



# On sampling uncertainty of satellite ozone profile measurements

V. F. Sofieva<sup>1</sup>, N. Kalakoski<sup>1</sup>, S.-M. Päivärinta<sup>1</sup>, J. Tamminen<sup>1</sup>, M. Laine<sup>1</sup>, and L. Froidevaux<sup>2</sup>

<sup>1</sup>Finnish Meteorological Institute, Helsinki, Finland

<sup>2</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

Correspondence to: V. F. Sofieva (viktorija.sofieva@fmi.fi)

Received: 7 February 2014 – Published in Atmos. Meas. Tech. Discuss.: 11 March 2014

Revised: 15 May 2014 – Accepted: 16 May 2014 – Published: 26 June 2014

**Abstract.** Satellite measurements sample continuous fields of atmospheric constituents at discrete locations and times. However, insufficient or inhomogeneous sampling, if not taken into account, can result in inaccurate average estimates and even induce spurious features. We propose to characterize the spatiotemporal inhomogeneity of atmospheric measurements by a measure, which is a linear combination of the asymmetry and entropy of a sampling distribution. It is shown that this measure is related to the so-called sampling uncertainty, which occurs due to non-uniform sampling patterns.

We have estimated the sampling uncertainty of zonal mean ozone profiles for six limb-viewing satellite instruments participating in the European Space Agency Ozone Climate Change Initiative project using the high-resolution ozone field simulated with the FinROSE chemistry-transport model. It is shown that the sampling uncertainty for the instruments with coarse sampling is not negligible and can be as large as a few percent. It is found that the standard deviation of the sampling uncertainty in the monthly zonal mean data allows for a simple parameterization in terms of the product of the standard deviation of natural variations and the proposed inhomogeneity measure. The sampling uncertainty estimates improve the uncertainty quantification and can be used in comprehensive data analyses.

The focus of this work is the vertical ozone distributions measured by limb-viewing satellite instruments, but the developed methods can also be applied to different satellite, ground-based and in situ measurements.

## 1 Introduction

Satellite data usually have very good global coverage, and they are therefore attractive for various analyses of spatiotemporal distributions of atmospheric parameters. However, measurements sample the continuous fields of atmospheric constituents, thermal structure and dynamic parameters only at discrete locations and times. Insufficient or inhomogeneous sampling (if not properly analyzed) can result in inaccurate average estimates and even induce spurious features.

The importance of sampling uncertainties is increasingly recognized. For example, to create temperature climatologies from radio-occultation measurements using different satellites, the error due to orbital sampling has been estimated using the high-resolution temperature field from ECMWF (European Centre for Medium-Range Weather Forecasts) and corrected in a deterministic way (Foelsche et al., 2011). Foelsche et al. (2011) estimated the sampling error by comparing climatologies derived from vertical ECMWF profiles at the radio-occultation times and locations with climatologies derived from the complete 4-D ECMWF field. This is the typical approach for estimating sampling uncertainties. Other recent works by Guan et al. (2013) and Aghedo et al. (2011) evaluate the impact of sampling patterns on satellite and climate model inter-comparisons, for several atmospheric parameters and for several satellites measuring in nadir-viewing geometry. The sampling uncertainty was found to be insignificant for the majority of variables, satellite instruments and types of averaging. Satellite measurements of atmospheric composition in a limb-viewing geometry have significantly coarser sampling than measurements from nadir-viewing satellite measurements. Recently, Toohey et al. (2013)

estimated the sampling uncertainty (referred to as “sampling bias”) of monthly zonal mean profiles in 5° latitude zones for the satellite instruments participating in the SPARC (Stratospheric Processes and their Role in Climate) Data Initiative project (Tegtmeier et al., 2013). As shown in this paper, stratospheric trace gas climatologies may contain significant sampling uncertainties (up to 20% in some regions for instruments with coarse sampling). The sampling uncertainty is found to be greatest when natural variability is strongest, as intuitively expected. Non-uniform sampling results not only in inaccurate averaged values, but it also affects the uncertainty estimates of the mean value, typically characterized by the standard error of the mean (Toohey and von Clarmann, 2013). A special sampling inhomogeneity may appear also in comparisons of data measured by different methods, e.g., in comparisons of satellite and ground-based observations.

The focus of our study is the vertical distribution of ozone measured by satellite instruments using the limb-viewing geometry. During our analyses of ozone distributions, we encountered some puzzling features, which turned out to be induced by non-uniform sampling patterns. We show examples of such spurious features, in order to attract the attention to the importance of sampling uncertainties in geophysical analyses. This is the first objective of our paper.

The second objective of our paper is characterizing the effect of orbital sampling on zonal monthly mean profiles by six limb and occultation satellite instruments participating in European Space Agency (ESA) Climate Change Initiative project (Ozone\_cci, <http://www.esa-ozone-cci.org>). Within this project, which is aimed at creating homogenized and merged ozone profile data sets from ESA Envisat satellite data and from ESA Third Party Missions, monthly zonal mean data sets in 10° latitude zones from each satellite instrument have been created. Analogously to the approaches used in previous studies, we have performed simulations by the chemistry-transport model (CTM) FinROSE for evaluating sampling errors in monthly zonal mean data sets. We found that the sampling error for several satellite instruments is non-negligible; it depends on the sampling pattern, the natural ozone variability, and the ozone distribution (the results are presented in Sect. 5). If a very good chemistry-transport model existed, it would allow for a correction of the sampling uncertainty (actually, sampling bias) in monthly zonal mean data in a deterministic way, analogous to the correction of sampling bias in radio occultation temperature data using ECMWF temperature fields (Foelsche et al., 2011). However, existing chemistry-transport models do not have sufficient accuracy for this purpose. Note also that such an approach cannot completely remove sampling biases. Therefore, there is a need for a robust and simple measure that can characterize the inhomogeneity of distributions and, at the same time, is related to the sampling uncertainty. Our other objectives are the creation of such an inhomogeneity measure and the

statistical characterization/parameterization of the sampling uncertainty.

This paper is organized as follows. In Sect. 2, we briefly describe the satellite data used in our analyses and the chemistry-transport model FinROSE. Examples of spurious ozone distribution features induced by non-uniform sampling are provided in Sect. 3. In Sect. 4, we propose an inhomogeneity measure for the characterization of sampling patterns. Section 5 is dedicated to the characterization of the sampling uncertainty in climatological ozone distributions and its parameterization. A summary section concludes the paper.

## 2 Data and models

### 2.1 Satellite ozone profile data

In this section, we briefly introduce the satellite ozone data that are used in our analyses and illustrations. We present here only the information that is important for studying the sampling uncertainty. More information about the satellite data can be found in the cited publications.

#### 2.1.1 Harmonized data set of ozone profiles

The HARMOnized data set of Ozone profiles (HARMOZ) has been created in the framework of the ESA Ozone\_cci project (Sofieva et al., 2013). It consists of user-friendly level-2 ozone profiles from six satellite instruments: Global Ozone Monitoring by Occultation of Stars (GOMOS), Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) and Scanning Imaging Spectrometer for Atmospheric Chartography (SCIAMACHY) on board Envisat, Optical Spectrograph and InfraRed Imaging System (OSIRIS) and Sub-Millimeter Radiometer (SMR) on board Odin, and Atmospheric Chemistry Experiment – Fourier Transform Spectrometer (ACE-FTS) on board SCISAT-1. Compared to the original Level 2 data, the HARMOZ data are screened for invalid values, presented on the same pressure grid and in the same format. More details on the HARMOZ data set can be found in Sofieva et al. (2013). The data are available online at <http://www.esa-ozone-cci.org/?q=node/161> or at [dx.doi.org/10.5270/esa-ozone\\_cci-limb\\_occultation\\_profiles-2001\\_2012-v\\_1-201308](http://dx.doi.org/10.5270/esa-ozone_cci-limb_occultation_profiles-2001_2012-v_1-201308).

Table 1 contains the parameters of the individual HARMOZ data sets that are important for sampling studies. MIPAS and SMR measure during day and night, while measurements from other instruments are obtained at specific local times. The best sampling is obtained by MIPAS and SCIAMACHY; ACE-FTS has the coarsest sampling.

#### 2.1.2 Microwave Limb Sounder

The Microwave Limb Sounder (MLS) on board the EOS Aura satellite has been measuring thermal microwave

**Table 1.** General information about the HARMOZ data sets.

	Vertical range in HARMOZ	Local time	Vertical resolution	Average number of profiles per day
GOMOS	250–1 × 10 <sup>-4</sup> hPa	~ 10 p.m.	2 km below 30 km, 3 km above 40 km, a linear transition between	~ 110
MIPAS	400–5 × 10 <sup>-2</sup> hPa	~ 10 p.m. and ~ 10 a.m.	profile-dependent, 3–5 km	~ 1000
SCIAMACHY	250–5 × 10 <sup>-2</sup> hPa	~ 10 a.m.	profile-dependent, 3–5 km	~ 1300
OSIRIS	450–1 × 10 <sup>-1</sup> hPa	~ 6 a.m. and ~ 6 p.m.	~ 2–3 km, altitude-dependent	~ 250
SMR	300–5 × 10 <sup>-2</sup> hPa	~ 6 a.m. and ~ 6 p.m.	profile-dependent, ~ 2.5–3.5 km	~ 250
ACE-FTS	450–2 × 10 <sup>-4</sup> hPa	sunrise and sunset	~ 3 km	~ 11

emission from the atmospheric limb in five spectral bands from 115 GHz to 2.5 THz from August 2004 to present (Waters et al., 2006, <http://mls.jpl.nasa.gov/index-eos-mls.php>). Aura is in a sun-synchronous orbit at about 705 km altitude with 98° inclination, a 98.8 min period, and an ascending equatorial crossing time of ~ 13:45 LT. MLS measures during day and night and provides near-global coverage from 82° S to 82° N with about 3500 limb scans daily. MLS v3.3 ozone profiles measured with a vertical resolution of ~ 3 km are used in our analyses. Differences between the v3.3 and v2.2 ozone data are typically less than a few percent, except in the upper troposphere and lower stratosphere, especially in the tropics, where the (finer resolution) v3.3 profiles exhibit larger vertical oscillations than the v2.2 profiles (Froidevaux et al., 2008; Livesey et al., 2013a, b).

## 2.2 Simulations with FinROSE chemistry-transport model

FinROSE is a global three-dimensional offline chemistry-transport model (CTM) developed for middle atmosphere studies (Damski et al., 2007). The model dynamics (i.e., temperature, horizontal winds and surface pressure) is taken from external sources. In this study FinROSE was run with the ECMWF ERA-Interim data, with 30 min time step, with 32 vertical levels from the surface up to 0.1 hPa (~ 65 km) and with a horizontal resolution of 3° longitude by 1.5° latitude.

FinROSE uses tropospheric abundances of chemical composition as boundary conditions. At the lower boundary, monthly averages are used for ozone and water vapor, and trends are imposed for long-lived gases. In the stratosphere FinROSE produces distributions of 40 species and families taking into account both chemistry and dynamics (only the long-lived constituents are transported). The model includes about 120 homogeneous reactions and 30 photodissociation processes. Chemical kinetic data, reaction rate coefficients and absorption cross sections are based on the Jet Propulsion Laboratory compilation by Sander et al. (2006), including updates from the available supplements. The

model photodissociation rates are calculated using a radiative transfer model PHODIS (Kylling et al., 1997). In addition, 30 heterogeneous reactions on/in liquid binary aerosols and type Ia, Ib and II polar stratospheric clouds are included in the model.

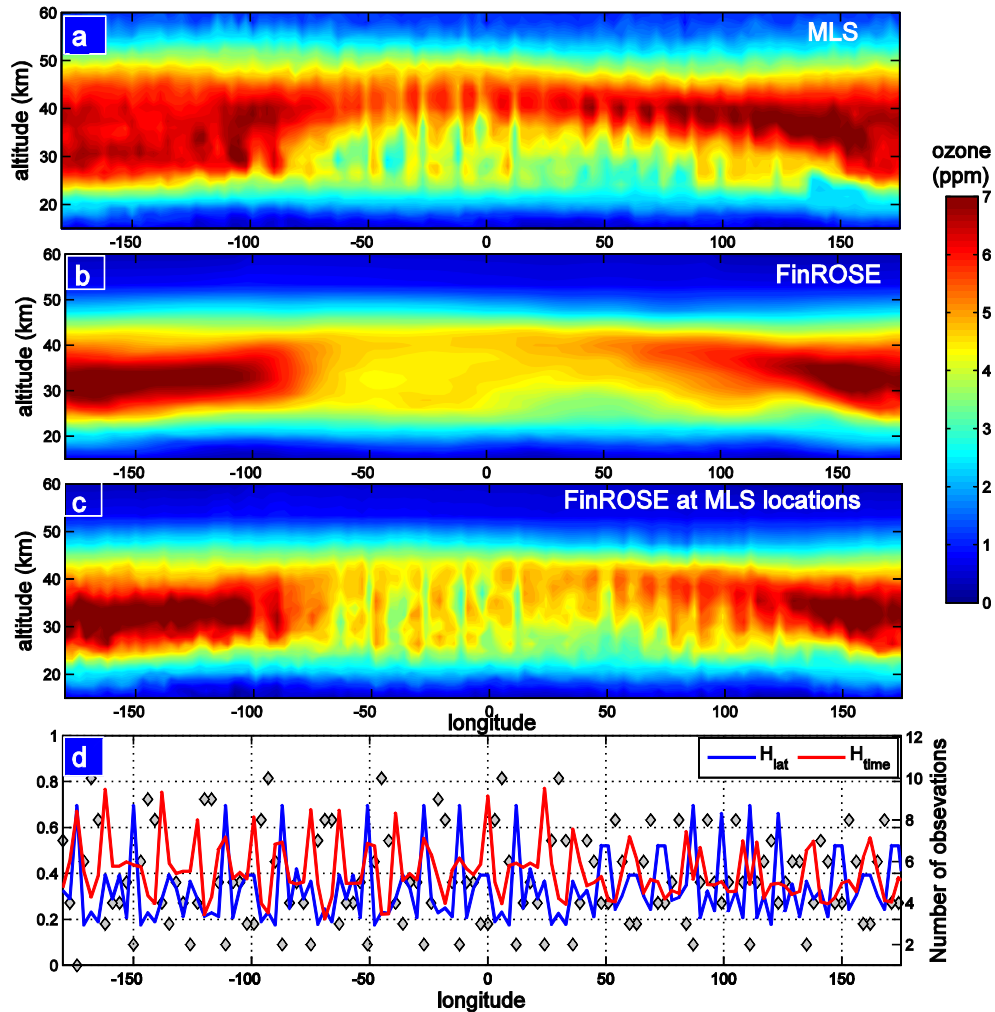
In this study, we compared the FinROSE ozone fields with the model data sub-sampled at locations and times corresponding to the satellite measurements. The spatiotemporal resolution of FinROSE is sufficient to capture most of the ozone variability, and it is therefore well suited for this study. Furthermore, the model's spatial resolution is similar to the effective resolution of the ozone profiles measured in the limb-viewing geometry. Small-scale ozone variability (e.g., perturbations due to gravity waves) is not considered in our analysis, neither in the model nor in the data.

## 3 Examples of spurious features induced by sampling patterns

In this section, we present examples of spurious features induced by inhomogeneous sampling patterns. We selected the examples where these features are clearly seen, and therefore they are “extreme” in the sense that they are related to highly non-uniform distributions. However, these examples are not artificially constructed, as they were encountered during our analyses of satellite data sets.

### 3.1 Quasi-periodic structures

During our analyses of longitudinal distributions of ozone using MLS observations, we noticed quasi-periodic oscillatory patterns when the data were averaged in small longitudinal bins. An example of such oscillations is shown in Fig. 1a. In this example, MLS ozone profiles at latitudes 60–70° N measured during three days, 14–16 October 2007, are averaged (median estimates) in 3° longitude bins (the bins are chosen to be small for a better view of the effect). The MLS measurements are located at the edge of the forming polar vortex, where ozone gradients are strong, even in a relatively small latitude zone 60–70° N



**Figure 1.** (a) Ozone mixing ratio from MLS observations at 60–70° N on 14–16 October 2007, averaged in 3° longitude bins; (b) the mean FinROSE field for 14–16 October 2007 at latitudes 60–70° N; (c) the FinROSE ozone field sampled at MLS locations, for the same dates 14–16 October 2007 and averaged in the same way as the MLS data; and (d) number of observations (grey diamonds) and inhomogeneity measures in latitude  $H_{lat}$  and in time  $H_{time}$ , for each longitude bin.

(examples of ozone distributions can be found for example at [http://mls.jpl.nasa.gov/plots/mls/mls\\_plot\\_locator.php](http://mls.jpl.nasa.gov/plots/mls/mls_plot_locator.php)). The observed quasi-periodic structures are not real; they are induced by the MLS sampling pattern in the presence of strong ozone gradients. Indeed, the ozone field from FinROSE simulations, being averaged over the considered period 14–16 October 2007, is smooth (Fig. 1b). However, if the FinROSE ozone field is sampled at the MLS locations and averaged in the same way as the MLS data, the longitudinal distribution has a very similar quasi-periodic pattern (Fig. 1c).

The reason for the appearance of such oscillations is that some of the longitudinal bins have a very small number of measurements (1 or 2), as shown in Fig. 1d. The distribution of measurements is very inhomogeneous for such small bins (the inhomogeneity measure will be discussed below

in Sect. 4). This inhomogeneity in combination with strong ozone gradients results in the observed spurious oscillations in these MLS ozone distributions. As expected, the oscillations disappear if larger longitude bins are used. We note that such spurious oscillatory patterns would also be observed with other satellite data sets (e.g., MIPAS or SCIAMACHY), and this sampling effect is not specific only to MLS.

### 3.2 Representation of annual cycles

For analyses of ozone time series and trends, monthly zonal mean data are usually used (e.g., Kyrölä et al., 2013; Newchurch et al., 2003; Randel and Wu, 2007; Staehelin et al., 2001). Sometimes (if the latitude zone is rather wide), the representation of the annual ozone cycle can differ significantly between various satellite measurements.

Such an example is presented in Fig. 2 (top panel), which shows monthly zonal mean ozone number densities at 15 hPa ( $\sim 30$  km) averaged in the latitude zone 40–60° N, for GOMOS and MIPAS. In summer, GOMOS and MIPAS data are rather close to each other, while in winter GOMOS data indicate significantly lower ozone values than MIPAS. This is seen especially clear in years 2005–2008; the relative difference  $\frac{\text{GOMOS}-\text{MIPAS}}{\text{GOMOS}+\text{MIPAS}} \cdot 200\%$  estimated from the monthly zonal mean data is shown by black lines in Fig. 2(bottom).

This difference in representation of the seasonal cycle is largely induced by inhomogeneous sampling. Indeed, when using only collocated data in this latitude zone, such seasonal cycle in difference between GOMOS and MIPAS ozone is not observed (Fig. 2, bottom). In our analysis, we used two collocation criteria: “strict”, based on the effective horizontal resolution of the measurements (distance  $d \leq 400$  km, time separation  $\Delta t \leq 4$  h), and a more relaxed (distance  $d \leq 1000$  km, time separation  $\Delta t \leq 24$  h, latitude difference  $|\Delta\theta| \leq 2^\circ$ ). These collocation criteria were also used while creating the bias tables for the HARMOZ data sets (Sofieva et al., 2013). As shown in Fig. 2 (bottom), the difference between collocated GOMOS and MIPAS ozone data at this altitude is nearly constant,  $\sim 4\%$ , with small variations. Therefore, the different representation of ozone annual cycles in this example is a feature induced by non-uniform sampling. GOMOS data do not provide a uniform sampling in this relatively wide latitude zone; thus the annual cycle is represented differently in MIPAS and GOMOS data. If smaller latitude zones are used, this spurious feature practically vanishes (not shown here).

#### 4 Measure of inhomogeneity

We will characterize the inhomogeneity of a sampling distribution by the linear combination of two classical inhomogeneity measures, asymmetry  $A$  and entropy  $E$ :

$$H = \frac{1}{2}(A + (1 - E)). \quad (1)$$

Let  $x$  be a coordinate that characterizes positions of measurements. Asymmetry is defined as

$$A = 2 \frac{|\langle x \rangle - x_0|}{\Delta x}, \quad (2)$$

where  $\langle x \rangle$  is the mean location of measurements in a cell of the width  $\Delta x$  with the central point  $x_0$ . Asymmetry ranges from 0 for symmetric distributions to 1 for strongly asymmetric distributions. This asymmetry measure has been used, for example, for characterization of inhomogeneity of ozone monthly zonal mean data in the recent study by Kyrölä et al. (2013). The entropy definition considers that  $\Delta x$  is divided into smaller bins. Then the entropy can be defined as (e.g., Shannon and Weaver, 1949)

$$E = \frac{-1}{\log(N)} \sum_i \frac{n(i)}{n_0} \log \left( \frac{n(i)}{n_0} \right), \quad (3)$$

where  $N$  is the number of bins,  $n(i)$  is the number of observations in the bin  $i$ , and  $n_0$  is the sample size. If  $n(i) = 0$  for some bin, the corresponding term in Eq. (3) contributes with its limit value, zero. The size of the bins for evaluating entropy should be selected according to the variability of a considered parameter: within a bin, the natural variations should be small. Perfectly homogeneous sampling patterns have  $E = 1$ .

Asymmetry or entropy separately does not always describe the inhomogeneity of distributions properly, as illustrated in Fig. 3 with an example of very simple artificial 1-D distributions in 10 bins. For example, the distributions in panels a and b of Fig. 3 are perfectly symmetric ( $A = 0$ ), but the distribution for Fig. 3b is significantly less homogeneous than that of Fig. 3a, as indicated by the different entropy (and thus inhomogeneity) values. The distributions in Fig. 3c and d have the same entropy but a different asymmetry. Our measure  $H$  characterizes the distribution for Fig. 3c as more homogeneous (smaller  $H$ ), as expected from visual inspection.

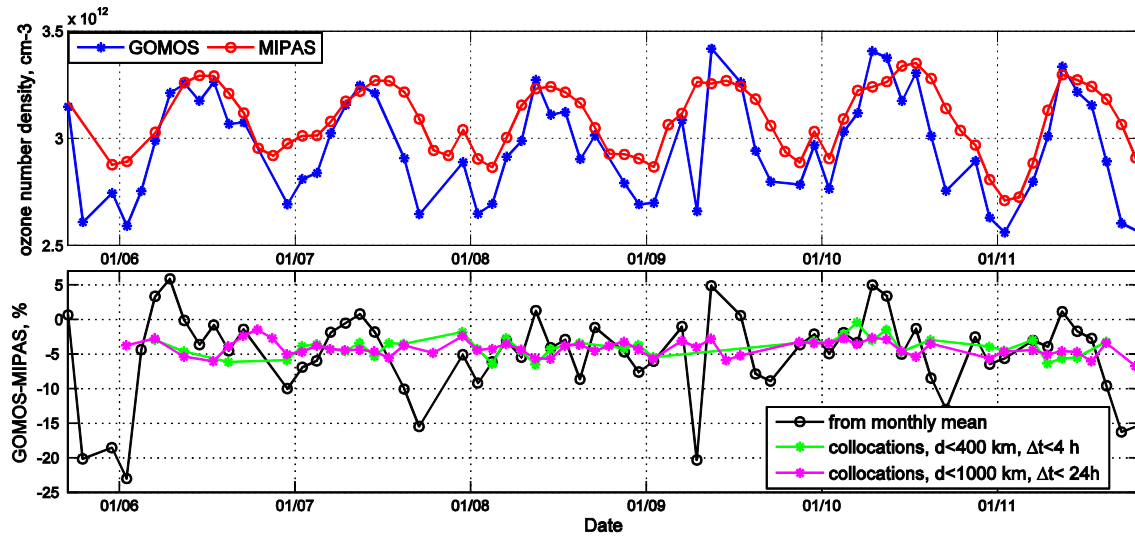
The linear combination,  $H$ , of asymmetry and entropy characterizes the inhomogeneity better than each component separately. Since asymmetry and entropy are conceptually different measures, we take  $A$  and  $(1 - E)$  with equal weights in Eq. (1). The inhomogeneity measure  $H$  ranges from 0 to 1: the larger  $H$  is, the less homogeneous the distribution is. In case of a single data point, its central location has the smallest inhomogeneity (Fig. 3e and f). However, in the case of several measurements, the locations close to the center have the same weight as others: the distributions shown in Fig. 3g and h have the same asymmetry and entropy values, thus the same inhomogeneity.

The spurious oscillations in MLS longitude–altitude distributions, which were discussed in Sect. 3.1 and shown in Fig. 1a, are well explained by the proposed inhomogeneity measure. Figure 1d shows the inhomogeneity measures in latitude  $H_{\text{lat}}$  and in time  $H_{\text{time}}$ , for each longitude bin. The same periods are observed in the inhomogeneity values, with peaks corresponding to small numbers of measurements.

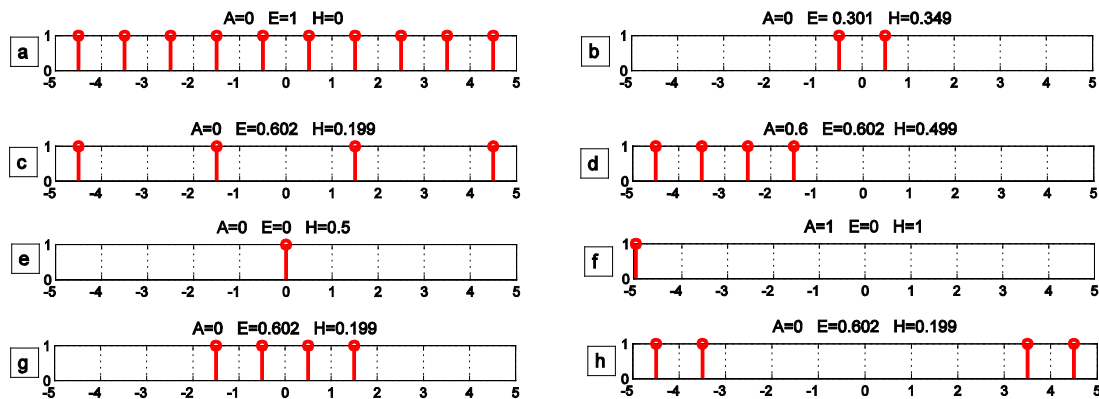
The simple inhomogeneity measure (Eq. 1) is one-dimensional. It could also be constructed for two- or multi-dimensional cases by combining the corresponding 1-D inhomogeneity measures. For example, a two-dimensional inhomogeneity measure can be constructed as a weighted mean  $H = w_1 H_1 + w_2 H_2$ , where  $w_1 + w_2 = 1$ . Then the range for  $H$  will be also from 0 to 1. This is discussed further in Sect. 5.

#### 5 Estimating sampling uncertainty in climatological data

In order to estimate sampling uncertainties in monthly zonal mean data for the HARMOZ instruments, we have performed simulations with the FinROSE chemistry-transport model for



**Figure 2.** Top: monthly zonal mean ozone number density at 15 hPa ( $\sim 30$  km), latitudes  $40\text{--}60^\circ$  N. Bottom: relative difference of GOMOS and MIPAS ozone at 15 hPa, for the same latitude zone  $40\text{--}60^\circ$  N estimated using monthly zonal mean data (black line) and using collocated GOMOS-MIPAS data (green and magenta lines).

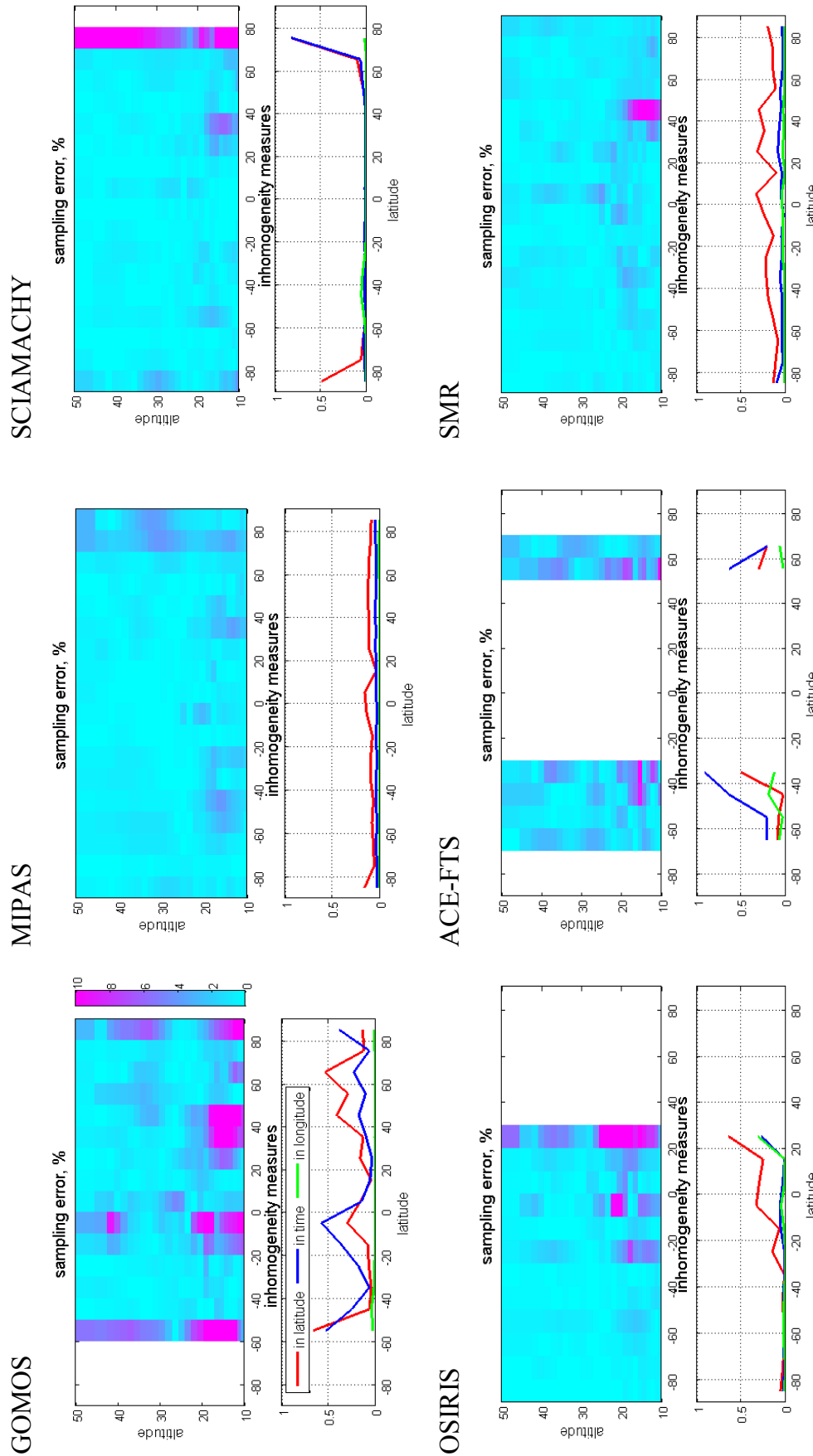


**Figure 3.** Examples of artificial distributions and their characterization by asymmetry  $A$ , entropy  $E$ , and inhomogeneity measure  $H$ , Eq. (1).

years 2005–2009. The ozone field was simulated with a high resolution of  $1.5^\circ$  in latitude and  $3^\circ$  in longitude. Then the monthly mean model data in  $10^\circ$  latitude bins from  $90^\circ$  S to  $90^\circ$  N were compared with those sampled according to the locations of the satellite measurements. An example of the absolute difference  $|\Delta| = |\langle x_{\text{FinROSE}} \rangle - \langle x_{\text{sat}} \rangle|$  between the full  $\langle x_{\text{FinROSE}} \rangle$  and sub-sampled  $\langle x_{\text{sat}} \rangle$  monthly zonal mean ozone field for January 2008 is shown in Fig. 4, for the sampling patterns of each HARMOZ data set. Large sampling errors are strongly correlated with inhomogeneous distributions in certain latitude ranges, as indicated by the inhomogeneity measures in latitude, longitude and time shown in the lower panels of Fig. 4. The inhomogeneity is computed using the profile locations; therefore, less dense sampling at lower altitudes is not taken into account in this illustration. For instruments with dense sampling (MIPAS and SCIAMACHY), the sampling errors are small in

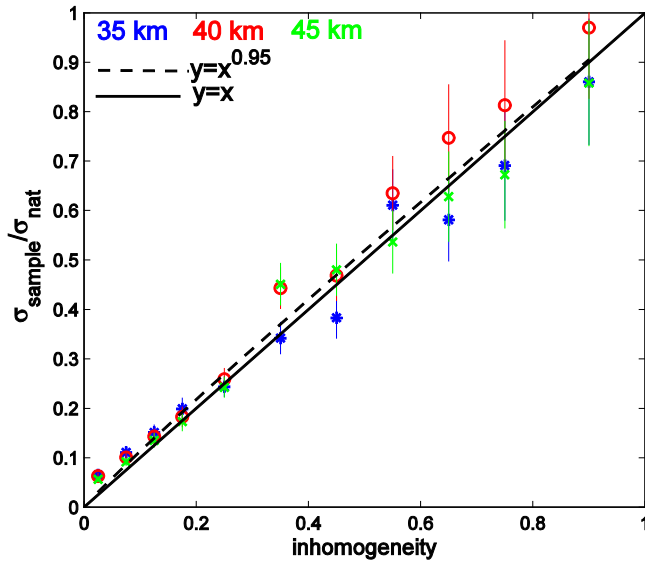
nearly all locations. For coarse samplers (GOMOS, OSIRIS, ACE-FTS), the sampling error is not negligible and can be as large as a few percent (sometimes, even up to  $\sim 20\%$ ). It can also be noticed that the sampling uncertainties are enhanced in the regions of high ozone variability (upper troposphere and lower stratosphere (UTLS), mid- and high northern latitudes), as expected. The sampling error magnitude is in perfect agreement with that reported by Toohey et al. (2013), who presented many illustrations of sampling bias statistics. In our paper, we will not show a similar characterization for each month and year, but we take a further step in the sampling uncertainty characterization via its parameterization, as described below.

The sampling error depends on the sampling pattern and the ozone distribution in a particular month (i.e., it depends on the spatiotemporal correlation of the ozone field). The absence of a sufficiently accurate CTM does not



**Figure 4.** Sampling errors estimated using FinROSE CTM (upper panels) and inhomogeneity measures (lower panels) in January 2008 for different instruments. Color scale is the same for all instruments.





**Figure 5.** The dependence of normalized sampling uncertainty  $\frac{\sigma_{\text{sample}}}{\sigma_{\text{nat}}}$  on the inhomogeneity measure  $H_{\text{tot}}$ , for 35 km (blue), 40 km (red) and 45 km (green); error bars correspond to  $1\sigma$  uncertainty. Black dashed line corresponds to the least-squares fit using the Eq. (4), i.e.,  $\sigma_{\text{sample}} = H_{\text{tot}}^{0.95} \cdot \sigma_{\text{nat}}$ . Black solid line corresponds to the relation  $\sigma_{\text{sample}} = H_{\text{tot}} \cdot \sigma_{\text{nat}}$ .

allow for a characterization and correction of the sampling error/bias in a deterministic way, as done by Foelsche et al. (2011). Our estimates of sampling uncertainty with the FinROSE model, as well as the analogous study by Tooney et al. (2013), clearly indicate enhancements of sampling uncertainty for coarse sampling patterns and for regions of high ozone variability. These (expected) findings prompt us to parameterize the sampling uncertainty as a function of the natural variability and the inhomogeneity measure. We propose to characterize the sampling uncertainty as an additional random error of monthly zonal mean data modeled as a Gaussian random variable with zero mean and standard deviation  $\sigma_{\text{sample}}$ . Our objective is to find a parameterization of  $\sigma_{\text{sample}}$ , which depends on the inhomogeneity measure  $H$  and the standard deviation of natural variations  $\sigma_{\text{nat}}$  in a spatiotemporal cell. Since  $\sigma_{\text{sample}}$  should approach zero when  $H \rightarrow 0$  or  $\sigma_{\text{nat}} \rightarrow 0$  and with dimensional analysis, it is reasonable to assume the following functional dependence:

$$\sigma_{\text{sample}} = \sigma_{\text{nat}} \cdot H^{\alpha}, \quad (4)$$

where  $\alpha$  is the parameter to be estimated.

In case of monthly zonal mean data, the inhomogeneity in longitude is very small for all HARMOZ instruments. Therefore, we define the combined (total) inhomogeneity of the considered satellite measurements as

$$H_{\text{tot}} = \frac{1}{2}(H_{\text{lat}} + H_{\text{time}}), \quad (5)$$

where  $H_{\text{lat}}$  and  $H_{\text{time}}$  are inhomogeneities in latitude and in time, respectively. Here we take  $H_{\text{lat}}$  and  $H_{\text{time}}$  with equal weights. In the future, when detailed information about the ozone spatiotemporal variability will be available, weighting of the one-dimensional inhomogeneities might be optimized. To find the sampling uncertainty parameterization, we used only the sampling patterns of coarse samplers GOMOS, OSIRIS and ACE-FTS, which provide a wide range of inhomogeneity values. For each of these three instruments, we computed the sampling error  $\Delta$  (i.e., the difference between the continuous and sub-sampled FinROSE ozone field averages) corresponding to each  $10^\circ$  latitude bin and each month in years 2005–2009. We divided the range of inhomogeneity values into intervals containing  $\sim 100$ – $200$  estimates of sampling error (the centers of the inhomogeneity bins are indicated by symbols in Fig. 5). Then we calculated the standard deviation of sampling errors normalized by ozone natural variability in the corresponding bins,  $\text{std}\left(\frac{\Delta}{\sigma_{\text{nat}}}\right)$ ,

which approximates the ratio  $\frac{\sigma_{\text{sample}}}{\sigma_{\text{nat}}}$ . The estimates of natural variability  $\sigma_{\text{nat}}$  for each month and each latitude bin are taken from the LLM climatology (McPeters et al., 2007). For stratospheric altitudes,  $\sigma_{\text{nat}}$  can also be estimated from the FinROSE fields (the results of sampling error parameterization will be the same). We considered three altitudes levels: 35 km, 40 km and 45 km, where the LLM climatology is represented by satellite data and thus can provide realistic estimates of ozone natural variability. Figure 5 shows the dependence of the estimates  $\frac{\sigma_{\text{sample}}}{\sigma_{\text{nat}}}$  as a function of  $H_{\text{tot}}$  for the three altitude levels. The experimental data shown in Fig. 5 are very close to the one-to-one line; the correlation coefficient between  $\frac{\sigma_{\text{sample}}}{\sigma_{\text{nat}}}$  and  $H_{\text{tot}}$  is very high,  $\sim 0.98$ . The non-linear least-squares fit of all experimental data points in Fig. 5 by the function in Eq. (4) gives the estimate of parameter  $\alpha = 0.95 \pm 0.02$ , which is very close to one. Therefore, we suggest parameterizing the sampling uncertainty with the following simple relation:

$$\sigma_{\text{sample}} = H_{\text{tot}} \cdot \sigma_{\text{nat}}. \quad (6)$$

This relation, which is intuitively expected, allows for a sufficiently accurate parameterization of the sampling uncertainty of monthly zonal mean ozone profiles. For the monthly zonal mean ozone data from the HARMOZ instruments, which are created in the framework of the Ozone\_cci project (<http://www.esa-ozone-cci.org/?q=node/166>), the profiles of inhomogeneity measure in latitude and in time are provided in the data files. Then the total uncertainty of the monthly zonal mean data  $\sigma_{\text{tot}}^2$  can be estimated as

$$\sigma_{\text{tot}}^2 = \sigma_{\text{mean}}^2 + \sigma_{\text{sample}}^2, \quad (7)$$

where  $\sigma_{\text{mean}}$  is the standard error of the mean, which can be computed as  $\sigma_{\text{mean}}^2 = \frac{s^2}{N}$  ( $s^2$  is the sample variance,  $N$  is the sample size), and  $\sigma_{\text{sample}}$  can be parameterized



by Eq. (5). The estimates of natural variability can be taken from a climatology, a reliable model or, in some applications, estimated directly from the data as  $\sigma_{\text{nat}}^2 = \text{var}(x) - \sigma_{\text{noise}}^2$ , where  $\text{var}(x)$  is the sample variance and  $\sigma_{\text{noise}}^2$  is the estimated data uncertainty variance. The estimates of natural variability from the data can be used only if the data distribution is homogeneous and if the uncertainty estimates are realistic.

## 6 Summary

The examples and discussions presented in our paper have shown that the sampling uncertainty of satellite instruments (especially those with coarse sampling) is not negligible and should be taken into account in the characterization of geophysical distributions. Sampling patterns might induce spurious features, and therefore it is worth accompanying geophysical analyses of distributions with an analysis of data representativeness.

We propose a characterization of the inhomogeneity of a distribution by a simple measure  $H$ , which is a linear combination of asymmetry and entropy.

For the instruments participating in the HARMOZ data set, we estimated the sampling error of monthly zonal mean ozone profiles using simulations with the FinROSE chemistry-transport model. It was demonstrated that the standard deviation of the sampling uncertainty  $\sigma_{\text{sample}}$  allows a simple parameterization as the product of natural variability  $\sigma_{\text{nat}}$  and the proposed inhomogeneity measure  $H$ :  $\sigma_{\text{sample}} = \sigma_{\text{nat}} \cdot H$ .

This parameterization of sampling uncertainty can be used in estimating the total uncertainty of the monthly zonal mean ozone profiles obtained from satellite observations. In particular, the sampling uncertainty can be taken into account in creating low-temporal-resolution combined data sets from different satellite measurements using ensemble-averaged estimates. Ideally, comparisons of distributions from models and measurements should take sampling uncertainties into account as well.

Although the focus of our analysis was the vertical distribution of ozone, the proposed inhomogeneity measure and the methods for parameterization of sampling uncertainty can be equally applied to the analyses of other (not only satellite) measurements.

*Acknowledgements.* The work of the FMI team has been supported by the ESA Ozone Climate Change Initiative project and the Academy of Finland (projects MIDAT, ASTREX and INQUIRE).

Edited by: R. Eckman

## References

- Aghedo, A. M., Bowman, K. W., Shindell, D. T., and Faluvegi, G.: The impact of orbital sampling, monthly averaging and vertical resolution on climate chemistry model evaluation with satellite observations, *Atmos. Chem. Phys.*, 11, 6493–6514, doi:10.5194/acp-11-6493-2011, 2011.
- Damski, J., Thölix, L., Backman, L., Kaurola, J., Taalas, P., Austin, J., Butchart, N., and Kulmala, M.: A chemistry-transport model simulation of middle atmospheric ozone from 1980 to 2019 using coupled chemistry GCM winds and temperatures, *Atmos. Chem. Phys.*, 7, 2165–2181, doi:10.5194/acp-7-2165-2007, 2007.
- Foelsche, U., Scherllin-Pirscher, B., Ladstädter, F., Steiner, A. K., and Kirchengast, G.: Refractivity and temperature climate records from multiple radio occultation satellites consistent within 0.05 %, *Atmos. Meas. Tech.*, 4, 2007–2018, doi:10.5194/amt-4-2007-2011, 2011.
- Froidevaux, L., Jiang, Y. B., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Browell, E. V., Hair, J. W., Avery, M. A., McGee, T. J., Twigg, L. W., Sunmicht, G. K., Jucks, K. W., Margitan, J. J., Sen, B., Stachnik, R. A., Toon, G. C., Bernath, P. F., Boone, C. D., Walker, K. a., Filipiak, M. J., Harwood, R. S., Fuller, R. A., Manney, G. L., Schwartz, M. J., Daffer, W. H., Drouin, B. J., Cofield, R. E., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., and Wagner, P. A.: Validation of Aura Microwave Limb Sounder stratospheric ozone measurements, *J. Geophys. Res.*, 113, D15S20, doi:10.1029/2007JD008771, 2008.
- Guan, B., Waliser, D. E., Li, J.-L. F., and da Silva, A.: Evaluating the impact of orbital sampling on satellite–climate model comparisons, *J. Geophys. Res. Atmos.*, 118, 1–15, doi:10.1029/2012JD018590, 2013.
- Kylling, A., Albold, A., and Seckmeyer, G.: Transmittance of a cloud is wavelength-dependent in the UV-range: Physical interpretation, *Geophys. Res. Lett.*, 24, 397–400, doi:10.1029/97GL00111, 1997.
- Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päivärinta, S.-M., Tukiainen, S., Zawodny, J., and Thomason, L.: Combined SAGE II-GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone, *Atmos. Chem. Phys.*, 13, 10645–10658, doi:10.5194/acp-13-10645-2013, 2013.
- Livesey, N. J., Logan, J. A., Santee, M. L., Waters, J. W., Doherty, R. M., Read, W. G., Froidevaux, L., and Jiang, J. H.: Interrelated variations of O<sub>3</sub>, CO and deep convection in the tropical/subtropical upper troposphere observed by the Aura Microwave Limb Sounder (MLS) during 2004–2011, *Atmos. Chem. Phys.*, 13, 579–598, doi:10.5194/acp-13-579-2013, 2013a.
- Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu, D. L.: EOS MLS Version 3.3 and 3.4 Level 2 data quality and description document, available at: [http://mls.jpl.nasa.gov/data/v3\\_data\\_quality\\_document.pdf](http://mls.jpl.nasa.gov/data/v3_data_quality_document.pdf) (last access: 24 June 2014), 2013b.
- McPeters, R. D., Labow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, *J. Geophys. Res.*, 112, D05308, doi:10.1029/2005JD006823, 2007.

- Newchurch, M. J., Yang, E.-S., Cunnold, D. M., Reinsel, G. C., Zawodny, J. M., and Russell, J. M.: Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery, *J. Geophys. Res. Atmos.*, 108, 4507, doi:10.1029/2003JD003471, 2003.
- Randel, W. J. and Wu, F.: A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, *J. Geophys. Res. Atmos.*, 112, D06313, doi:10.1029/2006JD007339, 2007.
- Sander, S. P., Golden, D. M., Kurylo, M. J., Moortgat, G. K., Wine, P. H., Ravishankara, A. R., Kolb, C. E., Molina, M. J., Finlayson-Pitts, B. J., Huie, R. E., and Orkin, V. L.: Chemical kinetics and photochemical data for use in Atmospheric Studies Evaluation Number 15, JPL, Pasadena, USA, available at: [http://jpldataeval.jpl.nasa.gov/previous\\_evaluations.html](http://jpldataeval.jpl.nasa.gov/previous_evaluations.html) (last access: 24 June 2014), 2006.
- Shannon, C. E. and Weaver, W.: The mathematical theory of communication, Univ. of Ill. Press, Urbana, 1949.
- Sofieva, V. F., Rahpoe, N., Tamminen, J., Kyrölä, E., Kalakoski, N., Weber, M., Rozanov, A., von Savigny, C., Laeng, A., von Clarmann, T., Stiller, G., Lossow, S., Degenstein, D., Bourassa, A., Adams, C., Roth, C., Lloyd, N., Bernath, P., Hargreaves, R. J., Urban, J., Murtagh, D., Hauchecorne, A., Dalaudier, F., van Roozendaal, M., Kalb, N., and Zehner, C.: Harmonized dataset of ozone profiles from satellite limb and occultation measurements, *Earth Syst. Sci. Data*, 5, 349–363, doi:10.5194/essd-5-349-2013, 2013.
- Staehelin, J., Harris, N. R. P., Appenzeller, C., and Eberhard, J.: Ozone trends: A review, *Rev. Geophys.*, 39, 231–290, doi:10.1029/1999RG000059, 2001.
- Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Lingenfelser, G., Lumpe, J., Nardi, B., Neu, J., Pendlebury, D., Remsberg, E., Rozanov, A., Smith, L., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, *J. Geophys. Res. Atmos.*, 118, 12229–12247, doi:10.1002/2013JD019877, 2013.
- Toohey, M. and von Clarmann, T.: Climatologies from satellite measurements: the impact of orbital sampling on the standard error of the mean, *Atmos. Meas. Tech.*, 6, 937–948, doi:10.5194/amt-6-937-2013, 2013.
- Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, *J. Geophys. Res. Atmos.*, 118, 11847–11862, doi:10.1002/jgrd.50874, 2013.
- Waters, J. W., Froidevaux, L., Harwood, R. S., Jarnot, R. F., Pickett, H. M., Read, W. G., Siegel, P. H., Cofield, R. E., Filipiak, M. J., Flower, D. A., Holden, J. R., Lau, G. K. K., Livesey, N. J., Manney, G. L., Pumphrey, H. C., Santee, M. L., Wu, D. L., Cuddy, D. T., Lay, R. R., Loo, M. S., Perun, V. S., Schwartz, M. J., Stek, P. C., Thurstans, R. P., Boyles, M. A., Chandra, K. M., Chavez, M. C., Chen, G. S., Chudasama, B. V., Dodge, R., Fuller, R. A., Girard, M. A., Jiang, J. H., Jiang, Y. B., Knosp, B. W., LaBelle, R. C., Lam, J. C., Lee, K. A., Miller, D., Oswald, J. E., Patel, N. C., Pukala, D. M., Quintero, O., Scaff, D. M., Van Snyder, W., Tope, M. C., Wagner, P. A., and Walch, M. J.: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, 44, 1075–1092, doi:10.1109/TGRS.2006.873771, 2006.