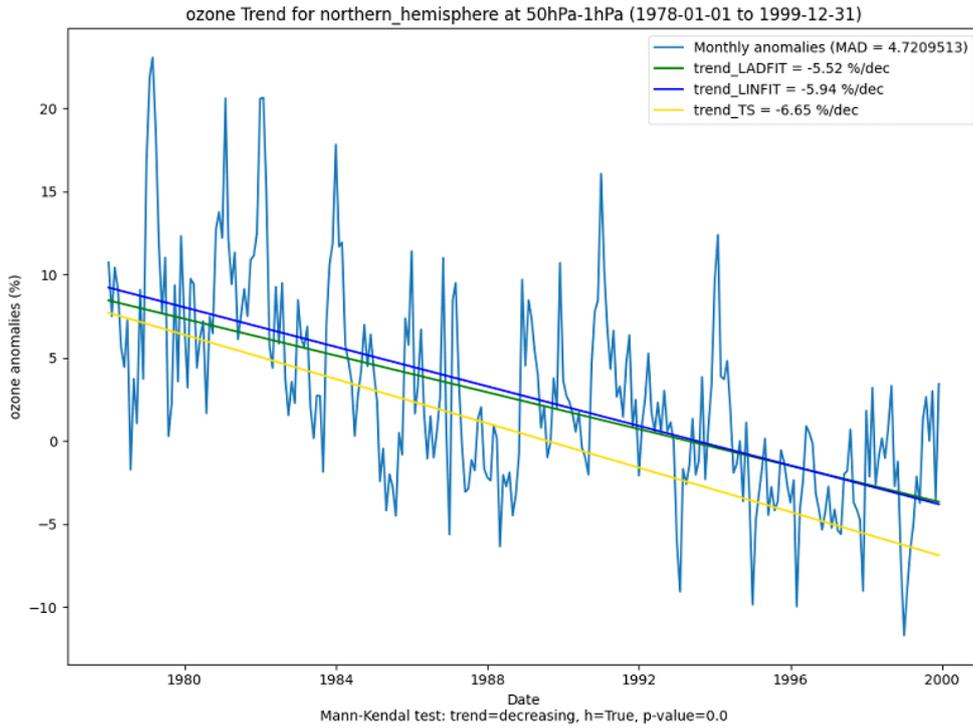
Figure 35. Trends estimations for the NH latitudinal sector at 50-1 hPa vertical range for the pre-2000 (1978-1999) period. On the top is the trend estimated with linear regression (LINFIT), least absolute deviation regression (LADFIT) and Theil-Sen regression (TS), and at the bottom is the trend estimated using the LOTUS regression, evaluated using LINFIT, LADFIT and TS linear regressors. The resulting percentage per decade (%/dec) for each regressor used and the MAD calculated on the average anomalies are shown in the figure legend. Finally, at the bottom centre is the result of the MK test.
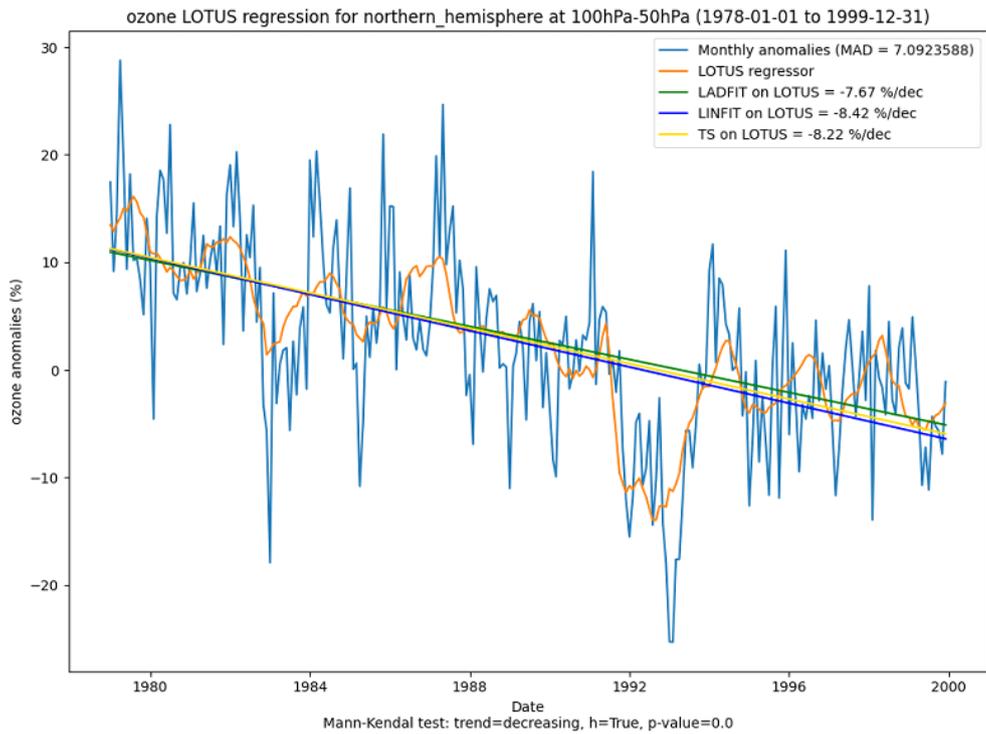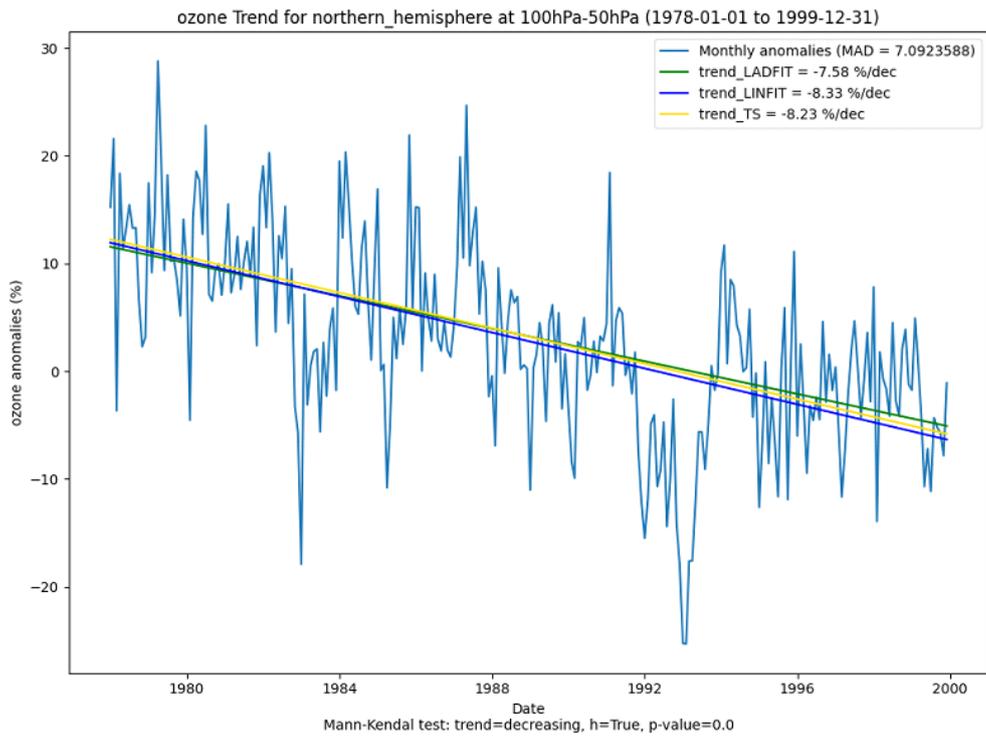
Figure 36. The same as Figure 35 but for the vertical range 100-50 hPa.
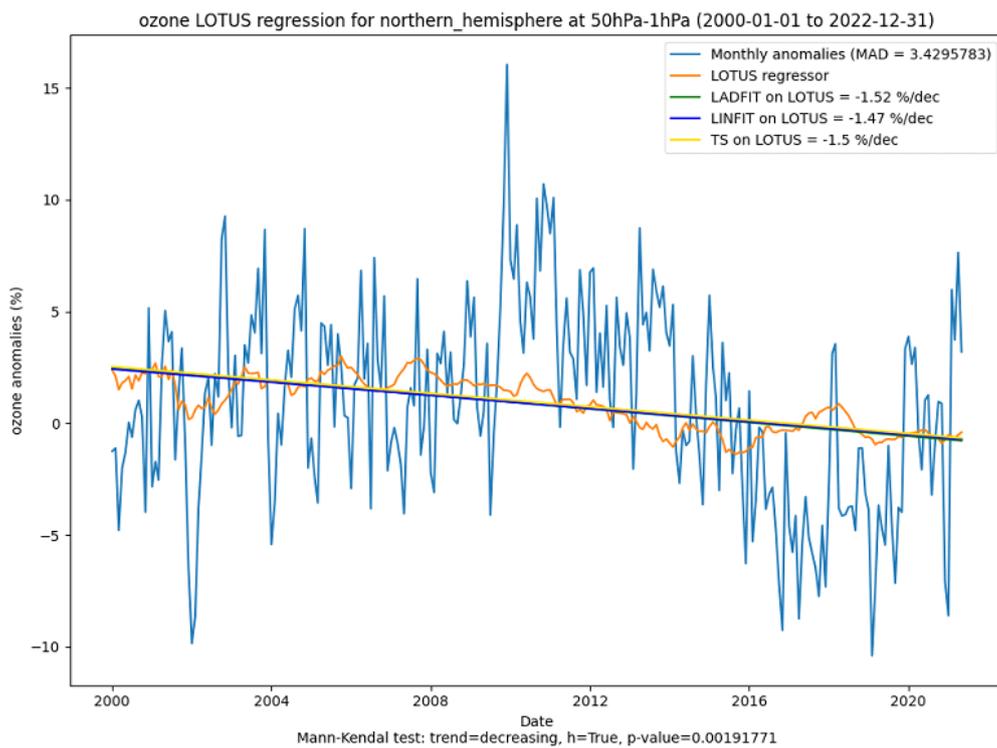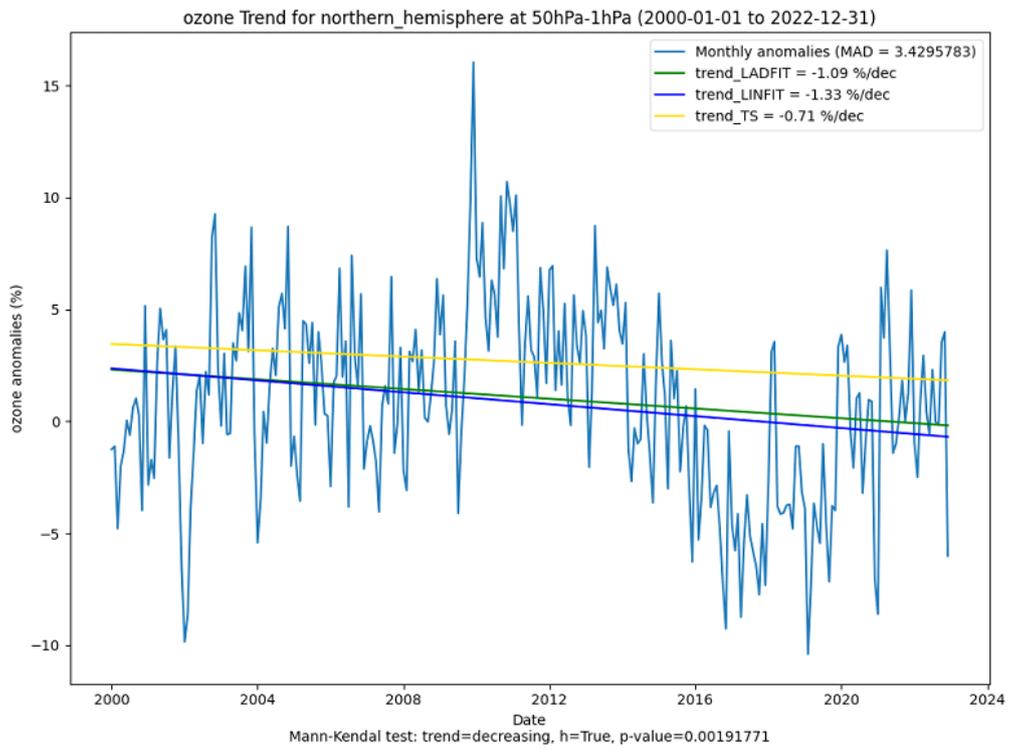
*Figure 37. The same as Figure 35 but for the post-2000 (2000-2022) period.*
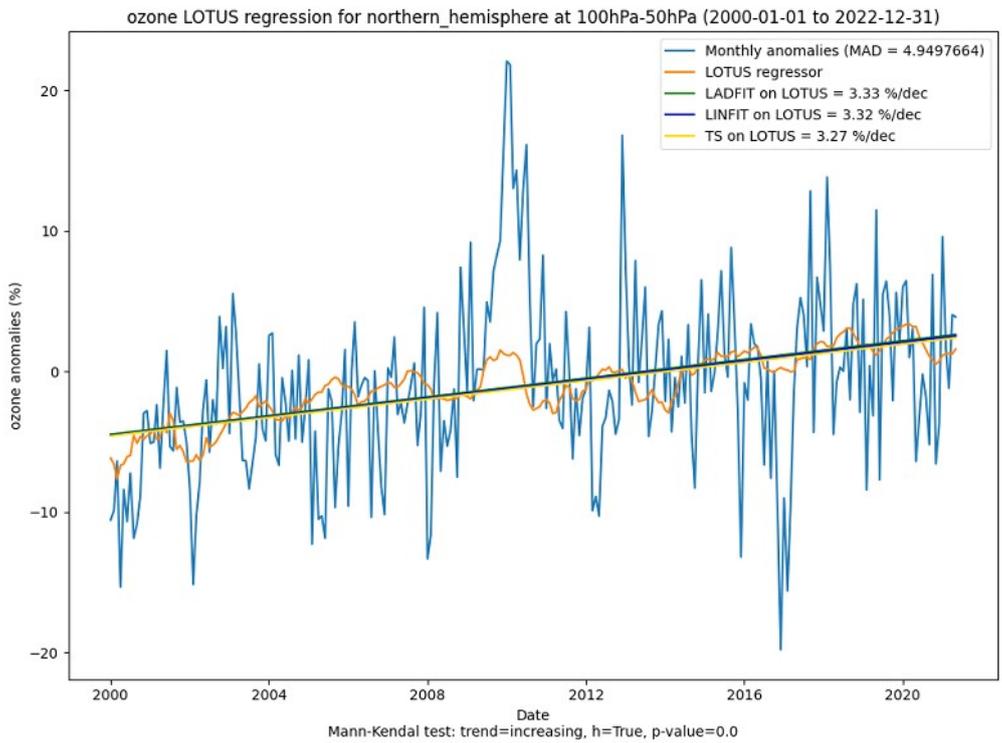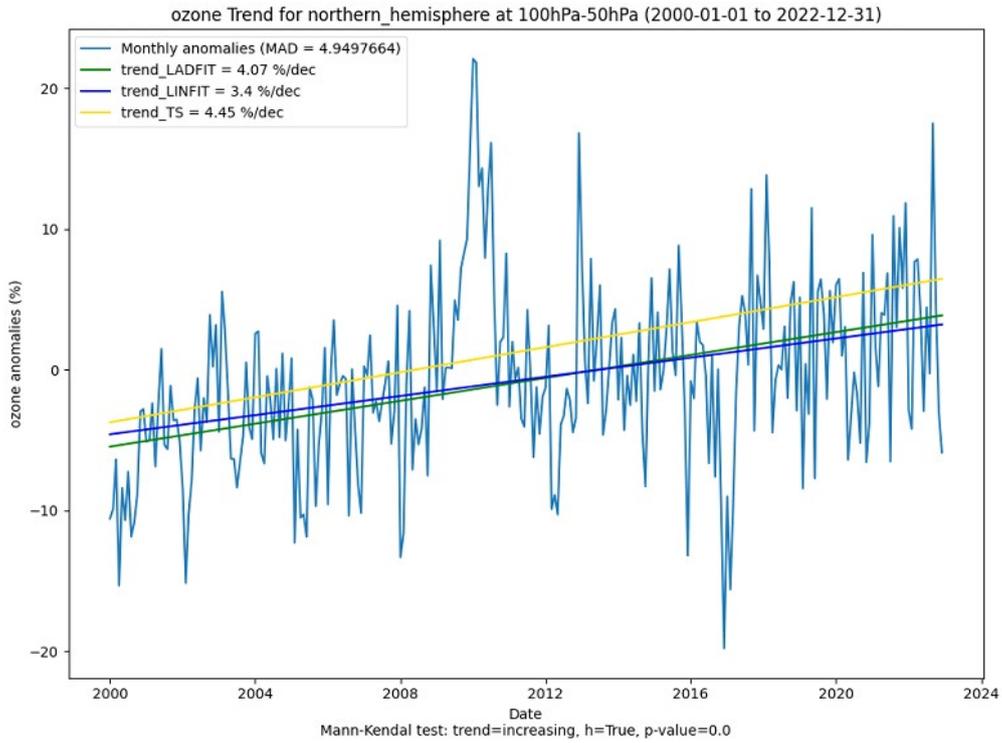
Figure 38. The same as Figure 35 but for the vertical range 100-50 hPa and for the post-2000 (2000-2022) period.

The ozone trends reported in Petropavlovskikh et al. (2019), which encompass the pre-2000 and post-2000 time series, can be compared with the patterns illustrated in Figures 35-38. The ozonesonde data show a negative trend of 5% per decade and more for the pre-2000 time series in the lower NH stratosphere, even reaching a negative trend of 10% at 100 hPa, which is very close to the trend percentage estimated in this work; for the post-2000 time series, the trend by Petropavlovskikh et al. (2019) is slightly negative, about 1%, in agreement with the trend estimated in the analysis presented here, and, at 100 hPa, a negative trend of 2%. This value has an uncertainty limit of ±7% and this makes the result calculated in this study (about 3-4% positive trend) within the uncertainty range.

Comparing the results with those from the post-2000 era provided by Sofieva et al. (2021) shows a 1-2% negative trend for the lower stratosphere, in agreement with what is shown with the unified dataset. The result shown by Sofieva et al. (2021) is estimated using a 15-year dataset from 2003 to 2018, which is compared with the trend estimate for the unified dataset from 2000 to 2022. Considering that the additional data in the unified dataset from 2000 to 2002 and from 2019 to 2022 are not significant for trend estimates due to their small size compared to those of the 2003-2018 time series, the comparison of trend estimates was considered acceptable.

Figures 35-38 also reveal that the difference between the trends calculated with the linear regressions and the LOTUS regression differ between 0.3 and 1 %/dec, therefore showing an agreement between the different types of regression. To calculate the trend in %/dec from the LOTUS regression, it was necessary to linearize the resulting regressor using the three linear techniques introduced previously. The three linearizations are in agreement for all vertical ranges and time series shown, differing within 0.8 %/dec.

4.4.2 Tropics

Figures 39-41 show the trends calculated for the TR sector for the 50-1 hPa and 100-50 hPa vertical ranges, distinguishing the pre-2000 and post-2000 time series. Unlike the NH, the estimated TR trend for the post-2000 period at the 100-50 hPa vertical range is not significant for the MK test, so it will not be reported here.
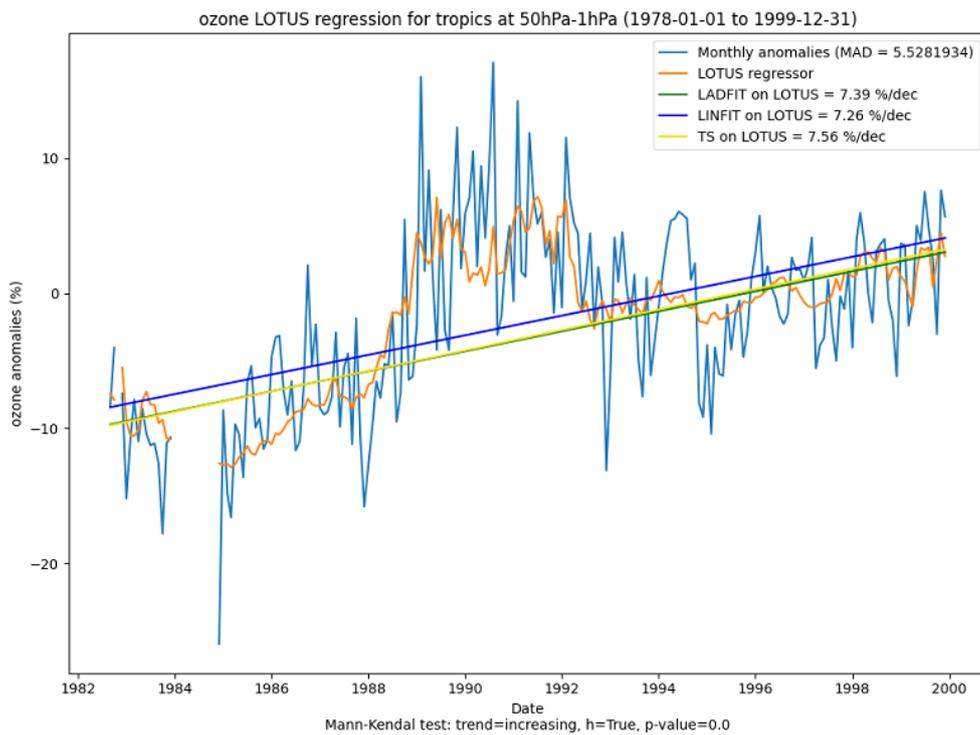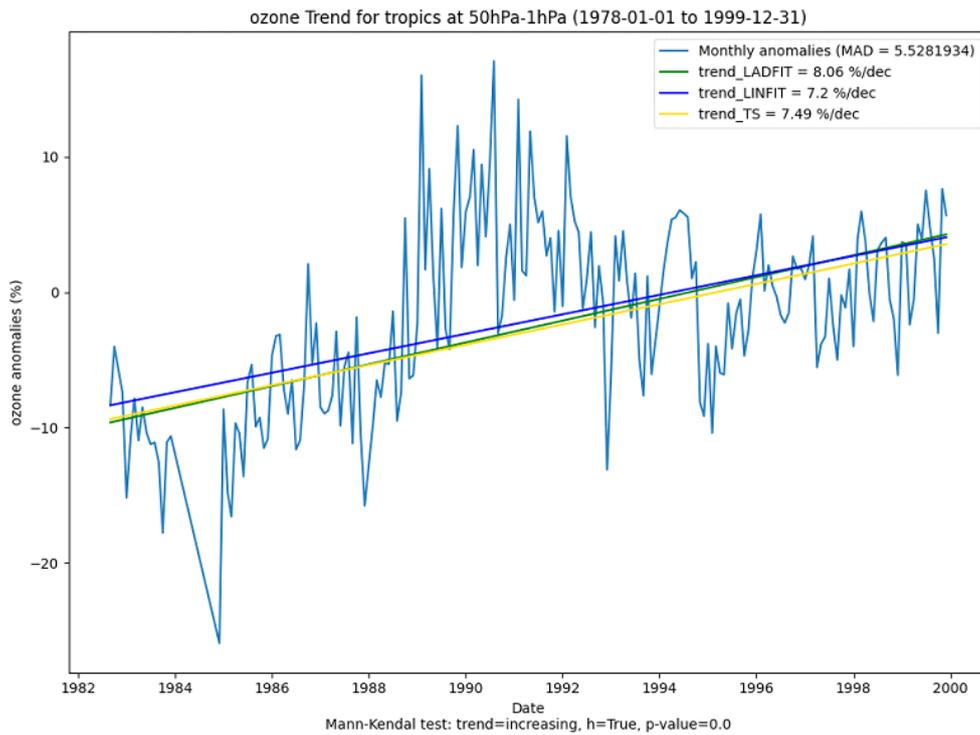
Figure 39. Trends estimations for TR latitudinal sector at 50-1 hPa vertical range for pre-2000 (1978-1999) period. On the top is the trend estimated with linear regression (LINFIT), least absolute deviation regression (LADFIT) and Theil-Sen regression (TS), and at the bottom is the trend estimated using the LOTUS regression, evaluated using LINFIT, LADFIT and TS linear regressors. The resulting percentage per decade (%/dec) for each regressor used and the MAD calculated on the average anomalies are shown in the figure legend. Finally, at the bottom centre is the result of the MK test.
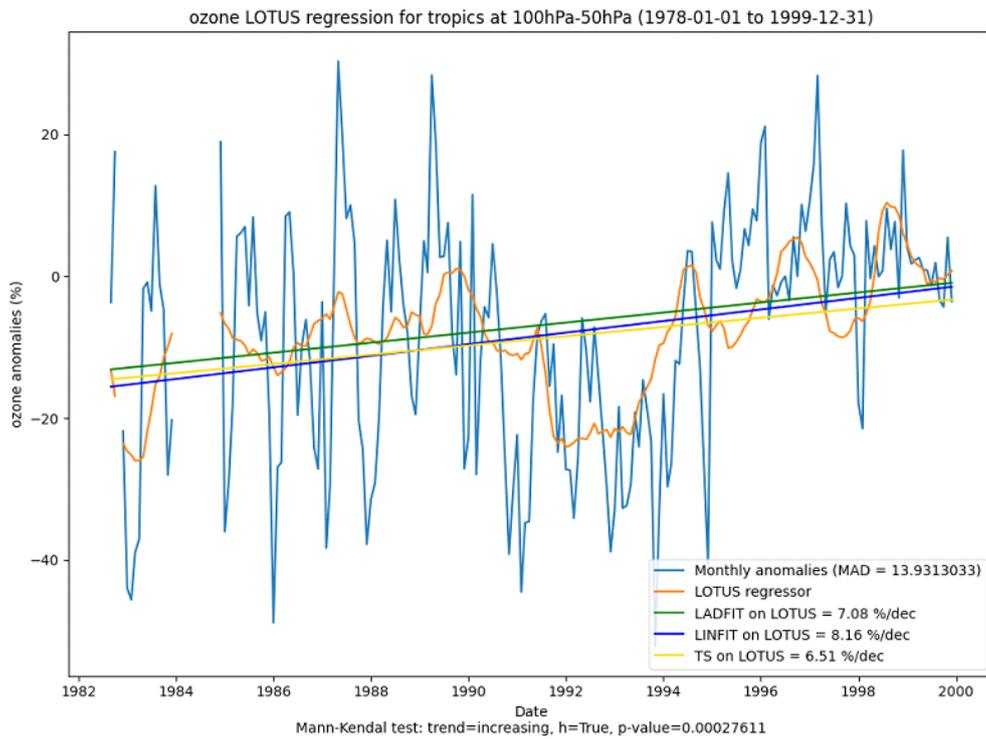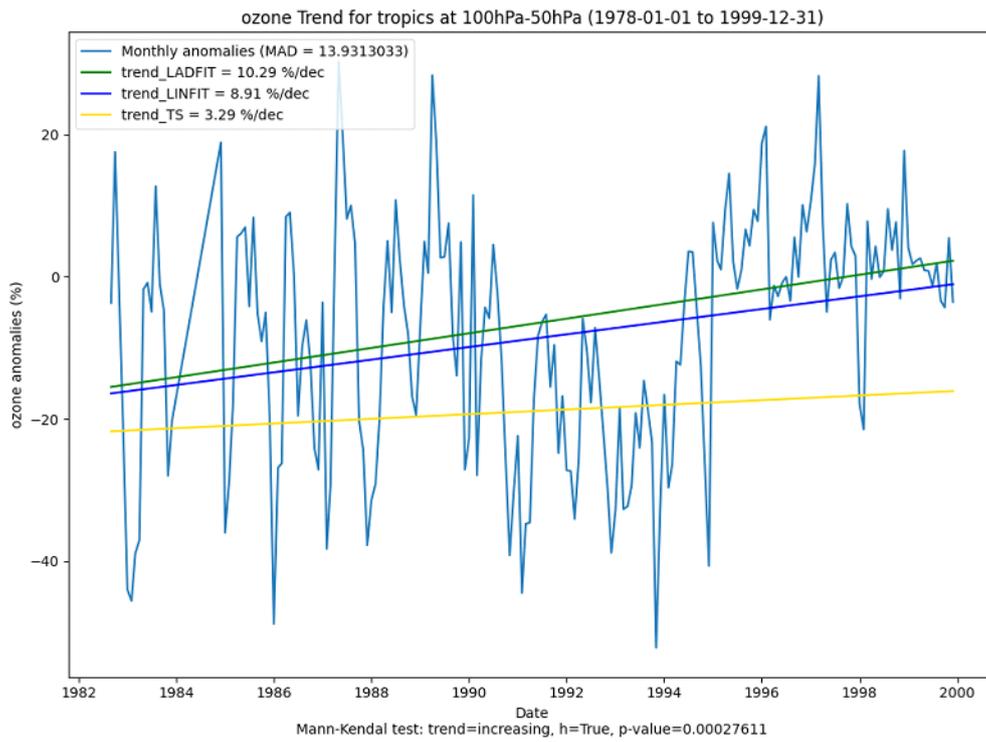
Figure 40. The same as Figure 39 but for the vertical range 100-50 hPa.
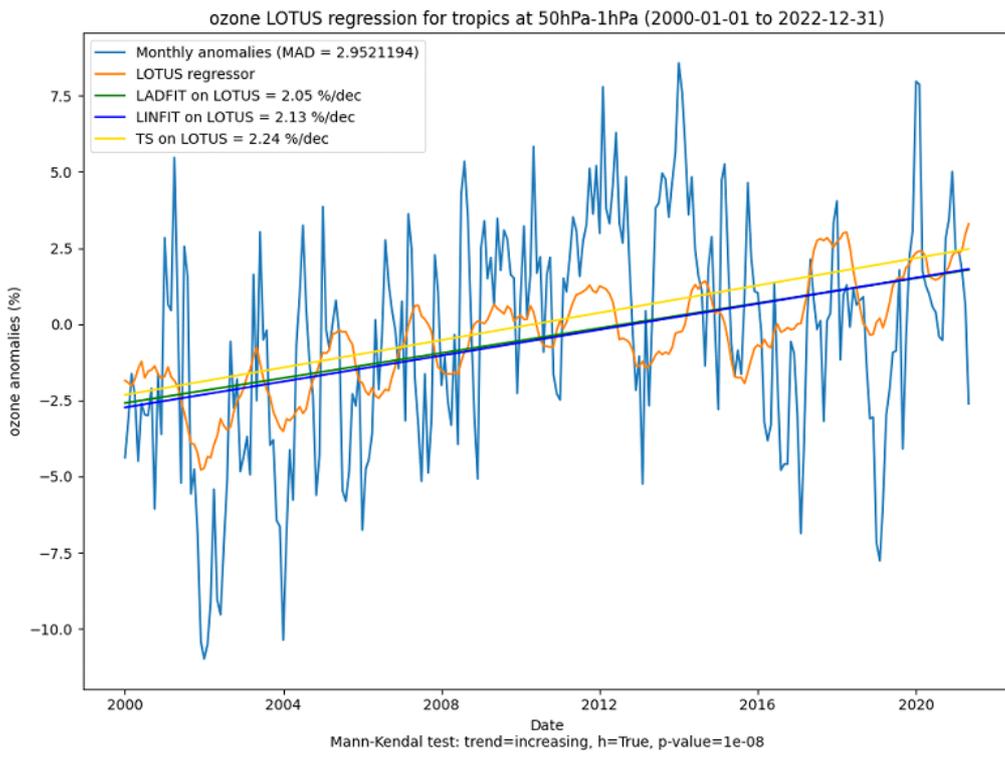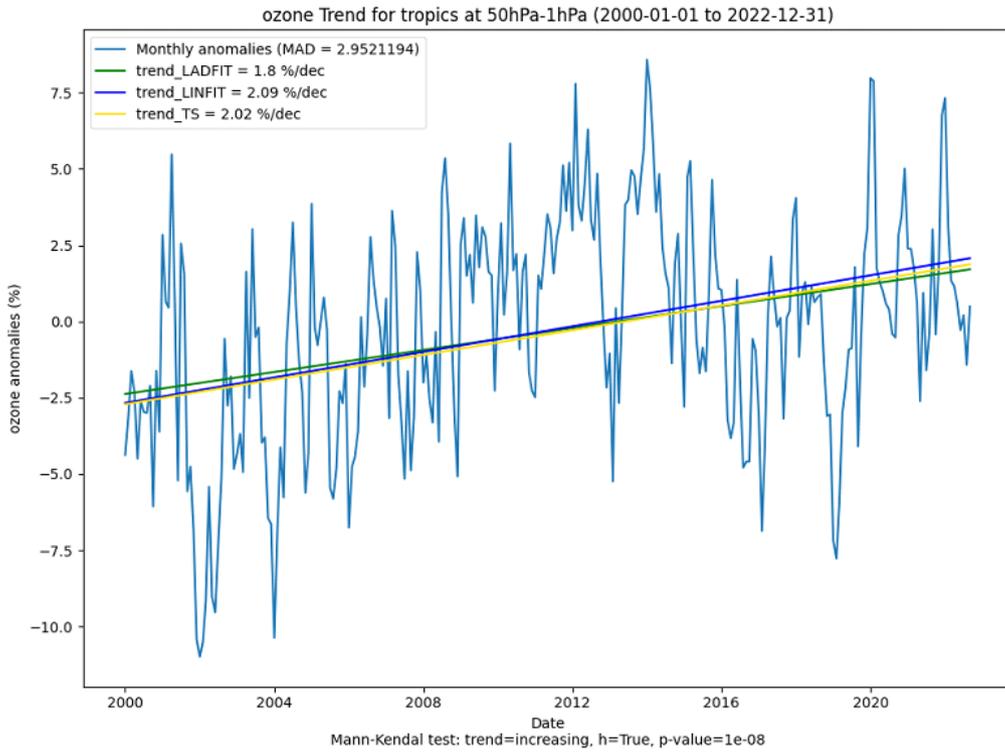
Figure 41. The same as Figure 39 but for the post-2000 (2000-2022) period.

Additionally, the patterns shown in Figures 39-41 for the TR sector may be contrasted with the ozone trends in Petropavlovskikh et al. (2019), for both the pre-2000 and post-2000 time periods. In contrast to the estimates discussed in this work, which shows a positive trend of about 8% in the lower stratosphere, reaching 10% at 100 hPa, the trends presented in Petropavlovskikh et al. (2019) indicate a negative trend of 2% per decade and more, reaching a negative trend of 10% at 100 hPa for time series pre-2000. The small sample size (just 4 stations) employed for the trend estimate from the unified dataset presented in this analysis is probably to blame for this disparity. However, concerning the post-2000 time series, a positive trend of 2% is seen in the lower TR stratosphere in Petropavlovskikh et al. (2019), as also demonstrated in this thesis, and a positive trend of 8% above 100 hPa (Petropavlovskikh et al., 2019). Similar to NH, this number has an uncertainty limit of 7%; nevertheless, as mentioned before, in this study, the MK test for the 100–50 hPa layer determined that the trend was not significant.

The lower stratosphere exhibits a positive trend of 2% for the post-2000 time series when compared to the data from Sofieva et al. (2021), confirming the estimated trend.

Differently from the NH, for the TR Figures 39-41 also show that the difference between the trends calculated with the linear regressions and the LOTUS regression is between 0.3 and 1 %/dec only for the vertical range 50-1 hPa (Figures 39 and 41), while for the vertical range 100-50 hPa (Figure 40) these differences exceed 3 %/dec for the LAD and TS regressors, while for the LIN it remains within 0.7 %/dec, therefore showing a better agreement. A possible cause of the greater differences compared to those shown for the NH is certainly the smaller amount of data available for estimating the trends. Finally, considering only the results obtained from the linearization of the LOTUS regressor, it is possible to say that they agree with all the vertical ranges and time series shown, differing between 0.2 and 1.6 %/dec.

Figure 41 reveals, after 2014, an En-Sci ozonesonde "dropoff" (Stauffer et al., 2022; Nakano and Morofuji, 2023), due to apparent anomalous losses of ozone in the lower and middle stratosphere, of several percentage points from the averages from 2014 to 2017 that could influence ozone trend calculations using tropical ozonesonde data. In this thesis, data from Izana (28.3°N, 16.48°W), Naha (26.21°N, 127.69°E), Hilo (19.43°N, 155.04°W) and Samoa (14.23°S, 170.56°W) stations were used to estimate ozone trends at post-2000 TR. As shown in Stauffer et al. (2022), not all tropical stations are affected by this sudden low bias; in fact, Izana and Naha do not suffer a "dropoff", unlike Hilo and Samoa, which are affected. Finally, considering that the trends estimated in this thesis follow those shown in Petropavlovskikh et al., (2019) and Sofieva et al., (2021), it is inferred that the influence of the "dropoff" is not significant for trend estimation at TR.

5. Upper-air trends of temperature and humidity

Observational data records are influenced by instrumental effects which may erode their reliability for climate applications (Kremser et al., 2018). Among in-situ measurement techniques, radiosounding balloons are widely used for the study of climate changes in the atmosphere. Several climate studies used radiosounding time series homogenized at mandatory pressure levels, based on a range of algorithms using reference time series (mainly atmospheric reanalysis data), information from the nearest neighbouring stations or kriging techniques (e.g., Free et al. 2004; Thorne et al., 2005; McCarthy et al., 2008; Sherwood et al. 2008; Dai et al., 2011; Haimberger et al., 2012). Some of these methods include ensembles to quantify parametric uncertainties (Haimberger et al., 2012; Thorne et al., 2011; Madonna et al..2022, Zhou et al., 2021). Only very few algorithms provide a quantification of measurement uncertainties.

Homogenization algorithms aim at detecting and adjusting for quantifiable systematic inhomogeneities in the observational time series, which for radiosoundings strongly depend on the sensor type and altitude level and vary substantially through time (e.g., changes in station location, instrumentation, calibration or drift issues, different instrument sensitivity concerning different networks, changes in the measurement procedures, etc.). Metadata can support the design or the validation of the algorithms, although for historical in-situ observations, the metadata availability is poor and known to be incomplete (Durre et al., 2006; Ferreira et al., 2019).

In this work, the novel algorithm named RHARM (Radiosounding HARMonization) for the homogenization of historical radiosounding data records available since 1978 was considered. The RHARM algorithm, designed and implemented under the Copernicus Climate Change Service (C3S), is based on a novel hybrid approach, combining physical (metadata-based) and statistical adjustments of radiosonde instrumental effects. RHARM is the first algorithm to provide in a single package homogenized time series of temperature, relative humidity and wind profiles alongside an estimation of the observational uncertainty for each observation and pressure level. Physical adjustments for the most recent radiosonde measurements (generally post-2004) are calculated following the data processing of radiosounding data adopted by GRUAN (Global Reference Upper-Air Network, Bodeker et al., 2016) and inferring information from the WMO 2010 radiosonde intercomparison (Nash et al., 2011). Historical observations before 2004 are homogenized using a statistical methodology allowing quantification of the measurement uncertainties. A detailed description of the RHARM approach is provided in Madonna et al. (2022).

5.1 Data sources and methodology

An important step in the performance assessment of the RHARM data is the comparison with independent datasets. The data records considered do not include any truly "Reference" datasets (Thorne et al., 2017) and therefore none of the compared datasets can be assumed as the truth. Hence these comparisons can elucidate only relative rather than absolute characteristics of the RHARM product.

The RHARM algorithm for the homogenization of global radiosounding temperature, humidity and wind profiles is applied to per-ascent (generally 00:00 and/or 12:00 UTC) radiosonde data on 16 mandatory pressure levels (10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, 1000 hPa). Relative humidity (RH) adjustments are limited to 250 hPa owing to pervasive sensor performance issues at higher altitudes. Profiles are adjusted at these mandatory levels. The applied adjustments are then interpolated to significant levels. Uncertainties are estimated for each processing step and then propagated to estimate the total uncertainty.

5.1.1 Data Sources Used

The RHARM approach is applied to the IGRA database which is the most comprehensive global collection of original "raw" historical and near-real-time radiosonde and pilot balloon observations. RHARM is applied to IGRA Version 2 (Durre et al., 2018) which incorporates data from a considerably greater number of data sources with an increased data volume by 30% compared to Version 1. RHARM is applied to a subset of 700 radiosounding stations and radiosoundings from ships. Only the records with documented metadata (i.e., including the radiosonde code according to WMO table 3685, describing the radiosonde type) since 2000 (for most of the stations) and for fewer stations since 1978 were selected. For these stations, depending on the radiosonde type, adjustments based on the application of GRUAN-like data processing or the comparison between GRUAN data and ID2010 can be applied to the post-2004 period, for which several instrumental effects are already corrected (e.g., the well-known solar radiation dry-bias).

The IGRA data v2 are the result of improved quality assurance procedures developed for the IGRA data v1 (Durre et al. 2006; Durre et al. 2008), which can be grouped into eight categories: fundamental "sanity" checks, checks on the plausibility and temporal consistency of surface elevation, internal consistency checks, checks for the repetition of values, checks for gross position errors in ship tracks, climatology-based checks, checks on the vertical and temporal consistency of temperature, and data completeness checks. The RHARM dataset thus inherits the IGRA quality assurance procedures, and additional quality checks are then applied. Tests were performed on metadata availability, physical plausibility, data completeness check, bias adjustment

accuracy, outlier removal, vertical correlation between structural breaks at the same station, and coherency check for the adjustments applied at significant levels.

As noted, the RHARM approach is applied to a subset of IGRA records, depending on the availability of the required metadata (Durre et al. 2008; Ferreira et al., 2019). For these stations, a quality-enhanced dataset with a sufficient number of radiosoundings available from 2004 to the present can be provided directly post-processing the profiles to account for several instrumental effects (e.g., the well-known solar radiation dry-bias). The post-processed profiles are then used as reference information to adjust the systematic effects in the historical data before 2004. For those stations where the number of post-processed radiosounding profiles is not sufficient for the homogenization algorithm before 2004, the post-processed profiles since 2004 are provided only. For the selected 700 IGRA stations, only measurements with the highest data quality according to the IGRA data quality system at each pressure level have been processed with the RHARM algorithm.

Figure 42 shows the locations of the stations processed herein and several launches available. In addition, the 1156 IGRA stations reporting data from 1978 to the present are also shown. The coverage of RHARM is reasonably homogeneous, except for Siberia where only a smaller number of launches is available. This is due to the limited information available on the main radiosonde type used in the region since 2004 (AVZ), which cannot be adjusted using RHARM to achieve the same quality as for the remaining radiosonde types. The station density in Canada, North East Asia, and East Africa is lower than in Europe, the U.S. and South America, but this is common to all datasets and reflects the inadequacies of the historical observing system. Table 18 confirms the low number of measurements available in the SH, although the number of measurements alone cannot address the value of the dataset for a specific study without considering representativeness (Weatherhead et al., 2017).
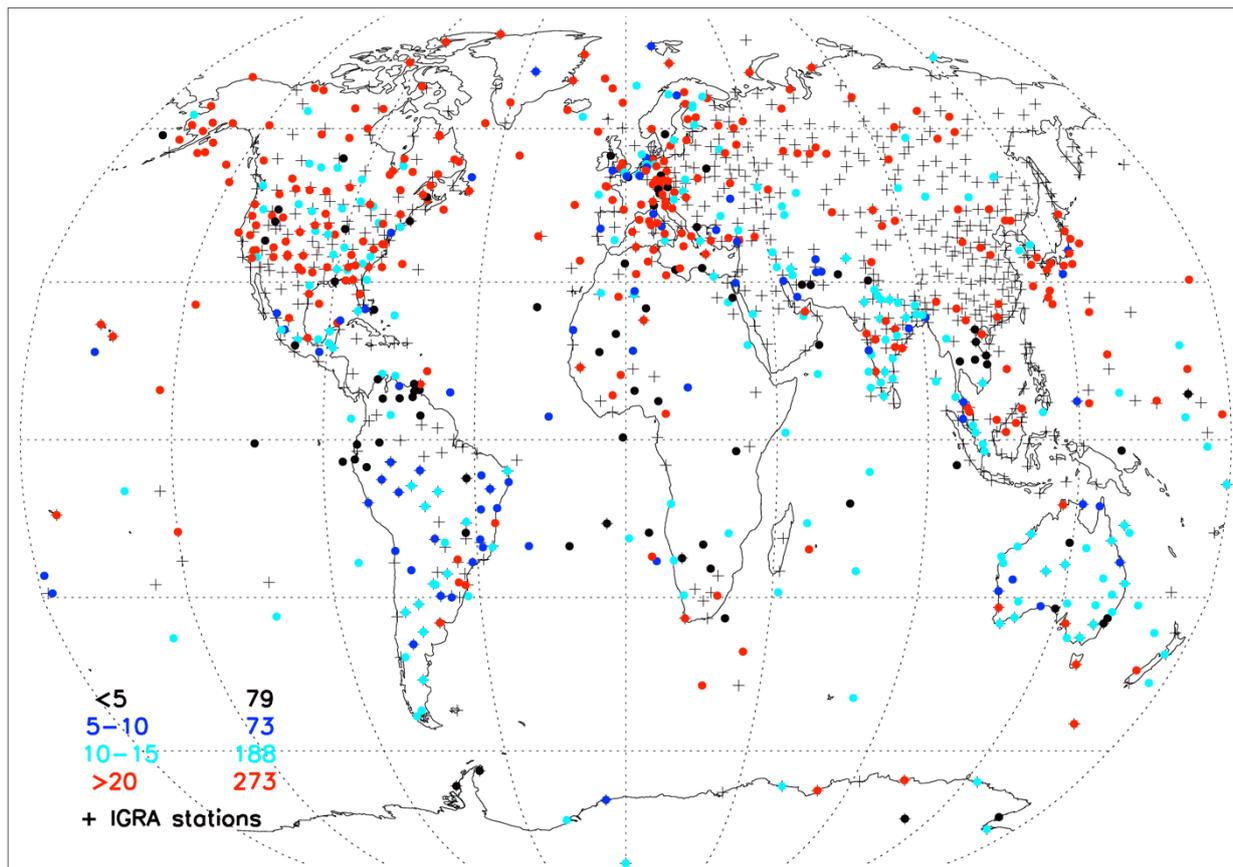
Figure 42. Global distribution and quantity of RHARM homogenized profiles. The scale in the bottom left corner denotes the available radiosoundings at each station (in millions of ascents) from 01-01-1978 to the present. The + symbol indicates the additional IGRA stations (1156) reporting data from 1978 to the present (last access to IGRA 31-12-2020).

Region	Latitude range	Number of launches (thousands)	Percentage
Arctic	70°N-90°N	316.1	2.5
Northern Hemisphere mid-latitudes	25°N-70°N	8203.7	65.4
Tropics	25°N-25°S	2979.3	23.8
Southern Hemisphere mid-latitudes	25°S-70°S	974.0	7.8
Antarctic	70°S-90°S	64.2	0.5
Total		**12537.3**	**100**

Table 18. The number and percentage of launches in different latitude bands for the stations shown in Figure 42.

5.1.2 Methodology

The RHARM homogenization of global radiosounding temperature, humidity and wind profiles is applied to per-ascent (generally 00:00 and/or 12:00 UTC) radiosonde data on 16 mandatory pressure levels (10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, 1000 hPa), because the frequency of reports from the stations is typically per ascent whereas significant level reports vary by definition per profile. Relative humidity (RH) adjustments are limited to 250 hPa owing to pervasive sensor performance issues at greater altitudes in almost all commercial sondes (Miloshevic et al., 2004). Profiles are adjusted at these mandatory pressure levels. The applied adjustments are then interpolated to the reported significant levels. Uncertainties are estimated for each processing step and propagated to estimate the total uncertainty.

For the sake of clarity, the RHARM-adjusted time series since 2004 (with starting time station-dependent) obtained by post-processing each single radiosounding profile using a GRUAN-like algorithm is labelled Post-Processed Time Series (PPTS). The PPTS is then used as a constraint for adjusting the preceding radiosounding time series, hereinafter Homogenized Time Series (HST). The concatenation of HTS and PPTS records provides the entire time series for each station, and only those stations satisfying the requirements for the production of a PPTS are considered for the HTS calculations. An overall scheme of the RHARM approach is shown in Figure 43.

The PPTS produced as step A1 for each station is merged with the prior period of record (step A2). The resulting time series (step B) is firstly divided into two sub-series to separate the nominal 00 UTC and 12 UTC launches, which are the two most frequent launch times in IGRA. Local nighttime and daytime conditions for each radiosounding launch are identified by calculating the solar zenith angle using the LOWTRAN module (available at http://ethangutmann.com/pages/idl/Utilities/zensun.pro), using as inputs each radiosonde launching time and the corresponding station geographical coordinates. The small number of radiosondes launches available at other synoptic hours has not been considered in the current RHARM data version. Such a step is critical mainly for temperature and humidity where radiation-heating effects can have substantive impacts on instrument performance (Miloshevic et al., 2004; Wang et al., 2013; Dirksen et al., 2014).
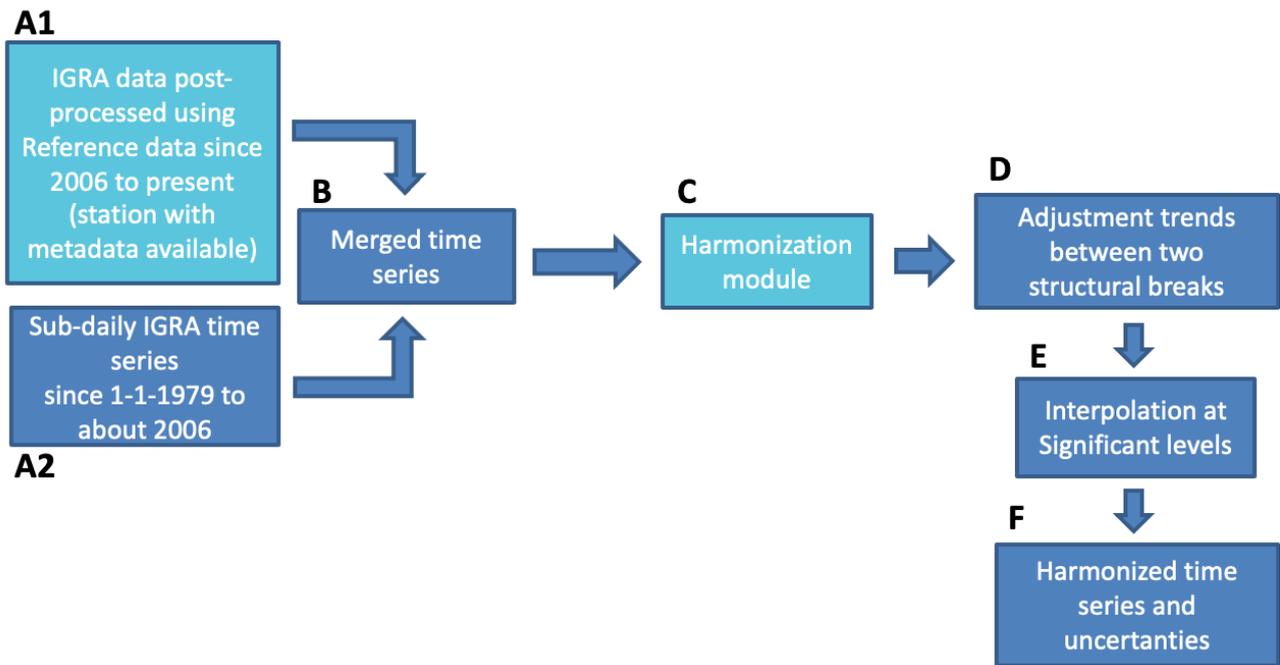
Figure 43. A schematic diagram describing the steps of the RHARM approach.

5.1.3 ERA5

The ECMWF ERA5 reanalysis (Hersbach et al., 2020) incorporates millions of observations into a data assimilation system, every 6-12 hours over the period being analysed, providing a systematic approach to produce a dataset for climate monitoring and research. ERA5 is the latest climate reanalysis produced by ECMWF providing hourly data on regular latitude-longitude grids at 0.25° x 0.25° resolution and on 37 pressure levels. ERA5 is publicly available through the Copernicus Climate Data Store (CDS, https://cds.climate.copernicus.eu). For the validation of uncertainties, we also use the ERA5 background (6-hour forecast) as a reference value. The various reanalysis products have proven to be valuable when used appropriately (Dee et al., 2011). Nevertheless, reanalysis reliability can considerably vary depending on the location, period, and variable considered (Dee and Fasullo, 2016). The changing mix of observations, and biases in observations and models, can introduce spurious variability and trends into reanalysis output (Dee and Fasullo, 2016). However, the ERA5 data is certainly one of the most widely used datasets for climate applications, with a growing popularity.

5.1.4 Satellite data

For water vapour version 4 of the data from the Microwave Limb Sounder (MLS) onboard the Aura satellite (Yan et al., 2016) has been used. For the comparison shown in this work with RHARM data, only the MLS data at 316 hPa RH time series were considered and these have been subsampled at the IGRA stations using the nearest grid points and compared to 300 hPa standard pressure level reports.

5.1.5 Other existing homogenized datasets

The RHARM approach increases the limited number of existing homogenized datasets, which include:

- homogenized radiosounding temperature measurements: Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC) by NOAA (Free et al., 2004), RAdiosonde OBservation COrrection using REanalyses (RAOBCORE), Radiosonde Innovation Composite Homogenization (RICH) by the University of Wien (Haimberger et al., 2012), Hadley Centre's radiosonde temperature product v2 (HadAT2) by Met Office (Thorne et al., 2005), Iterative Universal Kriging v2 (IUKv2) by University of New South Wales (Sherwood and Nishant et al., 2015), the State University of New York Albany dataset (Zhou et al., 2021);
- homogenized radiosounding humidity measurements: the Homogenized RS92 radiosounding humidity measurements (HomoRS92) by the State University of New York Albany (Dai et al., 2011) and the Hadley Centre's radiosonde temperature and humidity product (HadTH) (McCarthy et al., 2009); and
- homogenized radiosounding wind datasets: IUKv2 and GRASPA (Ramella-Pralungo et al., 2014 a,b).

5.2 Comparisons with ERA5

In this section, the tropospheric interannual variations for the period 1979-2018 depicted in the radiosonde datasets were compared with those from the ERA5 reanalysis. Monthly anomalies of temperature and RH from IGRA and RHARM are compared with the results obtained for ERA5, having selected the nearest grid point to each station. Simultaneous vertical profiles on 00 UTC and 12 UTC at 300 hPa and 500 hPa pressure levels are selected. Considering the high spatial resolution of ERA5 and its representativeness, the uncertainty due to the use of the nearest grid-point interpolation should be comparable with other methods (e.g., kriging, bilinear interpolation, etc.).

For temperature (Figures 44-45) in the NH, IGRA, RHARM and ERA5 show a similar positive decadal trend of 0.38, 0.39 and 0.43 K/decade, respectively, while in the TR at 300 hPa, the trend is of 0.17, 0.25, 0.20 K/decade, with a more pronounced trend increase starting around 1997. Similar results have been obtained considering European stations only (Madonna, 2020c). In the NH, the comparison of the anomalies at 300 hPa shows the evident adjustment applied by RHARM on the IGRA data over 1996-2005 (corresponding to the period of the RS90 radiosondes usage) which reduces the difference from ERA5 results. Differences between RHARM and ERA5 are generally smaller than 0.5 K in absolute value in both regions, with larger differences for the most extreme values especially in the TR. At 500 hPa, the situation is very similar although in the NH the differences are smaller, while in the TR results are in line with those for the 300 hPa pressure level.

For relative humidity (Figures 46-47), in the NH the substantive adjustment applied to IGRA by RHARM at 300 hPa before 1986 (up to 10 %RH) largely improves the agreement with ERA5. In 1986, a few major changes occurred in the global radiosounding data, the most relevant of which are: changes in several radiosonde models, such as MARS/MRZ and VIZ radiosondes; the adoption of new manufacturers at some stations, mainly changes from another manufacturer to Vaisala, and changes in the dewpoint depression algorithm, for example at U.K. stations; and, maybe the most important, the introduction of "pre-baselined" radiosondes, i.e. removal of the practice of applying a manual baseline lock for all temperature and RH profiles which was discovered to be prone to producing a wet bias for RH lower than 60% (more details at https://library.wmo.int/doc_num.php?explnum_id=9592). Since 2004 the adjustment applied by RHARM is smaller and further improves the agreement with ERA5 with a negative trend of -0.8 % RH/da. In the last decade, the trends show a change with a slight increase which has been already quantified in the European domain (Madonna, 2020c). At 500 hPa, the situation is very similar although the adjustments are much smaller. In the TR, the adjustments applied by RHARM at 300 and 500 hPa are smaller than in the NH. The comparison with ERA5 shows that the largest differences are at 500 hPa (up to 4-5 %RH). The comparison highlights major differences in three periods: before 1990, where ERA5 negative anomalies are smaller; after 2005, with RHARM anomalies larger than those of ERA5; and after 2015, when differences are larger, especially at 500 hPa. The comparison in these three periods generally reflects the higher variability of the observational times

series compared to ERA5 data. Large positive humidity anomalies observed in the TR for the period 2015-2019 are correlated with significant positive anomalies of the bi-monthly multivariate El Niño/Southern Oscillation (ENSO) index (Hu and Fedorov, 2017, available at https://www.esrl.noaa.gov/psd/enso/mei), with a warm event started in January 2015 and peaked within the same year. Boosted by this major El Niño, 2015 was the first of five consecutive years among the six warmest years in the 140-year observational record (e.g., https://www.ncdc.noaa.gov/sotc/global), which may be related to the observed strong positive anomalies of RH in the TR and the SH. A possible positive trend in upper-tropospheric absolute humidity has been noted in previous works (e.g., Dessler and Davis, 2010).

In the SH, where only 66 stations are available in RHARM, the comparison (not shown) results are similar to the TR, with the same strong positive humidity anomalies after 2015.

Despite a degree of non-independence, the comparison with ERA5 reveals discrepancies in the monthly anomalies and trends with both IGRA and RHARM, although the adjustments applied in the RHARM data allow to reduce the difference between ERA5 and the observations, especially for temperature and RH in the NH. ERA5 performances in reproducing the observed atmospheric variability appear to be higher in the NH than in the TR, likely due to the stronger observational constraints. Considering the difference in the RHARM homogenization approach for the data before and after 2004 (based on a statistical method and a GRUAN-like post-processing, respectively), the high-frequency variability is very similar for IGRA and RHARM: this is expected by construction for the data before 2004 (Figure 49), where only the bias in between each pair of breakpoints is adjusted, while for data after 2004 (Figure 48) the adjustments, although affecting the overall underlying data distribution, are typically of small magnitude due to the enhanced quality of recent radiosonde data compared to historical observations (Thorne et al., 2011).
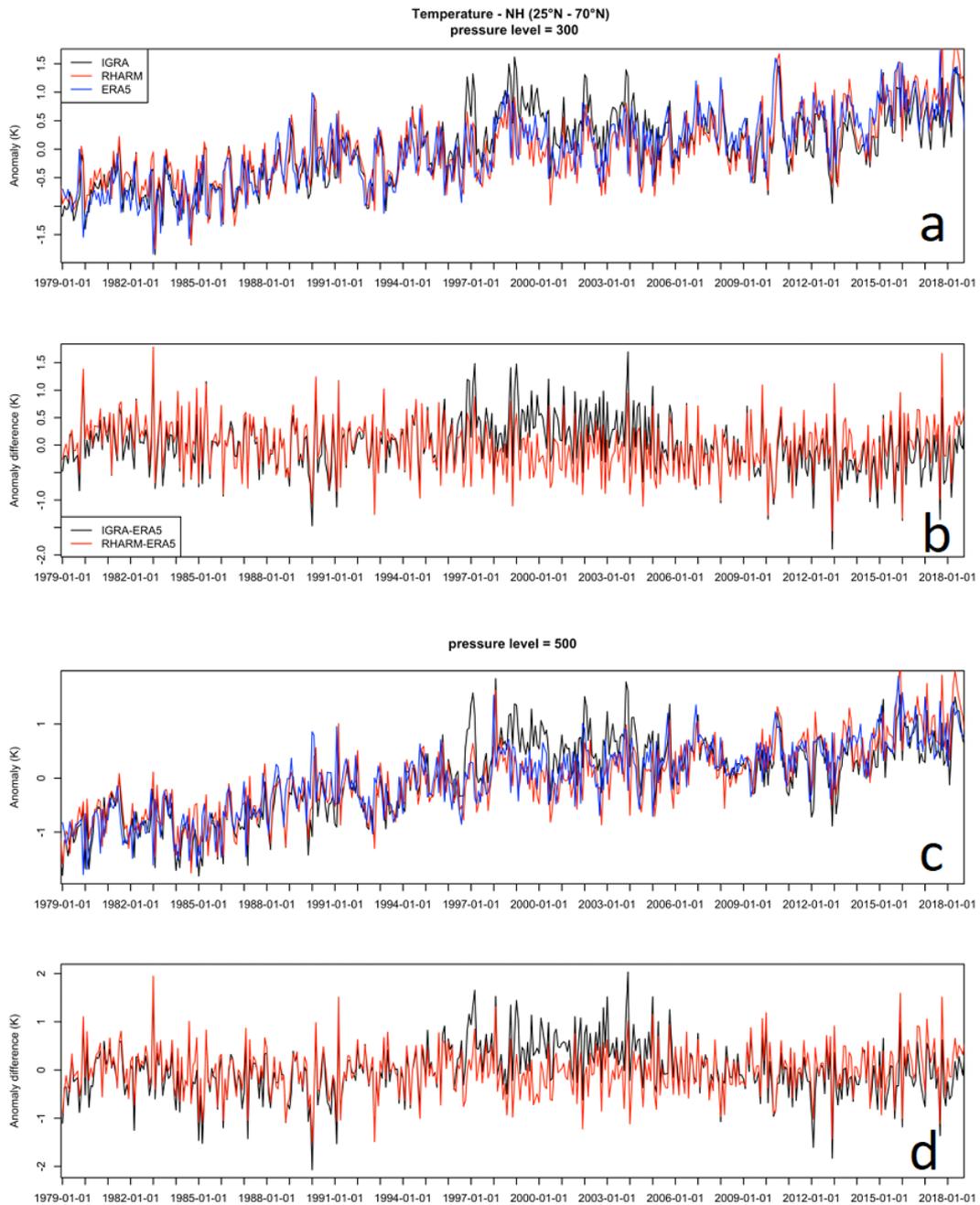
Figure 44. Monthly temperature anomalies (i.e., deviation from the mean created by subtracting climatological monthly means) and anomaly differences calculated between 01/01/2006 and 01/08/2019 for IGRA (black), RHARM (red) and ERA5 reanalysis (blue) at 300 and 500 hPa for NH. Anomalies are shown in panels a and c, while anomaly differences in panels b and d.
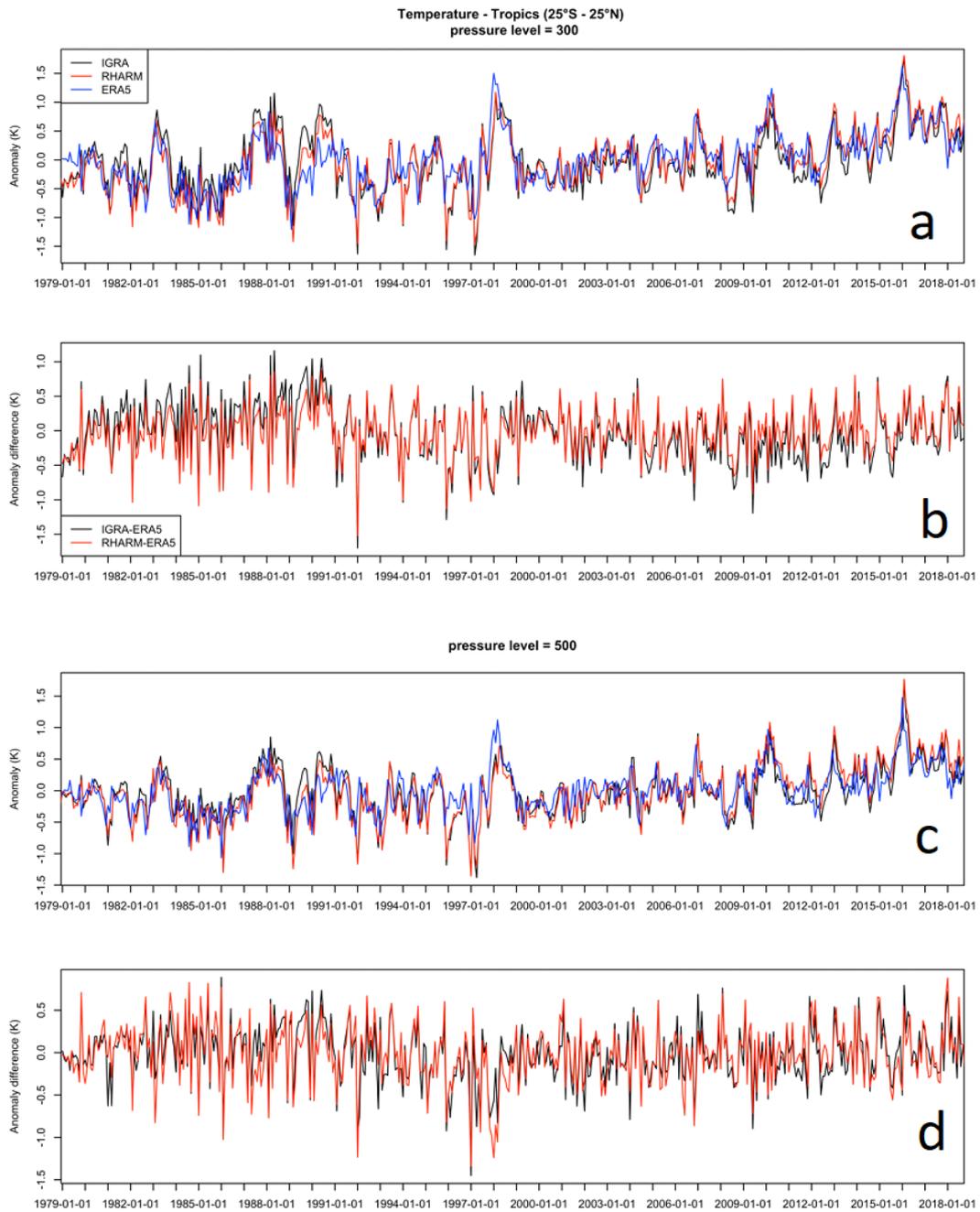
Figure 45. Same as Figure 44 but for the Tropics.
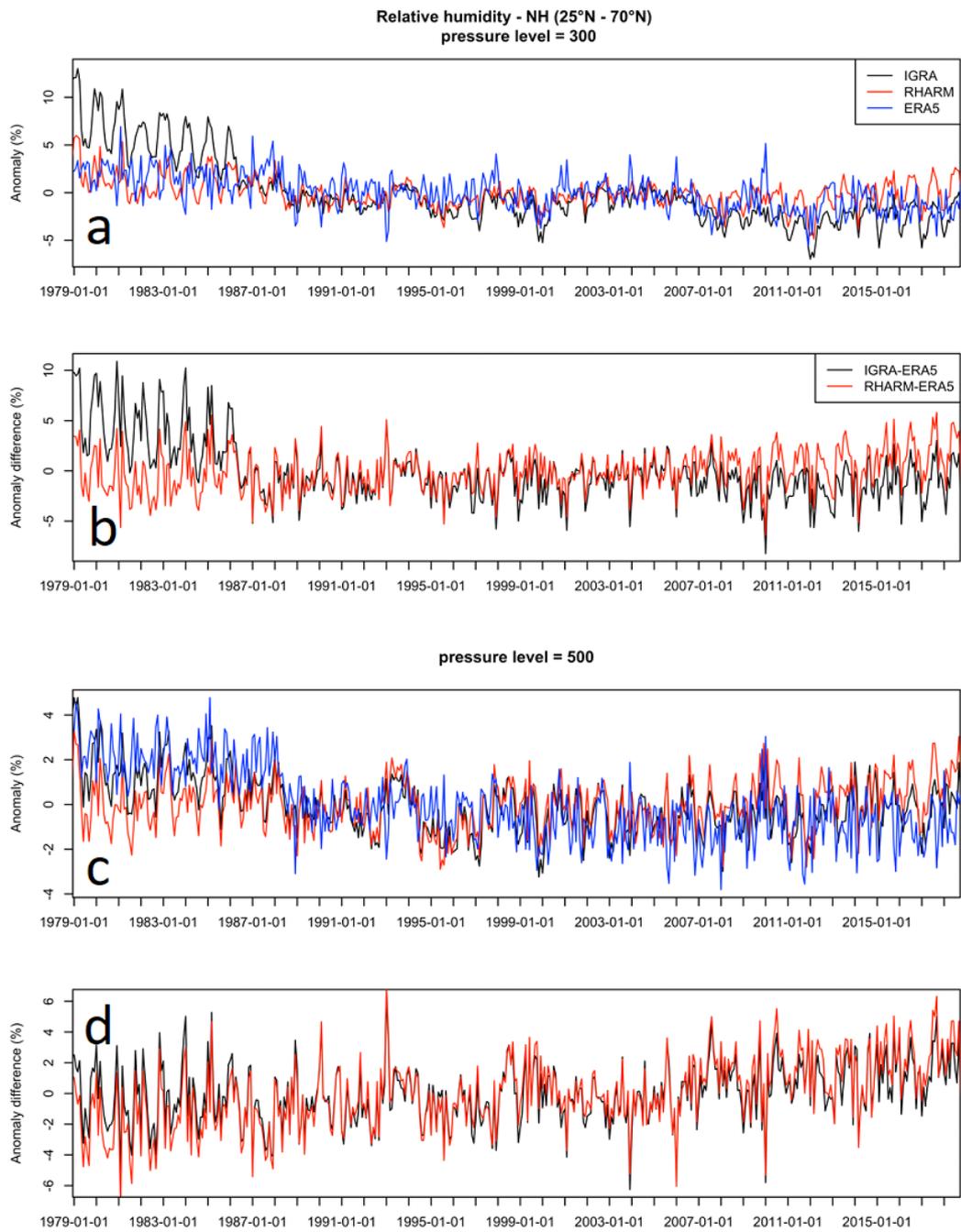
Figure 46. Same as Figure 44 but for relative humidity.
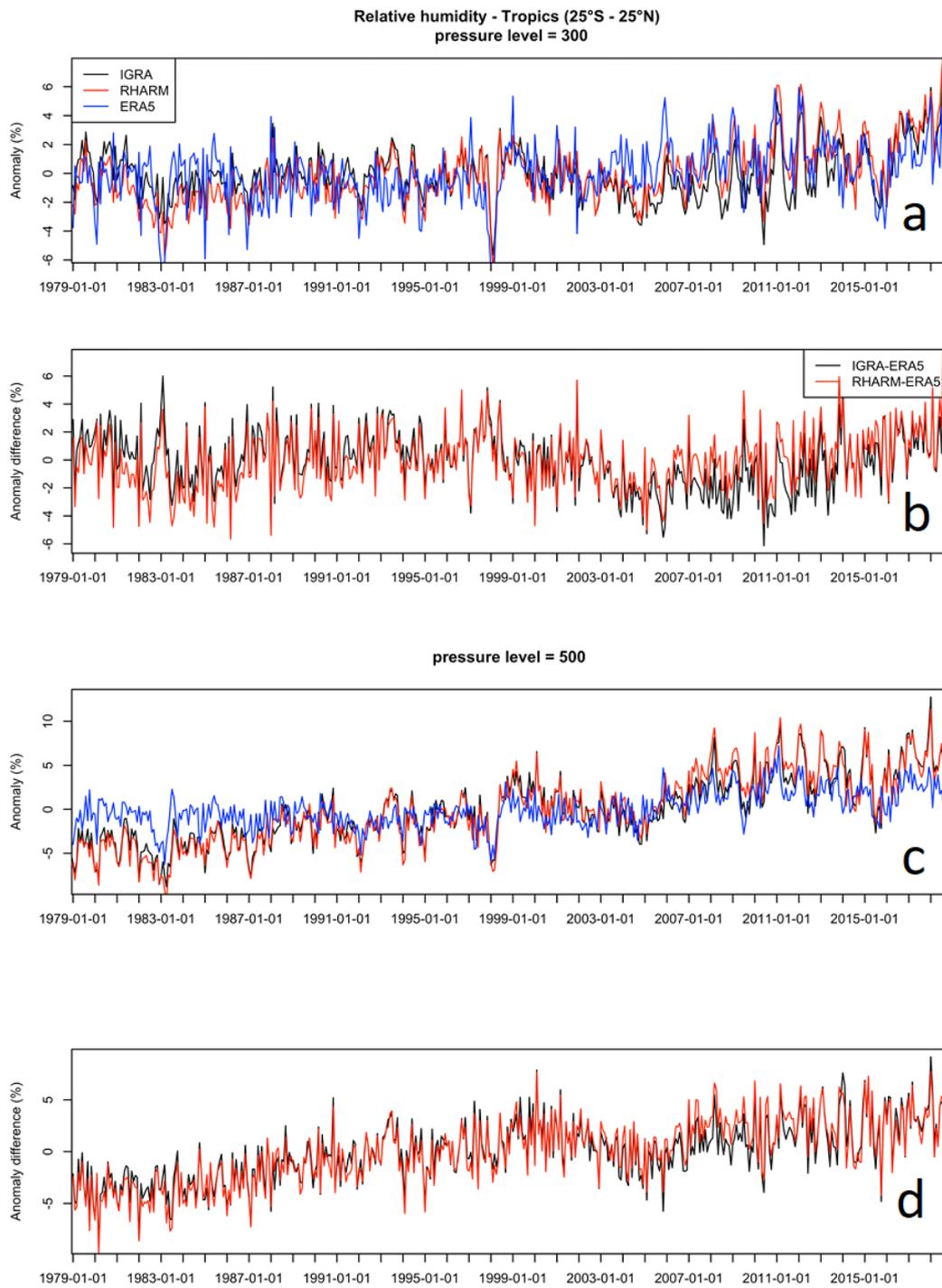
Figure 47. Same as Figure 45 but for relative humidity.
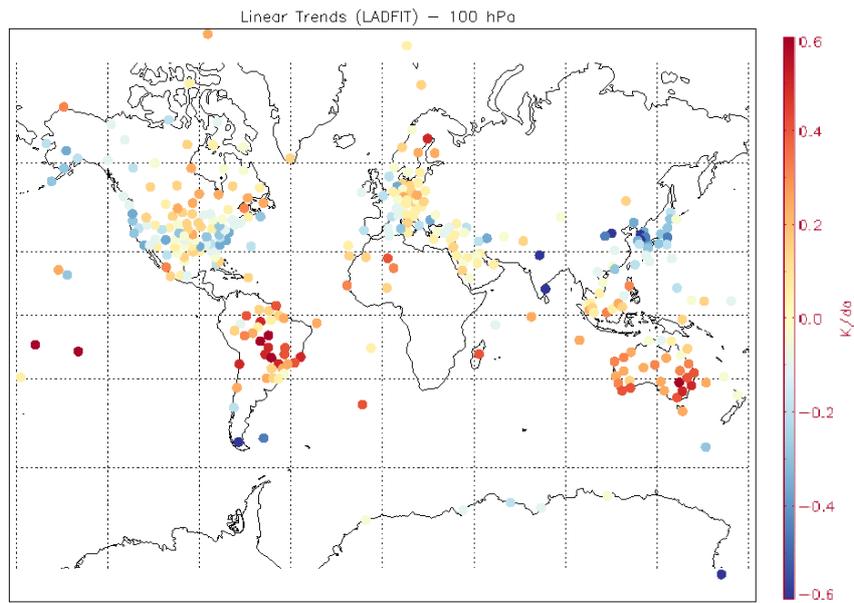
Figure 48. Global map of trends (per decade) at 100 hPa of temperature estimated by RHARM stations in the period 2004-2022.

Figure 48 provides the trends estimated from RHARM at 100 hPa for the period 2004-2022, considering only the stations post-processed by RHARM with more than 3000 ascents since 2004. Figure 49 provides the same as Figure 48 for the period 1978-2022.
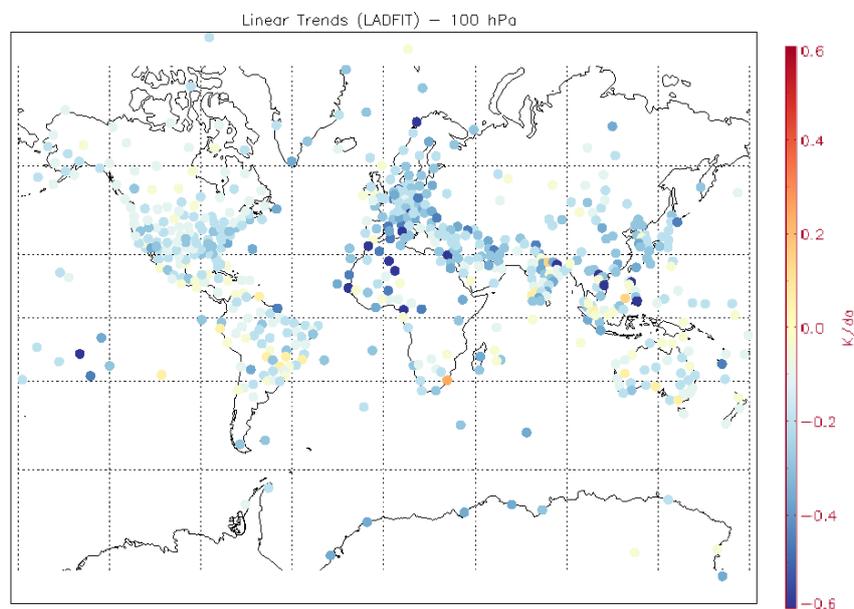
Figure 49. Same as Figure 48 but for the period 1978-2022.

5.3 Comparisons with existing homogenized datasets

5.3.1 Temperature

In the NH, the comparison with other radiosonde homogenized datasets for temperature shows a very good agreement overall at pressure levels up to 200 hPa (Figure 50, left panel), with relative differences within 0.1 K/decade, except for IUKv2 which oscillates between -0.5 and 1.5 K/decade for reasons that must be further investigated. Above 200 hPa, all the datasets are in agreement within 0.1 K/decade. Between 200 hPa and 70 hPa, RAOBCORE and RICH are the closest to ERA5, while above 70 hPa the ERA5 values overlap with IGRA. In the same altitude range RAOBCORE, RICH and RHARM are highly similar. IUKv2 data are limited to 30 hPa and this is the only observational gridded dataset, with a consequent additional uncertainty due to the interpolation to the nearest grid point. In the TR (25°S-25°N), the shape of the trend vertical profiles is similar for all datasets (Figure 50, right panel), with IGRA the coldest and IUKv2 the warmest at all pressure levels. Up to 300 hPa, trends are positive (tropospheric warming) and their difference does not exceed 0.25 K/decade. In the range 300 hPa – 70 hPa cooling trends for RHARM, ERA5, RAOBCORE and RICH are very close, within 0.1 K/decade, while above this level ERA5 are closer to RAOBCORE and RHARM are closer to RICH, respectively. The IUKv2 is warmer than any other dataset above 300 hPa.
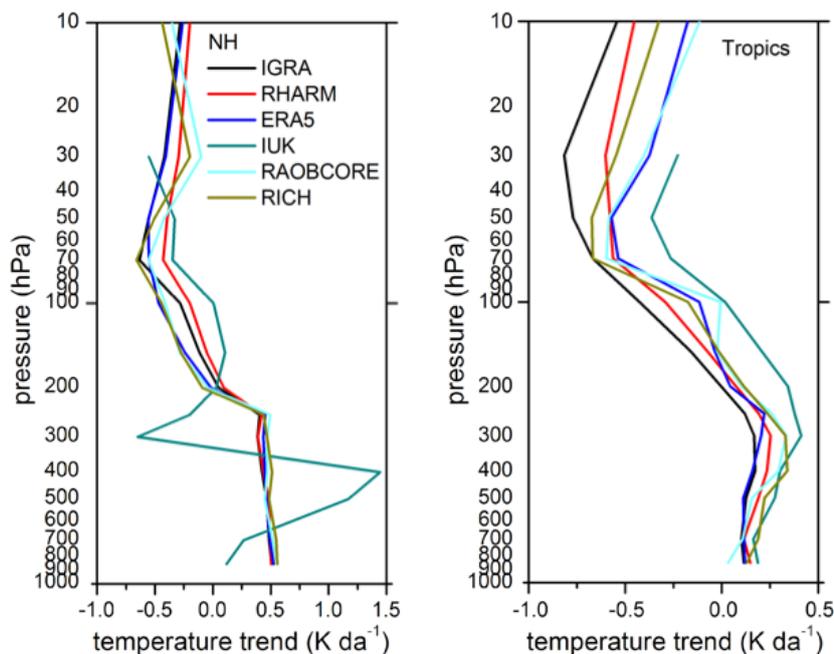

Figure 50. Profiles of temperature decadal trends at mandatory pressure levels between 850 and 10 hPa for the period 1979-2018, in the NH (left panel) and in the TR (right panel) for the unadjusted IGRA (black line), RHARM (red), ERA5 (blue), IUKv2 (green), RAOBCORE (cyan), RICH (dark yellow) datasets. The ordinate is logarithmic, and the abscissa differs between the two panels.

5.3.2 Relative humidity

In the literature, two examples of homogenized datasets of relative humidity radiosounding profiles are available: HadTH (McCarthy et al., 2009) and the HOMORS92 dataset (Dai et al. 2011). HadTH provides a time series of anomalies relative to the monthly 1981-2000 climatology at standard pressure levels up to 300 hPa for stations in the NH only. As stated in McCarthy et al. (2009), only in the NH there is sufficient data for the application of the HadTH homogenisation method. Considering that the HOMORS92 dataset has a longer time coverage and is not limited to the NH, this is chosen to perform a comparison with RHARM.

Relative humidity (RH) is adjusted directly in RHARM whereas dew point temperature (DPT) is adjusted in HOMORS92. RH is derived from DPT, which is the variable measured by radiosondes. For this comparison, the values of relative humidity have been used: for HOMORS92 these are available in the dataset (albeit the formula to calculate DPT is not declared), while RHARM converts the IGRA DPT using the Magnus-Teten formula. Moreover, the HOMORS92 data are available at about 1500 IGRA stations, therefore for the comparison RHARM stations only have been selected.

The comparison of the vertical profiles of RH trends, between 850 hPa and 300 hPa in the NH (Figure 51, left panel), shows that RHARM, ERA5 and HOMORS92 have a similar shape with relative differences around 1% RH/da throughout the entire vertical range, although RHARM is the only with positive values (near zero) above 500 hPa. Both datasets significantly differ from the unadjusted IGRA data for pressures below 500 hPa. In the TR (Figure 51, right panel), HOMORS92 displays negative trends (slightly below zero) at all pressure levels, while RHARM and ERA5 show a similar shape despite a difference of up to 1.5% RH/decade, increasing with height. Differently from temperature, ERA5 RH assimilated data are not bias-adjusted. HOMORS92 uses a quantile matching algorithm to adjust the DPT histograms in between different changepoints of the time series, not fully preserving seasonal variations (Wang et al., 2013). Instead, RHARM is designed to adjust systematic effects preserving trends and natural (i.e., internal) variability. The comparison cannot ascertain which of the datasets provides the best option to assess RH trends, although some features are apparent. The vertical profiles of ERA5 and RHARM are the most similar in terms of vertical structure while HOMORS92 and ERA5 have the smallest absolute differences, especially in the TR; in the NH, RHARM and the other dataset have opposite signs. Fundamentally, the paucity of available estimates makes it difficult to assess structural uncertainties.
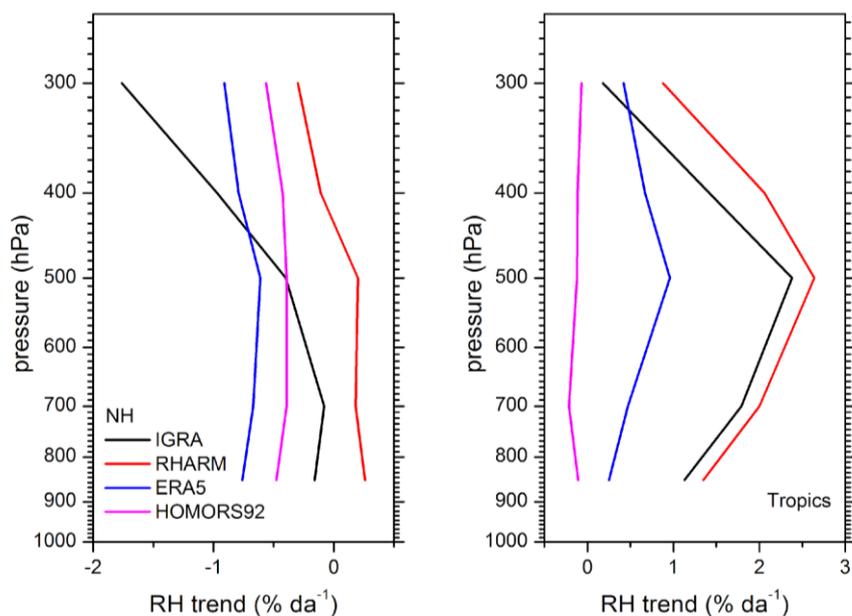
Figure 51. Profiles of relative humidity decadal trends at mandatory pressure levels between 850 and 300 hPa, in the period from 1979 to 2018, in the NH (left panel) and in the TR (right panel) for unadjusted IGRA (black line), RHARM (red), ERA5 (blue), and HOMORS92 (green). The ordinate is logarithmic, and the abscissa differs between the two panels. The comparison is limited to 300 hPa at lower pressure the comparison, and in general, the water vapour measurements might be compromised by pervasive issues on radiosonde sensors' performance in the upper troposphere and lower stratosphere, even larger for the oldest measurements.

5.3.3 Comparison with MLS

To quantify the adjustments of RHARM on RH, the water vapour mixing ratio was compared to data from the Aura MLS instrument. The Aura MLS time series at the nominal 316 hPa pressure level is chosen as being the closest to the mandatory level at 300 hPa for the IGRA/RHARM radiosondes, and because it is in good agreement with the Cryogenic FrostPoint Hygrometer (CFH) records from Boulder (Voemel et al., 2007). CFH is currently the only reference traceable instrument available for water vapour measurements in the Upper Troposphere/Lower Stratosphere (UT/LS). The possible difference in the water vapour content between the two levels at 316 and 300 hPa must be considered as an uncertainty contribution to the comparison. For the period 2006-2019, Figure 52 reveals a good agreement between RHARM and MLS, and it highlights the efficacy of the RHARM dry bias correction, especially tangible at water vapour concentrations below 0.2 g/kg. RHARM performs better in terms of mean offset, RMS differences and correlation than IGRA, as summarized in Table 19.

Tropics 0-25°N, 2006-2019
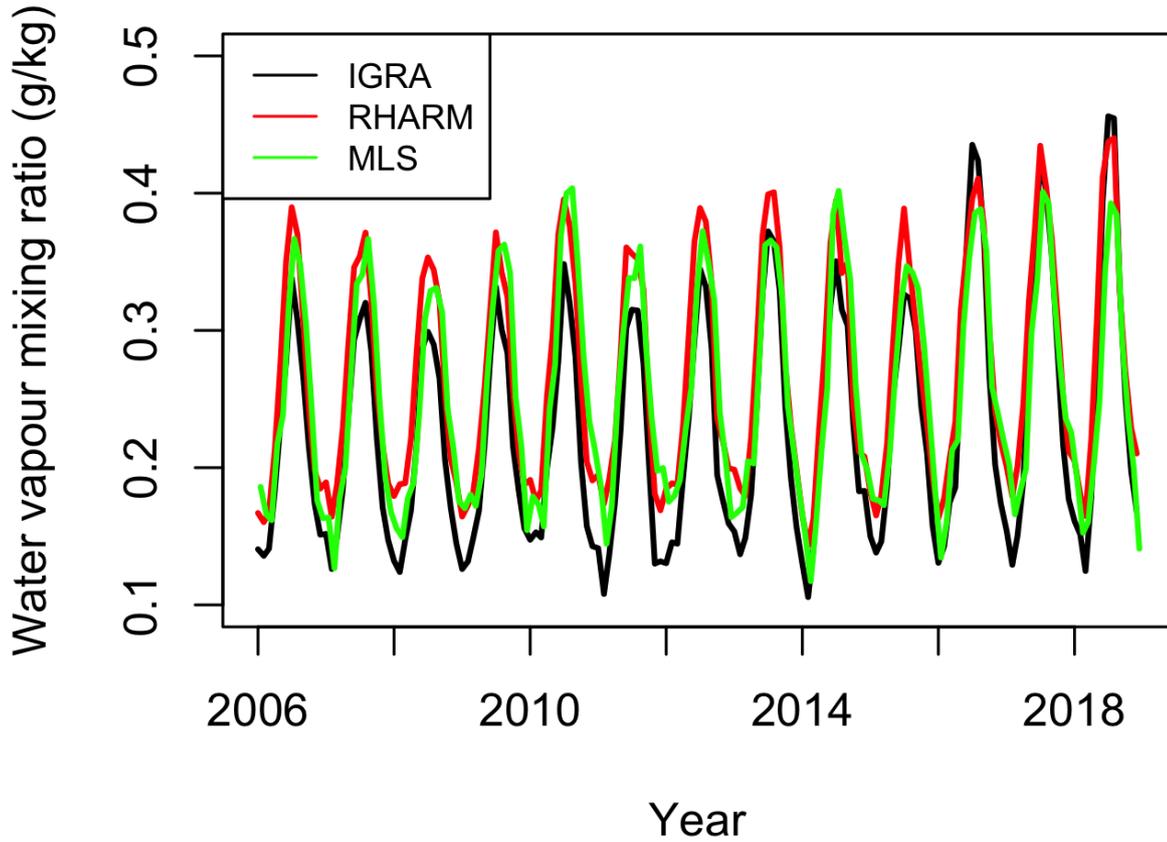
Figure 52. Comparison of monthly time series of the zonal mean water vapour mixing ratio in the northern Tropics (0°N-25°N) from 2006 to 2019. The black lines show the IGRA time series, while the red and the green lines are the corresponding RHARM and MLS time series. AURA MLS data have been subsampled at the IGRA stations within the considered domain. The MLS data are version 4 have been used (Yan et al., 2016).

	Mean difference (g/kg)	RMS difference (g/kg)	Correlation
IGRA-MLS	-0.03	0.03	0.95
RHARM-MLS	0.01	0.02	0.99

Table 19. Statistics on the comparison between IGRA-MLS and RHARM-MLS for the data are shown in Figure 52.

111

5.4 Quantification and presentation of uncertainties

A unique value of RHARM compared to other datasets is that, for the first time, an estimation of the uncertainty is provided for every single observation (i.e., at each pressure level). In this section, statistics on the estimated uncertainties are provided.

Considering data at the six stations shown in Table 20 only in the GRUAN era, the uncertainty for RHARM is generally larger than the uncertainties obtainable using the GDP as expected given the methodological considerations outlined in section 5.1 (Figure 53). In particular, for temperature (Figure 53, left panel), the median value of the GRUAN uncertainty is 0.16 K compared to 0.22 K for RHARM (median values are considered for the analysis, given the shape of the pdf). The interquartile range (IQR) for GRUAN is 0.20 K while for RHARM it is 0.26 K. These numbers confirm that on average the uncertainty estimation obtained for RHARM is greater than the GRUAN uncertainty. Nevertheless, due to the nature of the assumptions made within RHARM, its uncertainty may underestimate that of GRUAN, as seen for values below 0.1 K. These values are mainly related to nighttime measurements.

For RH (Figure 53, right panel), the median value of the GRUAN pdf is about 1.1% versus 3.6% for RHARM, with an IQR for GRUAN of 0.1% and 3.0% for RHARM. Maximum values observed with GRUAN are less than 8% while RHARM has values larger than 10% and very few values larger than 20%.

GRUAN code	Station name and country	Latitude	Longitude	Altitude	WMO index
CAB	Cabauw, Netherlands	51.97°	4.92°	1 m	06260
LIN	Lindenberg, Germany	52.21°	14.12°	98 m	10393
NYA	Ny-Ålesund, Norway	78.92°	11.92°	5 m	01004
SGP	Lamont, OK, USA	36.60°	-97.49°	320 m	74646
SOD	Sodankylä, Finland	67.37°	26.63°	179 m	02836
TAT	Tateno, Japan	36.06°	140.13°	25 m	47646

Table 20. List of the GRUAN stations used to calculate the additional calibration bias applied in the RHARM approach to adjust the Vaisala RS92 radiosoundings available from IGRA.
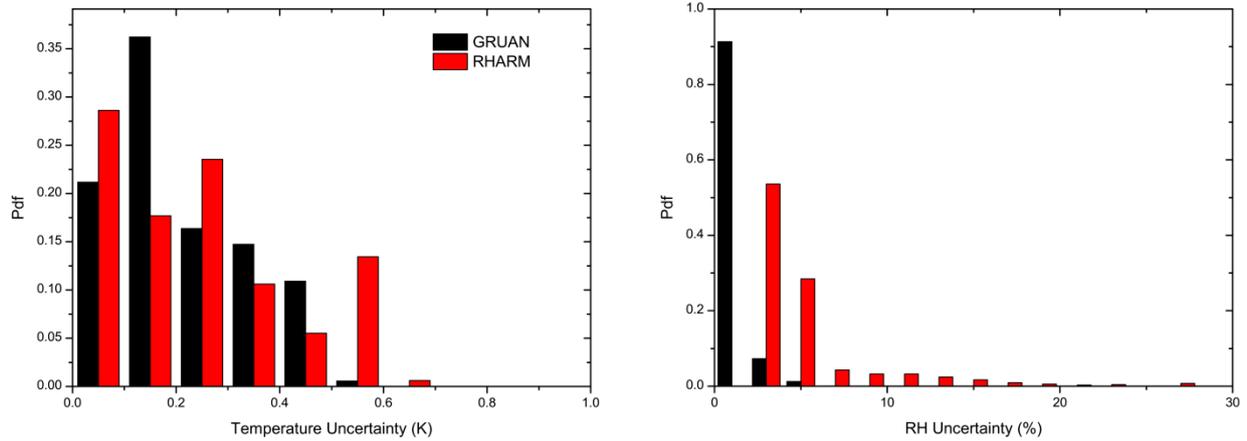
Figure 53. Comparison of pdfs of the uncertainty calculated using the GRUAN data processing (GDP) and the RHARM approach at the six stations shown in Table 20. Pdfs are relative to temperature (panel a) and relative humidity (panel b) for considering all pressure levels aggregated.

In Figure 54, the density function of the uncertainties estimated for the RHARM data is shown for the NH and the TR. The comparison for the temperature uncertainties shows that the density function in the NH is bimodal with modes centred around 0.5 K and 1.0 K, with most values smaller than 2.0 K. In the TR, values are smaller than 1.5 K. A large fraction of the values in both regions is around 0.25 K and these values refer to the values of the PPTS. For the RH, both the distributions are bimodal. A large fraction of the RH uncertainty values is smaller than 10%, while the second distribution mode is 14-15% RH uncertainty.
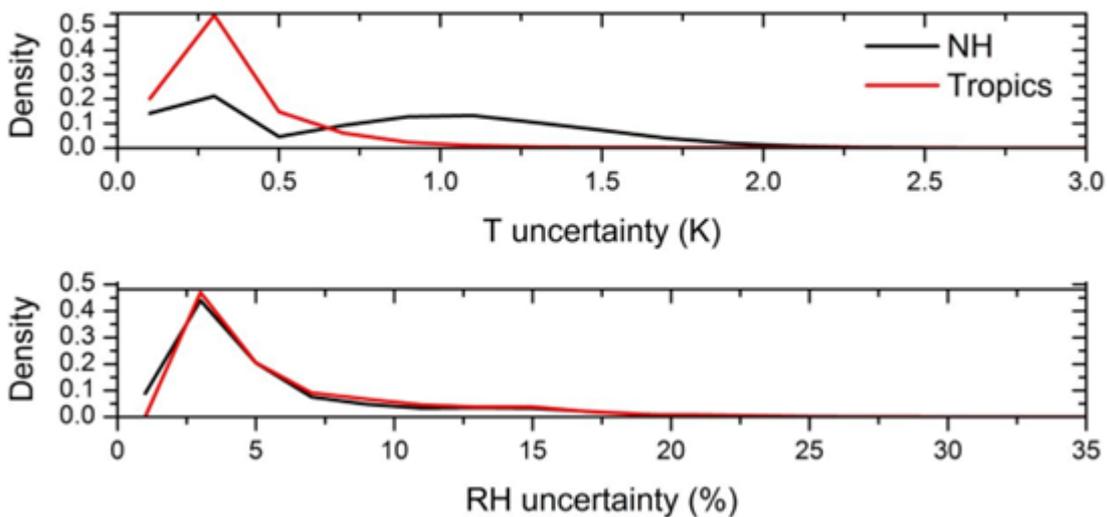
Figure 54. Comparison of the density function of the total uncertainties estimated for all the RHARM temperature (T) and relative humidity (RH) data since 01-01-1978 for the stations in the TR and in the NH.

113

To conclude the discussion on the RHARM uncertainties, a case study was discussed which illustrates the importance of uncertainties in comparison with other datasets: while a westerly wind regime characterizes the boreal circulation in the winter stratosphere (e.g., Waugh et al., 2017), occasional flow reversal may occur during sudden stratospheric warmings (SSWs). Here we consider the winters of 2017 and 2018, as in February of both years' SSWs occurred (Knight et al., 2021). Figure 55 shows daily mean temperatures at 100 hPa over the European Polar domain, showing sharp temperature increases (SSW are more evident at higher altitudes). The comparison has been carried out for the European Polar domain at 100 hPa in the lower stratosphere to benefit from the larger number of radiosounding data available than at higher pressure levels. For both the events, ERA5, IGRA and RHARM agree well and within the RHARM combined uncertainty (vertical dark grey lines) shown using k = 3, where k represents an uncertainty of 3 standard deviations equating approximately a 99% confidence level. Nevertheless, both events are preceded by a strong cooling of the lower stratosphere which in 2017 is not well captured by IGRA, due to the warm bias affecting the unadjusted radiosoundings temperature profiles. In 2018, the discrepancy between RHARM and IGRA is similar to 2017 due to the RHARM warm bias correction, while ERA5 is much warmer than both IGRA and RHARM. In this case, the estimation of the uncertainty is the only way to reveal how significant the difference is between the datasets. Other relevant differences among the three datasets can be pointed out in other months of the time series. In the right panel of Figure 55, zoom for the orange window in the left plot is shown to provide a more detailed quantification of the discrepancies among the datasets compared to the uncertainty provided by RHARM around the 2018 SSW central date.
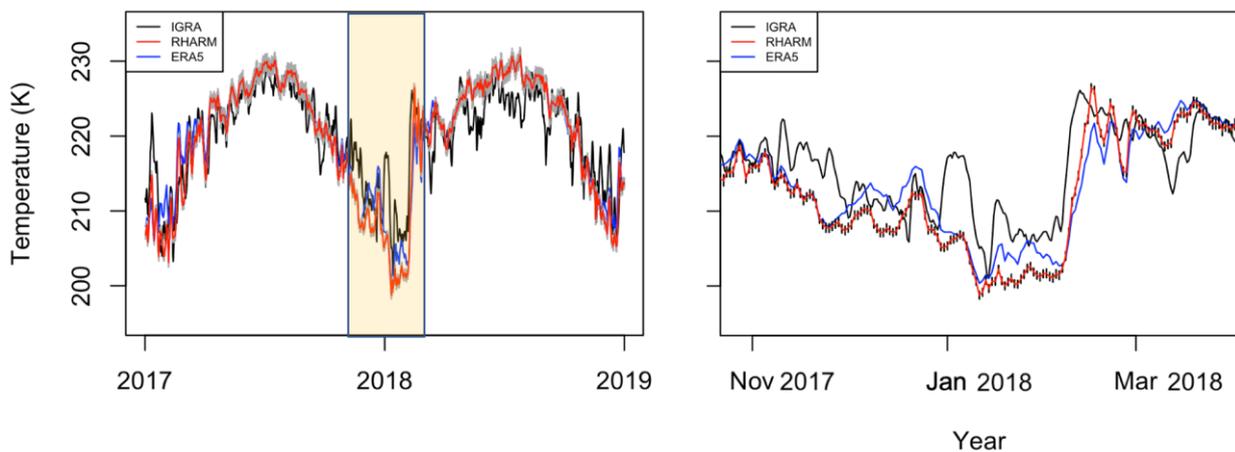
Figure 55. Left panel, comparison of time series of daily mean temperature at 100 hPa in the Polar European domain (70°N-90°N; -10°W-50°W) from 01-01-2017 to 31-12-2018; right panel, same as left panel but only for the orange window reported in the left panel. The blue lines are the ERA5 daily averages obtained by subsampling the reanalysis data at the IGRA stations within the considered domain; the black line shows the IGRA daily averages while the red line shows the corresponding RHARM averages. The combined measurement uncertainties for RHARM are also shown (vertical black error bars).

6. Summary and conclusion

The main aim of this work is the estimation of the ozone trends in the UT/LS for all latitudinal sectors, and in particular to investigate the effect of sampling and parametric uncertainties, for a new unified dataset merging the ozonesounding profiles provided by SHADOZ (https://tropo.gsfc.nasa.gov/shadoz/; Thompson et al., 2017), NDACC (https://ndacc.larc.nasa.gov/) and WOUDC (https://woudc.org/home.php). For this analysis, only profiles with sufficient quality in terms of plausibility, coherency, and consistency of the measurements of ozone partial pressure have been selected. Furthermore, temperature trends in the UT/LS and water vapour trends in the UT were also studied, based mainly on profiles obtained through radiosounding measurements provided by GRUAN reference, IGRA, and RHARM bias-adjusted datasets; comparisons with ERA5 atmospheric reanalysis have been also presented.

To study ozone trends, the stations belonging to the unified ozonesounding dataset are first classified into two clusters, based on the measurement of temporal coverage and completeness: a Long Coverage cluster (LC), and a medium coverage cluster (MC). A third cluster (long and medium coverage cluster, LMC) has been considered and obtained from the LC and MC combination to compare the relative effect of data completeness versus data spatial coverage. This comparison is made using the trends estimated for the various clusters. The trend estimates were carried out using LOTUS multiple linear regression (Petropavlovskikh et al., 2019; Godin-Beekmann et al., 2022) and three linear regression methods: least-square (Reinsel et al. 2002; Appenzeller et al., 2000; Weiss et al., 2001; Tiao et al., 1990; Weatherhead et al., 1998), least absolute deviation (Rice and White, 1964; Barrodale, 1968; Wong and Schneider Jr., 1989; Calitz and Rüther, 1996; Santer et al., 2000; Barrodale and Roberts, 1974) and Theil-Sen (Theil, 1950; Siegel and Benson, 1982; Helsel and Hirsch, 1992; Dang et al., 2008) linear methods. The significance of the trends was verified using the Mann-Kendall test (Kendall, 1975; Mann, 1945).

The comparison indicated that:

- The difference between the trends for the three clusters showed that the usage of the smallest but highest quality cluster, LC, provides the largest number of significant trends at the different vertical ranges and latitudes, with a small effect of spatial sampling and the smallest structural uncertainties. The latter was studied considering the differences among three linear regression methods: least-square, least absolute deviation and Theil-Sen.

- The study of sampling uncertainties shows that the NH polar latitudes, NH mid-latitudes and Tropics have uncertainties enabling a significant characterization of the ozone trends adequately, so the analysis was carried out only for these three sectors.

- The trend estimates on the LC cluster for the NH mid-latitudes and Tropics are also compared with other studies, already present in the literature (Petropavlovskikh et al., 2019; Sofieva et al., 2021), and show, for the NH mid-latitudes, a negative trend of 5% for the period pre-2000 at 50-1 hPa layer, reaching a negative trend of 10% at 100-50 hPa, and a negative trend of 1% for the period post-2000 at 50-1 hPa, with a positive trend of 4% at 100-50 hPa; for the Tropics and for the period pre-2000, a positive trend of about 8% at 50-1 hPa was estimated, reaching 10% at 100-50 hPa and, for post-2000, a positive trend of 2% at 50-1 hPa.

For temperature and water vapour, the analyses of interannual variations and trends from RHARM data in the period 1979–2018 show:

- Warming trends of temperature are smaller than 0.5 K da^{-1} at pressures higher than 250 hPa, while trends are cooling up to 0.25 K da^{-1} below that level. In the Tropics, trends are smaller at 0.25 K da^{-1} at a pressure higher than 250 hPa, while cooling and within 0.5 K da^{-1} below. Results are in good agreement with ERA5, especially for pressures higher than 200 hPa in the NH mid-latitudes and higher than 50 hPa in the Tropics;

- For RH, the NH mid-latitudes RHARM shows slightly positive or near-zero trends at pressures higher than 500 hPa, while trends are negative (up to 0.2% RH da^{-1}) below. In the Tropics, trends are positive and higher than 1.0% RH da^{-1} over the entire vertical range (larger than 2.0% RH da^{-1} at 500 hPa). Comparisons with ERA5 show differences are probably also due to the fact that, differently from temperature, ERA5 RH assimilated data are not bias-adjusted. The increasing humidity anomalies in the period 2015–2019 and the positive RH trend observed in RHARM data in the Tropics appear to be correlated with the warm ENSO event in 2015/16 (Madonna et al., 2022).

The future work will be extended to analyse synergically the trends of ozone, temperature and water vapour in the UT/LS (for water vapour in the UT only, due to the high uncertainties on the radiosonde measurements). This is a fundamental aspect because ozone is closely associated with temperature variations in the UT/LS, while the analysis of water vapor, in the same region of the atmosphere, is important to study the signals of climate variability and has also already been correlated to variations in ozone concentrations (Salby, 1995). Trends from in situ soundings will be also compared with satellite and reanalysis data. A key issue to investigate before further comparisons with other datasets will be to assess the effect of inhomogeneities due to changes over time in the measurement sensors and algorithms, which can influence the trend estimates. Moreover, the effect of serial measurement correlation will be added to the trend estimates, comparing the existing solution provided in the literature.

7. Appendix A: station list for the unified dataset

Station	Longest continuous data series	Classification
ABERYSTWYTH	1 year	Short coverage
AINSWORTH	1 month	Short coverage
ALAJUELA	4 years	Short coverage
ALBROOK	1 month	Short coverage
ALERT	33 years	Long coverage
AMUNDSEN-SCOTT	2 years	Short coverage
ANKARA	6 years and 5 months	Short coverage
ASCENSION_ISLAND	11 years and 3 months	Medium coverage
ASPENDALE	4 years and 8 months	Short coverage
BARBADOS	1 month	Short coverage
BARROW	1 month	Short coverage
BEDFORD	no data between 1978-01-01 and 2021-07-31	Short coverage
BELGRANO	4 years and 2 months	Short coverage
BELTSVILLE	1 month	Short coverage
BERLIN/TEMPLEHOF	no data between 1978-01-01 and 2021-07-31	Short coverage
BISCARROSSE/SMS	5 years	Short coverage
BOGOTA	2 years and 2 months	Short coverage
BOMBAY	no data between 1978-01-01 and 2021-07-31	Short coverage
BOULDER	29 years and 2 months	Long coverage
BRATTS_LAKE	8 years	Short coverage
BRAZZAVILLE	2 years and 6 months	Short coverage
BROADMEADOWS	21 years and 11 months	Long coverage
BYRD	no data between 1978-01-01 and 2021-07-31	Short coverage
CAGLIARI/ELMAS	2 years and 6 months	Short coverage
CANTON_ISLAND	no data between 1978-01-01 and 2021-07-31	Short coverage
CARIBOU	1 month	Short coverage
CHEJU	4 months	Short coverage
CHILCA	no data between 1978-01-01 and 2021-07-31	Short coverage
CHRISTCHURCH	no data between 1978-01-01 and 2021-07-31	Short coverage
CHURCHILL	19 years	Medium coverage
COLD_LAKE	4 years	Short coverage
COOLIDGE_FIELD	no data between 1978-01-01 and 2021-07-31	Short coverage
COSTA RICA	11 years	Medium coverage

COTONOU	2 years	Short coverage
CUIABA	2 months	Short coverage
DAVIS	6 years and 8 months	Short coverage
DEBILT	28 years and 1 month	Long coverage
DENVER	no data between 1978-01-01 and 2021-07-31	Short coverage
DUMONT	11 years	Medium coverage
EASTER_ISLAND	2 years	Short coverage
EDMONTON STONY_PLAIN	42 years and 11 months	Long coverage
EGBERT	8 years	Short coverage
EL_ARENOSILLO	1 month	Short coverage
ETOSHA_PAN	2 months	Short coverage
EUREKA	27 years and 2 months	Long coverage
FAIRBANKS	no data between 1978-01-01 and 2021-07-31	Short coverage
FT._SHERMAN	no data between 1978-01-01 and 2021-07-31	Short coverage
GIMLI	5 years and 1 month	Short coverage
GOOSE_BAY	35 years and 1 month	Long coverage
GREAT_FALLS	no data between 1978-01-01 and 2021-07-31	Short coverage
HALLETT	no data between 1978-01-01 and 2021-07-31	Short coverage
HANOI	5 years	Short coverage
HEREDIA	1 year and 2 months	Short coverage
HILO	38 years	Long coverage
HOHENPEISSENBERG	43 years and 5 months	Long coverage
HOLTVILLE	1 month	Short coverage
HONG_KONG_OBSERVATORY	12 years and 11 months	Medium coverage
HOUSTON	5 months	Short coverage
HUNTSVILLE	8 years and 8 months	Short coverage
IQALUIT	5 months	Short coverage
IRENE	7 years and 7 months	Short coverage
ISFAHAN	2 years and 1 month	Short coverage
IZANA	25 years and 7 months	Long coverage
JOKIOINEN	2 years	Short coverage
KAASHIDHOO	2 months	Short coverage
KAGOSHIMA	14 years	Medium coverage
KELOWNA	11 years and 7 months	Medium coverage
KOUROU	no data between 1978-01-01 and 2021-07-31	Short coverage
KUALA LUMPUR	12 years and 1 month	Medium coverage

LA_REUNION_ISLAND	12 years and 7 months	Medium coverage
LAUDER	35 years and 2 months	Long coverage
LAVERTON	6 years and 4 months	Short coverage
LEGIONOWO	30 years and 8 months	Long coverage
LERWICK	12 years and 10 months	Medium coverage
LINDENBERG	37 years	Long coverage
LONG_VIEW	no data between 1978-01-01 and 2021-07-31	Short coverage
MACQUARIE_ISLAND	17 years and 1 month	Medium coverage
MADRID BARAJAS	17 years	Medium coverage
MAITRI	4 years and 11 months	Short coverage
MALINDI	6 years and 10 months	Short coverage
MARAMBIO	15 years and 7 months	Medium coverage
MCDONALD_OBSERVATORY	no data between 1978-01-01 and 2021-07-31	Short coverage
MCMURDO	11 months	Short coverage
MIRNY	2 years and 5 months	Short coverage
MOUNT_ABU	no data between 1978-01-01 and 2021-07-31	Short coverage
NAHA	29 years and 8 months	Long coverage
NAIROBI	13 years and 7 months	Medium coverage
NARRAGANSETT	5 months	Short coverage
NATAL	12 years and 6 months	Medium coverage
NEUMAYER	28 years and 7 months	Long coverage
NEW_DELHI	6 years	Short coverage
NOVOLASAREVSKAYA/FORSTER	5 years and 9 months	Short coverage
NY_ALESUND	30 years	Long coverage
OHP	16 years and 9 months	Medium coverage
OVEJUYO	no data between 1978-01-01 and 2021-07-31	Short coverage
PAGO PAGO AMERICAN SAMOA	25 years	Long coverage
PALESTINE	6 years and 8 months	Short coverage
PAPEETE	4 years and 5 months	Short coverage
PARADOX	2 months	Short coverage
PARAMARIBO	13 years and 2 months	Medium coverage
PAYERNE	42 years and 8 months	Long coverage
PELLSTON	2 months	Short coverage
PETALING_JAYA	2 years and 3 months	Short coverage
POHANG	3 years	Short coverage
POKER_FLAT	3 years	Short coverage

PORT_HARDY	2 years and 6 months	Short coverage
PORTO_NACIONAL	2 months	Short coverage
PRAHA	3 years	Short coverage
PUERTO_MONTT	no data between 1978-01-01 and 2021-07-31	Short coverage
PUNE	6 years and 10 months	Short coverage
RESOLUTE	15 years and 10 months	Medium coverage
RICHLAND	2 months	Short coverage
S.PIETRO_CAPOFIUME	2 years and 9 months	Short coverage
SABLE_ISLAND	2 months	Short coverage
SALEKHARD	2 years and 1 month	Short coverage
SAN_CRISTOBAL	3 years and 6 months	Short coverage
SAN_DIEGO	no data between 1978-01-01 and 2021-07-31	Short coverage
SAN_JUAN	no data between 1978-01-01 and 2021-07-31	Short coverage
SAPPORO	28 years and 2 months	Long coverage
SCORESBYSUND	26 years and 10 months	Long coverage
SEPANG_AIRPORT	12 years and 1 month	Medium coverage
SINGAPORE	4 years	Short coverage
SODANKYLA	23 years and 2 months	Long coverage
SOFIA	9 years and 10 months	Short coverage
SOUTH POLE	29 years and 9 months	Long coverage
SPOKANE	no data between 1978-01-01 and 2021-07-31	Short coverage
STERLING	no data between 1978-01-01 and 2021-07-31	Short coverage
SUMMIT	11 years and 8 months	Medium coverage
SUVA	8 years and 9 months	Short coverage
SYOWA	34 years and 7 months	Long coverage
TABLE_MOUNTAIN	6 months	Short coverage
TAIPEI	7 years and 2 months	Short coverage
TATENO	31 years and 6 months	Long coverage
TECAMEC	6 months	Short coverage
THALWIL	no data between 1978-01-01 and 2021-07-31	Short coverage
THIRUVANANTHAPURAM	7 years	Short coverage
THULE	4 months	Short coverage
TOPEKA	no data between 1978-01-01 and 2021-07-31	Short coverage
TORONTO	3 months	Short coverage
TRINIDAD_HEAD	2 years and 5 months	Short coverage
UCCLE	33 years and 5 months	Long coverage

USHUAIA	3 years and 8 months	Short coverage
VALENTIA	16 years	Medium coverage
VALPARAISO	4 months	Short coverage
VANSCOY	2 years and 2 months	Short coverage
VIGNA_DI_VALLE	4 years 10 months	Short coverage
WALLOPS_ISLAND	26 years and 2 months	Long coverage
WALSINGHAM	2 months	Short coverage
WATUKOSEK	15 years and 2 months	Medium coverage
WILKES	no data between 1978-01-01 and 2021-07-31	Short coverage
YAKUTSK	4 months	Short coverage
YARMOUTH	17 years and 2 months	Medium coverage
YORKTON	1 month	Short coverage

Table 21. Information about the station's coverage classification.

8. Bibliography

Andrews, D. G., Holton, J. R., & Leovy, C. B. (1987). Middle Atmosphere Dynamics. *Academic Press, Orlando*, 489 pp.

B., H. (1998). ITS-90 FORMULATIONS FOR VAPOR PRESSURE, FROSTPOINT TEMPERATURE, DEWPOINT TEMPERATURE, AND ENHANCEMENT FACTORS IN THE RANGE –100 TO +100 C. *The Proceedings of the Third International Symposium on Humidity & Moisture*.

B., K.-D. D., Hintsa, E. J., Anderson, J. G., & Keith, D. W. (1999). The effect of climate change on ozone depletion through changes in stratospheric water vapour. *Nature, 402*, 300-401.

Baldissera Pacchetti, M. (2021). Structural uncertainty through the lens of model building. *Synthese, 198*, 10377–10393.

Barrodale, I. (1968). L1 Approximation and the Analysis of Data. *Royal Statistical Society, 17*, 51-57.

Bodeker, G. E., Bojinski, S., Cimini, D., Dirksen, R. J., Haeffelin, M., Hannigan, J. W., . . . Wang, J. (2016). Reference Upper-Air Observations for Climate: From Concept to Reality. *Bulletin of the American Meteorological Society, 97*, 123–135.

Bojilova, R., Mukhtarov, P., & Miloshev, N. (2022). Latitude Dependence of the Total Ozone Trends for the Period 2005–2020: TOC for Bulgaria in the Period 1996–2020. *Atmosphere, 13*(6).

Brasseur, G. P. (2008). Climate Variability and Extremes during the Past 100 years: Creating Knowledge from the Controntation of Observations and Models: The Case of Stratospheric Ozone. *Advances in Global Change Research, 33*, 303–316.

Brewer, A. W. (1949). Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteor. Soc., 75*(6), 351–363.

Butchart, N., Cionni, I., Eyring, V., Shepherd, T. G., Waugh, D. W., Akiyoshi, H., . . . Tian, W. (2010). Chemistry-climate model simulations of twenty-first century stratospheric climate and circulation changes. *J. Clim., 23*, 5349–5374.

Calitz, M. F., & Rüther, H. (1996). Least absolute deviation (LAD) image matching. *ISPRS Journal of Photogrammetry and Remote Sensing, 51*, 223-229.

Cardinali, C., & Healy, S. B. (2014). Impact of GPS radio occultation measurements in the ECMWF system using adjointbased diagnostics. *Q. J. Roy. Meteor. Soc., 140*, 2315–2320.

Chapman, S. (1930). A theory of upper-atmospheric ozone. *Memoirs of the Royal Meteorological Society, 3(26)*(3).

co-authors:, K. (2021). Predictability of European Winters 2017/2018 and 2018/2019: Contrasting influences from the Tropics and stratosphere. *RMetS*, e1009.

Connor, B. J., Parrish, A., Tsou, J. J., & McCormick, M. P. (1995). Error analysis for the ground-based microwave ozone measurements during STOIC. *J. Geophys. Res., 100*, 9283-9291.

Cucurull, L., Derber, J. C., Treadon, R., & Purser, R. J. (2007). Assimilation of Global Positioning System Radio Occultation Observations into NCEP's Global Data Assimilation System. *Mon. Weather Rev., 135*, 3174–3193.

Dai, A., Wang, J., Thorne, P. W., Parker, D. E., Haimberger, L., & Wang, X. L. (2011). A New Approach to Homogenize Daily Radiosonde Humidity Data. *Journal of Climate, 24*, 965–991.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., . . . al., e. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc., 137*, 553–597.

Dee, D., Fasullo, J., Shea, D., Walsh, J., & (Eds)., N. C. (2016). The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables.

Deshler, T., Mercer, J., Smit, H. J., Stuebi, R., Levrat, G., & Johnson, B. J. (2008). Atmospheric comparison of electrochemical cell ozonesondes from different manufacturers, and with different cathode solution strengths: The Balloon Experiment on Standards for Ozonesondes. *Journal of Geophysical Research, 113*, D04307.

Deshler, T., Stübi, R., Schmidlin, F. J., Mercer, J. L., Smit, H. J., & Johnson, B. J. (2017). Methods to homogenize ECC ozonesonde measurements across changes in sensing solution concentration or ozonesonde manufacturer. *Atmospheric Measurement Techniques, 10*, 2012–2043.

Dessler, A. E., Schoeberl, M. R., Wang, T., Davis, S. M., & Rosenlof, K. H. (2013). Stratospheric water vapor feedback. *PNAS*.

Dessler, A., & Davis, S. (2010). Trends in tropospheric humidity from reanalysis systems. *J. Geophys. Res., 115*, D19127.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., & Vömel, H. (2014). Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmos. Meas. Tech., 7*, 4463–4490.

Dirksen, R., & al., e. (2020). Managing the transition from Vaisala RS92 to RS41 radiosondes within the Global Climate Observing System Reference Upper-Air Network (GRUAN): a progress report. *Geosci. Instrum. Method. Data Syst., 9*, 337–355.

Dobson, G. B. (1956). Origin and distribution of polyatomic molecules in the atmosphere. *The Royal Society, A, Mathematical and Physical Sciences, 236, 1205*(6), 187–193.

Durre, I., Menne, M. J., & Vose, R. S. (2008). Strategies for Evaluating Quality Assurance Procedures. *Journal of Applied Meteorology and Climatology, 47*, 1785–1791.

Durre, I., Vose, R. S., & Wuertz, D. B. (2006). Overview of the Integrated Global Radiosonde Archive. *Journal of Climate, 19*, 53–68.

Durre, I., Yin, X., Vose, R. S., Applequist, S., & Arnfield, J. (2018). Enhancing the Data Coverage in the Integrated Global Radiosonde Archive. *Journal of Atmospheric and Oceanic Technology, 35*, 1753–1770.

F., M. (2020c). Can reference radiosounding measurements be used to improve historical time series? *Il Nuovo Cimento C, 43*, 1-10.

Farman, J. C., Gardiner, B. G., & Shanklin, J. D. (1985). Large losses of total ozone in the Antarctica reveal seasonal ClOx/NOx interaction. *Nature, 315*, 207–210.

Fleming, E. L., Chandra, S., Barnett, J. J., & Corney, M. (1990). Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. *Advances in Space Research, 12*, 1211-1259.

Fortuin, J. F., & Kelder, H. (1998). An ozone climatology based on ozonesonde and satellite measurements. *Journal of Geophysical Research – Atmospheres, 103(D24)*(31), 709-734.

Fortuin, J. F., & Langematz, U. (1994). An update on the global ozone climatology and on concurrent ozone and temperature trends. *SPIE, Atmospheric Sensing and Modeling*(2311), 207-216.

Free, M., Angle, J. K., Durre, I., Lanzante, J., Peterson, T. C., & Seidel, D. J. (2004). Using first differences to reduce inhomogeneity in radiosonde temperature datasets. *J. Clim., 17*, 4171–4179.

Girolamo, P. D., Cuomo, V., Pappalardo, G., & Velotta, R. (1994). Lidar validation of temperature and water vapor satellite measurements. *PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING*.

Godin, S. e. (1999). Ozone Differential Absorption Lidar Algorithm Intercomparison. *Appl. Opt., 38*(30), 6225-6236.

Godin-Beekmann, S., Azouz, N., Sofieva, V. F., Hubert, D., Petropavlovskikh, I., Effertz, P., . . . Sussmann, R. (2022). Updated trends of the stratospheric ozone vertical distribution in the 60° S–60° N latitude range based on the LOTUS regression model. *ACP, 22*, 11657–11673.

Godin-Beekmann, S., Porteneuve, J., & Garnier, A. (2003). Systematic DIAL lidar monitoring of the stratospheric ozone vertical distribution at Observatoire de Haute-Provence (43.92°N, 5.71°E). *J. Env. Monitoring, 5*, 57-67.

Haigh, J. D., & Pyle, J. A. (1982). Ozone perturbation experiments in a two-dimensional circulation model. *Q. J. R. Meteorol. Soc., 108*, 551–574.

Haimberger, L., Tavolato, C., & Sperka, S. (2012). Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations. *Journal of Climate, 25*, 8108–8131.

Hajj, G. A., Kursinski, E. R., Romans, L. J., Bertiger, W. L., & Leroy, S. S. (2002). A Technical Description of Atmospheric Sounding By Gps occultation. *J. Atmos. Sol.-Terr. Phy., 64*, 451–469.

Harris, N. P., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., . . . Zawodny, J. M. (2015). Past changes in the vertical distribution of ozone – Part 3: Analysis and interpretation of trends. *Atmos. Chem. Phys., 15*, 9965-9982.

Hartmann, J., West, A., Renforth, P., Köhler, P., De La Rocha, C. L., Wolf-Gladrow, D. A., . . . Scheffran, J. (2013). Enhanced chemical weathering as a geoengineering strategy to reduce atmospheric carbon dioxide, supply nutrients, and mitigate ocean acidification. *Reviews of Geophysics, 51*, 113-149.

Hase, F. (2000). Inversion von Spurengasprofilen aus hochaufgelösten bodengebundenen FTIR-Messungen in Absorption, Dissertation. *Wissenschaftliche Berichte Forschungszentrum Karlsruhe*.

Hase, F., Blumenstock, T., & Paton-Walsh, C. (1999). Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software. *Appl. Opt., 38*, 3417-3422.

Hassler, B. e. (2014). Past changes in the vertical distribution of ozone – Part 1: Measurement techniques, uncertainties and availability. *Atmos. Meas. Tech., 7*, 1395-1427.

Haynes, P. (2005). Stratospheric dynamics. *Annu. Rev. Fluid Mech., 37*(6), 263–293.

Healy, S. B., & Eyre, J. R. (2000). Retrieving temperature, water vapour and surface pressure information from refractive index profiles derived by radio occultation: A simulation study. *Q. J. Roy. Meteor. Soc., 126*, 1661-1683.

Healy, S. B., Jupp, A. M., & Marquardt, C. (2005). Forecast impact experiment with GPS radio occultation measurements. *Geophys. Res. Lett., 32*, L03804.

Helsel, D. R., & Hirsch, R. M. (1992). Statistical methods in water resources. *Elsevier*.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J. N. (2020). The ERA5 global reanalysis. *RMetS, 146*, 1999-2049.

Hocke, K., Kämpfer, N., Ruffieux, D., Froidevaux, L., Parrish, A., Boyd, I., . . . Kyrölä, E. (2007). Comparison and synergy of stratospheric ozone measurements by satellite limb sounders and the ground-based microwave radiometer SOMORA. *Atmos. Chem. Phys., 7*, 4117-4131.

Hoshino, S., & al., e. (2022). Comparison of GRUAN data products for Meisei iMS-100 and Vaisala RS92 radiosondes at Tateno, Japan. *Atmos. Meas. Tech., 15*, 5917–5948.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., . . . Johnson, C. A. (2001). IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*, 944 pp.

Hu, S., & Fedorov, A. (2017). The extreme El Niño of 2015-2016 and the end of global warming hiatus: GLOBAL WARMING HIATUS AND 2015 EL NIÑO. *Geophys. Res. Lett., 44*, 3816–3824.

IPCC/TEAP. (2005). Safeguarding the Ozone Layer and the Global Climate System.

Johnson, B. J., Oltmans, S. J., Vömel, H., Smit, H. J., Deshler, T., & Kroeger, C. (2002). ECC ozonesonde pump efficiency measurements and tests on the sensitivity to ozone of buffered and unbuffered ECC sensor cathode solutions. *Journal of Geophysical Research, 107*, D19.

Jonsson, A. I., Fomichev, V. I., & Shepherd, T. G. (2009). The effect of nonlinearity in CO2 heating rates on the attribution of stratospheric ozone and temperature changes. *Atmos. Chem. Phys., 9*, 8447–8452.

Kendall, M. G. (1975). Rank Correlation Methods. *Charles Griffin 4th edition*.

Ko, M. W., Newman, P. A., Reimann, S., & Strahan, S. E. (2013). Lifetimes of Stratospheric Ozone-Depleting Substances, Their Replacements, and Related Species. *World Climate Research Programme, SPARC Report No. 6*, WCRP-15/2013.

Kremser, S., Tradowsky, J. S., Rust, H. W., & Bodeker, G. E. (2018). Is it feasible to estimate radiosonde biases from interlaced measurements? *AMT, 11*, 3021–3029.

Kursinski , E. R., Hajj, G. A., Hardy, K. R., Schofield, J. T., & Linfield, R. (1997). Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. *JGR, 102*, 23429-23465.

Leblanc, T. e. (2016b). Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 2: Ozone DIAL uncertainty budget. *Atmos. Meas. Tech., 9*, 4051-4078.

Leblanc, T., & McDermid, I. S. (2000). Stratospheric ozone climatology from lidar measurements at Table Mountain (34.4°N, 117.7°W) and Mauna Loa (19.5°N, 155.6°W). *J. Geophys. Res., 105(D11)*, 14613–14623.

Lossow, S. e. (2019). The SPARC water vapour assessment II: profile-to-profile comparisons of stratospheric and lower mesospheric water vapour data sets obtained from satellites. *AMT, 12*, 2693–2732.

Madonna, F., Tramutola, E., Sy, S., Serva, F., Proto, M., Rosoldi, M., . . . Thorne, P. W. (2022). The New Radiosounding HARMonization (RHARM) Data Set of Homogenized Radiosounding Temperature, Humidity, and Wind Profiles With Uncertainties. *JGR, 127*.

Maillard Barras, E., Ruffieux, D., & Hocke, K. (2009). Stratospheric ozone profiles over Switzerland measured by SOMORA, ozonesonde, and MLS/Aura satellite. *Int. J. Remote Sens., 30*, 4033–4041.

Mann, H. B. (1945). Non-parametric tests against trend. *Econometrica, 13*, 163-171.

McCarthy, M. P., Titchner, H. A., Thorne, P. W., Tett, S. B., Haimberger, L., & Parker, D. E. (2008). Assessing Bias and Uncertainty in the HadAT-Adjusted Radiosonde Climate Record. *Journal of Climate, 21*, 817–832.

McCarthy, M., Thorne, P., & Titchner, H. (2009). An Analysis of Tropospheric Humidity Trends from Radiosondes. *Journal of Climate, 22*, 5820–5838.

McKenzie, R. L., Aucamp, P. J., Bais, A. F., Björn, L. O., & Ilyas, M. (2007). Changes in biologically-active ultraviolet radiation reaching the Earth's surface. *Photochem. Photobiol. Sci., 6*, 218–231.

Mégie, G., & Menzies, R. T. (1980). Complementarity of UV and IR differential absorption lidar for global measurements of atmospheric species. *Appl. Opt., 19*, 1173-1183.

Metz, B., Kuijpers, L., Solomon, S., Andersen, S. O., Davidson, O., Pons, J., . . . Meyer, L. (2005). IPCC/TEAP. *Cambridge University Press, UK*, 478.

Michelsen, H. A., Salawitch, R. J., Gunson, M. R., Aellig, C., Kämpfer, N., Abbas, M. M., . . . Zander, R. (1996). Stratospheric chlorine partitioning: Constraints from shuttle-borne measurements of [HCl], [ClNO3], and [ClO]. *Geophys. Res. Lett., 23*, 2361–2364.

Miloshevich, L. M., Paukkunen, A., Vömel, H., & Oltmans, S. J. (2004). Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements. *Journal of Atmospheric and Oceanic Technology, 21*, 1305–1327.

Molina, M. J., & Rowland, F. S. (1974). Stratospheric sink for chlorofluoromethanes, chlorine atom catalysed destruction of ozone. *Nature, 249*(4), 810–812.

Montzka, S. A., & al., e. (1999). Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature, 398*, 690–694.

Montzka, S. A., Butler, J. H., Elkins, J. W., Thompson, T. M., Clarke, A. D., & Lock, L. T. (1999). Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature, 398*, 690–694.

Montzka, S. A., Butler, J. H., Hall, B. D., Mondeel, D. J., & Elkins, J. W. (2003). A decline in tropospheric organic bromine. *Geophysical Research Letters, 30*.

Nakano, T., & Morofuji, T. (2023). Development of an automated pump-efficiency measuring system for ozonesondes utilizing an airbag-type flowmeter. *AMT, 16*, 1583–1595.

Nash, J., Oakley, T., Vömel, H., & Wei, L. (2011). WMO intercomparison of high quality radiosonde systems. *WMO/TD-No. 1580*.

Nedoluha, G. e. (2015). Unusual stratospheric ozone anomalies observed in 22 years of measurements from Lauder, New Zealand. *Atmos. Chem. Phys., 15*, 6817–6826.

Newchurch, M. J., Eun-Su Yang, Cunnold, D. M., Reinsel, G. C., Zawodny, J. M., & Russell III, J. M. (2003). Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery. *J. Geophys. Res., 108*, 4507.

Newman, P. A., Daniel, J. S., Waugh, D. W., & Nash, E. R. (2007). A new formulation of equivalent effective stratospheric chlorine (EESC). *Atmos. Chem. Phys., 7*, 4537–4552.

Nisbet, E. G., Dlugokencky, E. J., Manning, M. R., Lowry, D., Fisher, R. E., France, J. L., . . . Ganesan, A. L. (2016). Rising atmospheric methane: 2007–14 growth and isotopic shift. *Glob. Biogeochem. Cycles, 30*, 1356–1370.

O'Doherty, S., & al., e. (2004). Rapid growth of hydrofluorocarbon 134a and hydrochlorofluorocarbons 141b, 142b, and 22 from Advanced Global Atmospheric Gases Experiment (AGAGE) observations at Cape Grim, Tasmania, and Mace Head, Ireland. *J. Geophys. Res., 109*, D06310.

Oort, A., & Peixóto, J. (1983). Global Angular Momentum and Energy Balance Requirements from Observations. *Advances in Geophysics, 25*, 355-490.

Pascoe, C., Gray, L., Crooks, S., Juckes, M., & Baldwin, M. (2005). The quasi-biennial oscillation: Analysis using ERA-40 data. *JGR, 110*.

Pelon, J., Godin, S., & Mégie, G. (1986). Upper stratospheric (30–50 km) lidar observations of the ozone vertical distribution. *Journal of Geophysical Research: Atmospheres, 91*, 8667-8671.

Petropavlovskikh, I., Godin-Beekmann, S., Hubert, D., Damadeo, R., Hassler, B., & Sofieva, V. (2019). SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere. *SPARC Report No. 9, GAW Report No. 241,*, WCRP-17/2018.

Philipona, R., & Kräuchi, A. (2016). Return glider radiosonde for in situ upper-air research measurements. *AMT, 9*, 2535–2544.

Polvani, L. M., Wang, L., Aquila, V., & Waugh, D. W. (2017). The Impact of Ozone-Depleting Substances on Tropical Upwelling, as Revealed by the Absence of Lower-Stratospheric Cooling since the Late 1990s. *Journal of Climate, 30*, 2523-2534.

Pougatchev, N. S., Connor, B. J., & Rinsland, C. P. (1995). Infrared measurements of the ozone vertical distribution above Kitt Peak. *J. Geophys. Res., 100*, 16689–16697.

Ramella Pralungo, L., & Haimberger, L. (2014). A "Global Radiosonde and tracked-balloon Archive on Sixteen Pressure levels" (GRASP) going back to 1905 – Part 2: homogeneity adjustments for pilot balloon and radiosonde wind data. *Earth Syst. Sci. Data, 6*, 297–316.

Randel, W. J., Shine, K. P., Austin, J., Barnett, J., Claud, C., Gillet, N. P., . . . Yoden, S. (2009). An update of observed stratospheric temperature trends. *JGR, 114*, D02107.

Reinsel, G. C., Weatherhead, E., Tiao, G. C., Miller, A. J., Nagatani, R. M., Wuebbles, D. J., & Flynn, L. E. (2002). On detection of turnaround and recovery in trend for ozone. *JGR, 107*, ACH 1-1-ACH 1-12.

Rice, J. R., & White, J. S. (1964). Norms for smoothing and estimation. *SIAM Review, 6*(3).

Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M., & Forster, P. (2012). Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects. *JGR, 117*.

Rodgers, C. D. (2000). Inverse Methods for Atmospheric Sounding: Theory and Practice, Series on Atmospheric. *Oceanic and Planetary Physics, 2*.

Santer, B. D., Wigley, T. L., Boyle, J. S., Gaffen, D. J., Hnilo, J. J., Nychka, D., . . . Taylor, K. E. (2000). Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *JGR, 105*, 7337-7356.

Sarra, A. d., Cacciani, M., Girolamo, P. D., Fiocco, G., Fuà, D., Knudsen, B., . . . Joergensen, T. S. (1992). Observations of correlated behavior of stratospheric ozone and aerosol at Thule during winter 1991-1992. *Geophysical Research Letters, 19*, 1823-1826.

Sherwood, S. C., & Nishant, N. (n.d.).

Sherwood, S. C., & Nishant, N. (2015). Atmospheric changes through 2012 as shown by iteratively homogenised radiosonde temperature and wind data (IUKv2). *Env. Res. Lett., 10*, 054007.

Sherwood, S. C., Meyer, C. L., Allen, R. J., & Titchner, H. A. (2008). Robust Tropospheric Warming Revealed by Iteratively Homogenized Radiosonde Data. *Journal of Climate, 21*, 5336-5352.

Siegel, A. F., & Benson, R. H. (1982). A Robust Comparison of Biological Shapes. *Biometrics, 38*(2), 341-350.

Smit, H. J., & ASOPOS panel. (2014). ASOPOS panel (Assessment of Standard Operating Procedures for Ozonesondes): Quality assurance and quality control for ozonesonde measurements in GAW. *World Meteorological Organization, GAW Report #201*.

Smit, H. J., & O3S-DQA panel. (2012). O3S-DQA panel (Ozone Sonde Data Quality Assessment): Guidelines for homogenization of ozonesonde data, SI2N/O3S-DQA activity as part of "Past changes in the vertical distribution of ozone assessment".

Smit, H. J., Straeter, W., Johnson, B., Oltmans, S., Davies, J., & Tarasick, D. W. (2007). Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Juelich Ozone Sonde Intercomparison Experiment (JOSIE). *Journal of Geophysical Research, 112*, D19306.

Solomon, S., Garcia, R. R., Rowland, F. S., & Wuebbles, D. J. (1986). On the depletion of Antarctic ozone. *Nature, 321*(9), 755–758.

Stauffer, R. M., A. M., T., D. E., K., D. W., T., R., V. M., H. G. J., S., . . . M. M., Y. (2022). An Examination of the Recent Stability of Ozonesonde Global Network Data. *AGU, 9*, e2022EA002459.

Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., . . . Tummon, F. (2017). An update on ozone profile trends for the period 2000 to 2016. *Atmos. Chem. Phys., 17*, 10675-10690.

Sterling, C. W., Johnson, B. J., Oltmans, S. J., Smit, H. J., Jordan, A. F., Cullis, P. D., . . . Witte, J. C. (2018). Homogenizing and estimating the uncertainty in NOAA's long-term vertical ozone profile records measured with the electrochemical concentration cell ozonesonde. *AMT, 11*, 3661–3687.

Stolarski, R. A., & Cicerone, R. J. (1974). Stratospheric chlorine: A possible sink for ozone. *Can. J. Chem., 52*(4), 1610–1615.

Studer, S. e. (2013). Intercomparison of stratospheric ozone profiles for the assessment of the upgraded GROMOS radiometer at Bern. *Atmos. Meas. Tech., 6*, 6097–6146.

Studer, S. e. (2014). A climatology of the diurnal variations in stratospheric and mesospheric ozone over Bern, Switzerland. *Atmos. Chem. Phys., 14*, 5905–5919.

Tarasick, D. W., Davies, J., Smit, H. J., & Oltmans, S. J. (2016). A re-evaluated Canadian ozonesonde record: Measurements of the vertical distribution of ozone over Canada from 1966 to 2013. *Atmospheric Measurement Techniques, 9*, 195–214.

Tarasick, D. W., Smit, H. J., Thompson, A. M., Morris, G. A., Witte, J. C., Davies, J., . . . Vömel, H. (2019). Improving ECC Ozonesonde Data Quality: Assessment of Current Methods and Outstanding Issues. *Earth and Space Science, 8*.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae, 12*.

Thompson, A., & al., e. (2019). Ozonesonde Quality Assurance: The JOSIE–SHADOZ (2017) Experience. *Bulletin of the American Meteorological Society, 100*, 155–171.

Thompson, A., Witte, J., Sterling, C., Jordan, A., Johnson, B., Oltmans, S., . . . al., e. (2017). First Reprocessing of Southern Hemisphere Additional Ozonesondes (SHADOZ) Ozone Profiles (1998–2016): 2. Comparisons With Satellites and Ground-Based Instruments. *JGR Atmosphere, 122*, 13,000-13,025.

Thorne, P. W., & Brohan, P. (2011). A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. *Journal of Geophysical Research - Atmospheres*.

Thorne, P. W., & Parker, D. E. (2005). Uncertainties in climate trends - Lessons from upper-air temperature records. *Bulletin of the American Meteorological Society, 10*, 1437-+.

Thorne, P. W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., . . . Pappalardo, G. (2017). Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment. *Geosci. Instrum. Method. Data Syst., 6*, 453-472.

Van Malderen, R., Allaart, M. F., De Backer, H., Smit, H. J., & De Muer, D. (2016). On instrumental errors and related correction strategies of ozonesondes: Possible effect on calculated ozone trends for the nearby sites Uccle and De Bilt. *Atmospheric Measurements Techniques, 9*, 3793–3816.

Vömel, H., David, D. E., & Smith, K. (2007). Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations. *JGR, 112*.

Wang, R. (2013). GOZCARDS merged data for ozone monthly zonal means on a geodetic latitude and pressure grid v1.01. *NASA Goddard Earth Sciences Data and Information Services Center*.

Waugh, D. W., Sobel, A. H., & Polvani, L. M. (2017). What Is the Polar Vortex and How Does It Influence Weather? *AMS, 98*, 37–44.

Weatherhead, E. C., Bodeker, G. E., Fassò, A., Chang, K. L., Lazo, J. K., Clack, C. M., . . . Yorgun, S. (2017). Spatial Coverage of Monitoring Networks: A Climate Observing System Simulation Experiment. *Journal of Applied Meteorology and Climatology, 56*, 3211–3228.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Jackman, C. H., Bishop, L., Hollandsworth Frith, S. M., . . . Frederick, J. E. (2000). Detecting the recovery of total column ozone. *J. Geophys. Res., 105*, 22201–22210.

Witte, J. C., Thompson, A. M., Smit, H. J., Fujiwara, M., Posny, F., Coetzee, G. R., . . . da Silva, F. R. (2017). First reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) profile records (1998–2015): 1. Methodology and evaluation. *JGR, 122*, 6611-6636.

Witte, J. C., Thompson, A. M., Smit, H. J., Vömel, H., Posny, F., & Stübi, R. (2018). First Reprocessing of Southern Hemisphere ADditional OZonesondes Profile Records: 3. Uncertainty in Ozone Profile and Total Column. *JGR, 123*, 3243-3268.

WMO. (2007). Scientific Assessment of Ozone Depletion: 2006. *Global Ozone Research and Monitoring Project—Report No. 50*.

WMO. (2014). Scientific Assessment of Ozone Depletion: 2014. *Global Ozone Research and Monitoring Project-Report No. 55*.

Wong, R. W., Schneider, C., & Mielke Jr, P. W. (1989). Geometric consistency for regression model estimation and testing in climatology and meteorology. *Atmosphere-Ocean, 27*, 508-520.

Yamazaki, K., Nakamura, T., Ukita, J., & Hoshi, K. (2020). A tropospheric pathway of the stratospheric quasi-biennial oscillation (QBO) impact on the boreal winter polar vortex. *ACP, 20*, 5111–5127.

Yan, X., Wright, J. S., Zheng, X., Livesey, N., Vömel, H., & Zhou, X. (2016). Validation of Aura MLS retrievals of temperature, water vapour and ozone in the upper troposphere and lower–middle stratosphere over the Tibetan Plateau during boreal summer. *AMT, 9*, 3547–3566.

Zhou, C., Wang, J., Dai, A., & Thorne, P. W. (2021). A New Approach to Homogenize Global Subdaily Radiosonde Temperature Data from 1958 to 2018. *Journal of Climate, 34*, 1163–1183.