**Atmospheric
Chemistry
and Physics**

Statistical analysis of water vapour and ozone in the UT/LS observed during SPURT and MOZAIC

A. Kunz[1,2], **C. Schiller**[1], **F. Rohrer**[2], **H. G. J. Smit**[2], **P. Nedelec**[3], **and N. Spelten**[1]

[1]Institut für Chemie und Dynamik der Geosphäre: Stratosphäre (ICG-1), Forschungszentrum Jülich, Jülich, Germany
[2]Institut für Chemie und Dynamik der Geosphäre: Troposphäre (ICG-2), Forschungszentrum Jülich, Jülich, Germany
[3]Laboratoire d'Aerologie, CNRS, Observatoire Midi-Pyrenees, Toulouse, France

Abstract. A statistical analysis for the comparability of water (H_2O) and ozone (O_3) data sets sampled during the SPURT aircraft campaigns and the MOZAIC passenger aircraft flights is presented. The Kolmogoroff-Smirnoff test reveals that the distribution functions from SPURT and MOZAIC trace gases differ from each other with a confidence of 95%. A variance analysis shows a different variability character in both trace gas data sets. While the SPURT H_2O data only contain atmospheric processes variable on a diurnal or synoptical timescale, MOZAIC H_2O data also reveal processes, which vary on inter-seasonal and seasonal timescales. The SPURT H_2O data set does not represent the full MOZAIC H_2O variance in the UT/LS for climatological investigations, whereas the variance of O_3 is much better represented. SPURT H_2O data are better suited in the stratosphere, where the MOZAIC RH sensor looses its sensitivity.

1 Introduction

The composition of the tropopause region is strongly determined by large and small scale transports of trace gases. One governing process is the exchange of air masses between the stratosphere and the troposphere. Diabatic ascent or descent like convection or stratospheric intrusions from the overworld (above 380 K isentrope) lead to a vertical exchange and rapid exchanges by quasi-isentropic transport from and to the upper troposphere across the extratropical tropopause to a horizontal exchange (Stohl et al., 2003; Holton et al.,

1995). Mixing of stratospheric and tropospheric air leads to a so-called mixing layer around the tropopause (Hoor et al., 2002). These processes result in a highly variable trace gas distribution in the upper troposphere and lower stratosphere (UT/LS). The strong variability of these processes in time and space thus imply a highly variable composition of the tropopause region in different seasons and different geographical regions. Thus several airborne projects, e.g. SPURT and MOZAIC, were performed to measure the large-scale distribution of trace gases in the UT/LS.

Within the MOZAIC (Measurement of Ozone and Water Vapour by Airbus In-Service Aircraft) programme civil aircrafts are in regular service for making routinely measurements of chemical species in the atmosphere with almost global coverage. The project was initiated in 1993 with automatic in-situ H_2O and O_3 measurements onboard of up to five long-range A340 aircraft (Marenco et al., 1998). To date at least four flights are performed each day.

The SPURT (Trace gas transport in the tropopause region) campaigns between November 2001 and July 2003 deliver the distribution of a wide range of trace gases in the UT/LS region above Europe. As the campaigns equally cover all seasons, an accurate data set with climatological character should have been obtained to study atmospheric transport and to investigate seasonal variability of trace gases in the UT/LS (Engel, 2006).

A crucial question of this paper is on the representativeness of the limited SPURT data. Are they really suited for a climatological investigation on a seasonal and annual timescale and do they represent the full atmospheric variability of trace gases in the UT/LS? To answer this question we will investigate the comparability of trace gas mixing ratios observed during the limited number of flights in SPURT with
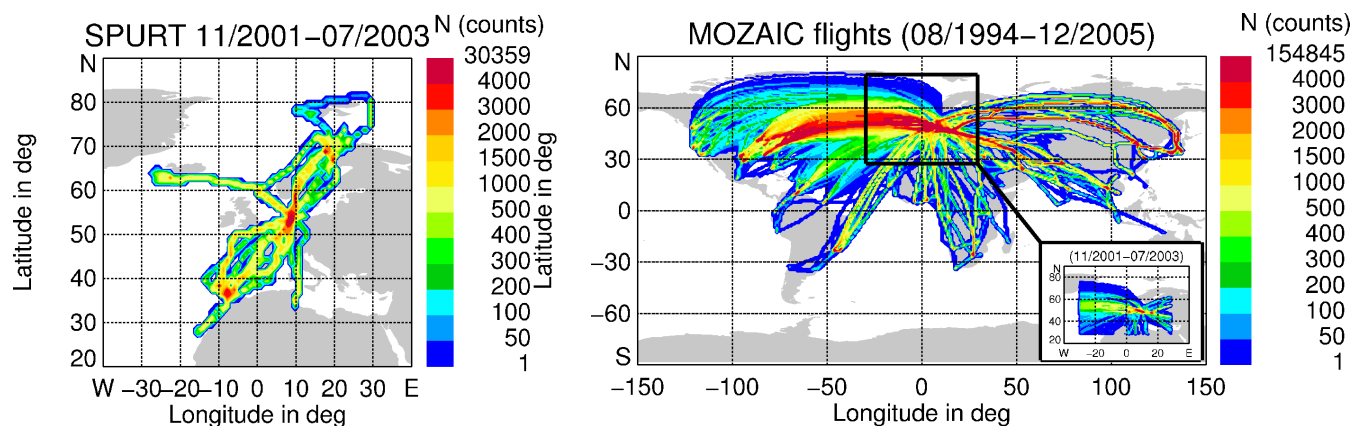
Fig. 1. Geographical distribution of flights during SPURT (left) and MOZAIC (right). The frequency of 1 Hz (SPURT) or 1-min-averaged (MOZAIC) data points in each geographical 1° lat×1° lon bin is colour-coded. The extension of SPURT flights is marked as black box in the MOZAIC plot and additionally the frequency of MOZAIC flights in this European sector during 2001 and 2003 can be seen down right.

those of the climatological data set obtained during the frequent MOZAIC flights. A statistical analysis of H_2O and O_3 follows to show in an objective manner the strengths and weaknesses of the two data sets. The analysis tools developed are not restricted to these particular data sets and are applicable for the comparison of different data sets, including model results, in a general sense.

2 Characteristics of the data sets

2.1 Geographical and vertical distribution

The SPURT project was performed to investigate the upper troposphere (UT) and lower stratosphere (LS). From November 2001 to July 2003, eight measurement campaigns were carried-out using a Learjet 35 A with a ceiling altitude of 13 km as measurement platform. A typical campaign consisted of 2–3 consecutive mission days. The data set is based on 36 flight missions and 147 flight hours. Each season during the SPURT period is captured by two measurement campaigns in subsequent years in order to investigate the seasonality of the trace gas concentrations (e.g., Krebsbach et al., 2006; Hoor et al., 2004; Hegglin et al., 2006). A description of the SPURT campaigns, the project strategy and performance is given in Engel (2006).

Figure 1 (left) shows the geographical distribution of the SPURT flights in 1 s data points. The aircraft was based at the Hohn military base in northern Germany. Southbound flights usually used Faro in southern Portugal for refueling and northbound flights Tromsø in Norway. Around the three stations the data density is very high because of slow ascents and descents.

The geographical distribution of MOZAIC measurements between 1994 and 2005 is displayed as one minute averages of 5 s measurements in Fig. 1 (right). MOZAIC flights cover

almost all continents. The Northern Hemisphere is better covered than the Southern Hemisphere, with more than 40% of MOZAIC flights in the North Atlantic flight corridor, more than 30% in Asia and around 10% of flights above Africa. Most of the measurements (90%) correspond to cruise altitudes 9–12 km (Marenco et al., 1998), lying in the troposphere in the tropics and subtropics and in the UT/LS at mid latitudes. The European region of SPURT campaigns is highlighted as black box and the measurement frequency between 2001 and 2003 in this region can be seen in the right bottom.

Figure 2 displays the vertical data coverage of SPURT and MOZAIC in Europe (see black box in Fig. 1) in 5 K potential temperature bins in reference to the tropopause (2 PVU surface). The distance of the trace gas data from tropopause (DTP) is derived with the help of potential vorticity and potential temperature, calculated from ECMWF output fields. The measurement frequency of MOZAIC in Europe (red line) peaks at a potential temperature of 330 K which corresponds to the vicinity of the tropopause. The maximum measurement frequencies of SPURT (black line) range between 335 K and 350 K, i.e. around 5 K below to 25 K above the tropopause. The average ceiling altitudes of the MOZAIC flights are lower and hence the maximum percentage of measurements appears at lower altitudes. More than 50% of MOZAIC flights and more than 75% of SPURT flights are performed in the lower stratosphere, so data should allow an investigation of trace gases in the tropopause region (e.g., Thouret et al., 2006; Law et al., 1998) and of exchange processes between the troposphere and the stratosphere.

2.2 Measurement systems

2.2.1 O_3 measuring instrument

MOZAIC O_3 is measured with a modified commercial dual beam UV-absorption photometer (Thermo-Electron,

Model 49-103). The measuring system and its performance are reported in detail by Thouret et al. (1998). The response time is better than 4 s with a detection limit of about ± 2 ppbv. The overall uncertainty is estimated to be about $\pm(2$ ppbv+2% of the observed reading). This corresponds to ± 2 ppbv for an O_3 mixing ratio of 10 ppbv, ± 4 ppbv at 100 ppbv, ± 6 ppbv at 200 ppbv (Thouret et al., 1998).

O_3 during SPURT was measured by UV absorption using the JOE (Jülich Ozone Experiment) instrument. The instrument is based on a Thermo Environmental Instrument ozonometer similar to that used for the MOZAIC programme. The instrument was operated with a time resolution of 10 s and has an accuracy of 5% (Mottaghy, 2001). The MOZAIC and SPURT O_3 instruments are regularly calibrated in the Jülich laboratories against the same reference instrument.

2.2.2 H₂O measuring instruments

During the SPURT campaigns H_2O mixing ratio was measured in-situ using the FISH (Fast In Situ Stratospheric Hygrometer) instrument (Zöger et al., 1999) which is based on the Lyman α photofragment fluorescence technique. The FISH instrument has a foreward facing inlet and measures total water, i.e. the sum of the gaseous phase and the condensed phase. The response time is 1 s, which allows also the detection of small-scale variations of H_2O mixing ratios in the vicinity of the tropopause, in clouds and contrails. The instruments accuracy is approximately 6% and the detection limit is better than 0.2 ppmv.

On board of the five MOZAIC airbuses relative humidity with respect to liquid water RH is measured with compact airborne humidity sensing devices (Helten et al., 1998). The sensing element consists of a capacitive sensor (Humicap-H, Vaisala, Finland) with a hydroactive polymer film as dielectric material whose capacitance depends on the relative humidity, and a platinum resistance sensor (PT100) for direct measurement of temperature at the humidity sensor. The sensor mounted in an appropriate Rosemount housing is designed for measurement of gas-phase water which is calculated from the relative humidity measurement. Adiabatic compression leads to a temperature increase of the sampled air and thus to a reduction of the dynamic range of the sensor and sufficient time response at low static air temperatures. In the middle troposphere the overall uncertainty is within ± 4% RH and around ± 7% RH between 9 and 13 km. This implies a limited use of the MOZAIC H_2O sensor in the stratosphere dominated by low RH and thus an increasing large uncertainty. The response time is around 10 s in the lower and middle troposphere and increases up to 1–3 min in the upper troposphere at 10–12 km altitude (Helten et al., 1998). After 500 operation hours the MOZAIC sensor is calibrated in the laboratory in Jülich.
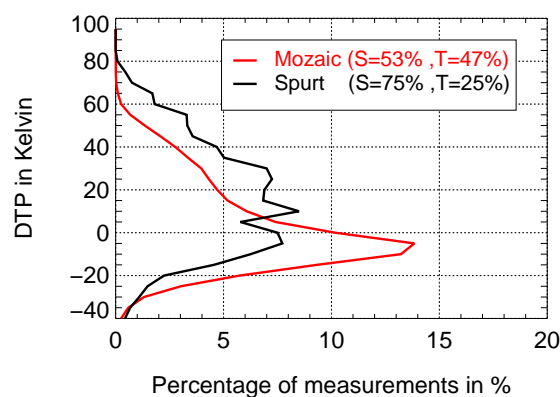

Fig. 2. Vertical distribution of percentage of data points dependent on potential temperature in distance to tropopause (DTP) during SPURT (black line) and MOZAIC in the European region (red line). Averages in 5 K bins are shown in reference to the tropopause (PV=2 PVU, DTP=0 K). The legend contains the percentage of data points in the stratosphere (S) and troposphere (T).

3 Statistical analysis

Both data sets are statistically analysed in order to assess the comparability of H_2O and O_3 data in SPURT and MOZAIC. A crucial question is whether or under which constraints the data sets with different coverage in space (region and altitude), time and with different instrument characteristics represent the same population in the atmospheric system. This includes the investigation whether the SPURT campaigns, with around eight flight missions in each season, are as representative as the MOZAIC daily flights for specific regions and whether the mixing ratios observed within the European sector during SPURT represent the seasonal trace gas variability.
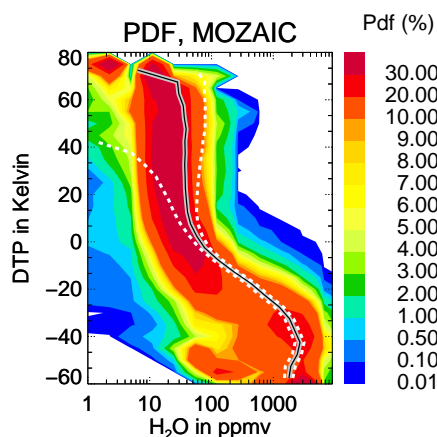
The following statistical analysis is performed for MOZAIC data observed in the same geographical region where the SPURT campaigns were carried out and for the same period from November 2001 until July 2003 (black box in Fig. 1 right). The MOZAIC and SPURT data sets are split according the distance to local tropopause (2 PVU surface): upper troposphere UT (DTP<–5 K) and lower stratosphere LS (DTP>5 K). So different sampling strategies and different trace gas characteristics should be accounted for. Influences by the large trace gas gradient in the vicinity of the tropopause (–5 K<DTP<5 K) are excluded.

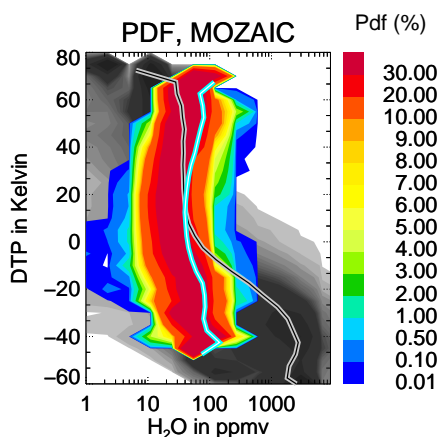### 3.1 Probability distribution and selection of data

Figure 3 shows the probability distribution functions (PDF) of H_2O data and Fig. 4 those of O_3 data dependent on the distance to tropopause for MOZAIC and SPURT (panels A and D respectively). The trace gas frequencies are calculated in 5 K bins relative to tropopause.

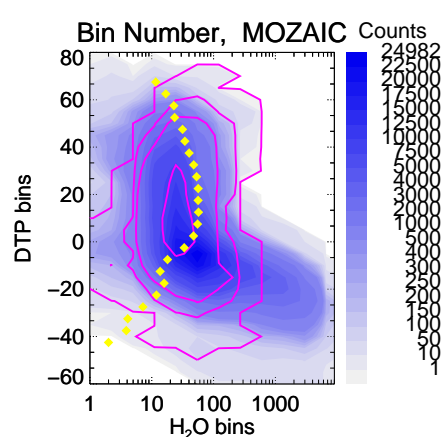**Fig. 3.** Probability distribution functions of MOZAIC (top) and SPURT (down) H_2O mixing ratio related to the distance to the local tropopause in K, considered as the 2 PVU surface. H_2O is binned in the logarithmical space between 0 and 9.6 with a bin size of 0.8, the distance to local tropopause in 5 K bins. Left panels: PDF of original H_2O data. The mean vertical profile (grey-black solid line) and the uncertainty of 5% RH (white dashed lines) are shown for the MOZAIC PDF. SPURT accuracy of H_2O data is 6% of concentration (not shown). Middle panels: The distribution of original data (panels A and D) is shadowed and those of selected H_2O data set (RH<10%, $RH_{ice} \leq 100\%$, $H_2O < 500$ ppmv, $p < 250$ hPa, see text) is colour coded. The mean PDFs are also shown as black-grey line (original data) and blue-white line (selected data). Right panels: Number of original data points per bin (blue shaded) and of selected data (pink non filled contours 0, 100, 500, 5000 data per DTP bin). The fraction of selected data relative to the original number in each DTP bin in percent is shown as yellow diamonds for all DTP bins with more than 1% data.

However, these probability distributions of H_2O reveal some differences between SPURT and MOZAIC. A very high probability of SPURT H_2O data lower than 10 ppmv occurs in the stratosphere more than 20 K above the tropopause (panel D). Most strikingly there is only a very low probability of H_2O data in the respective mixing ratio bins in MOZAIC (panel A). The MOZAIC H_2O probability becomes largest at higher mixing ratios in the stratosphere. Further there are no SPURT H_2O values larger than 2000 ppmv in the tro-

posphere more than 45 K below the tropopause, where the MOZAIC H_2O still contains up to 10 data points per bin (see density plots, panels C and F of Fig. 3). This is due to the measurement discrepancy with MOZAIC data sampled from the ground and SPURT data above the 400 hPa level. Hence there is a higher mean PDF (grey-black solid line) corresponding to a higher mean vertical H_2O profile both in the troposphere and in the stratosphere in MOZAIC than the SPURT. The MOZAIC mean H_2O profile remains nearly

MOZAIC - Ozone

A: Original data   **B: Selected data**   **C: Data density**
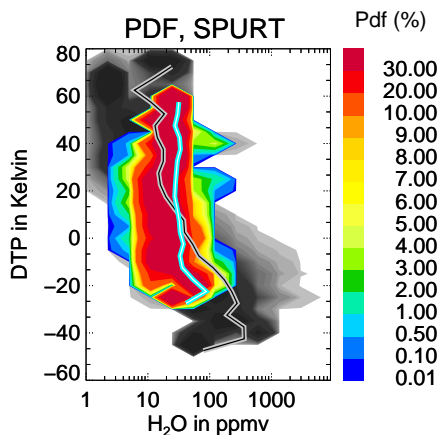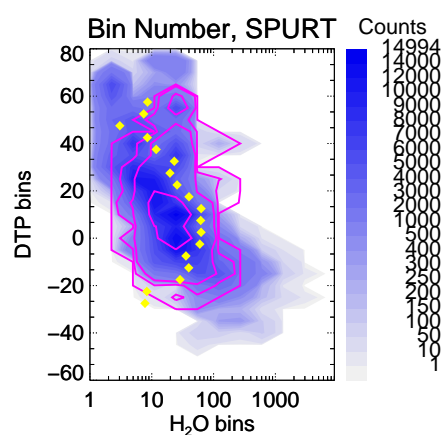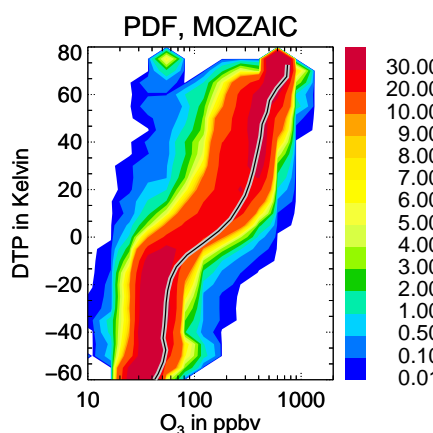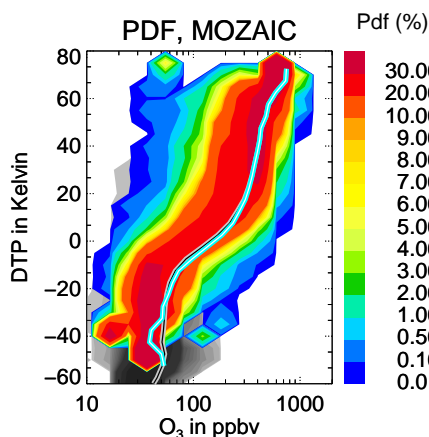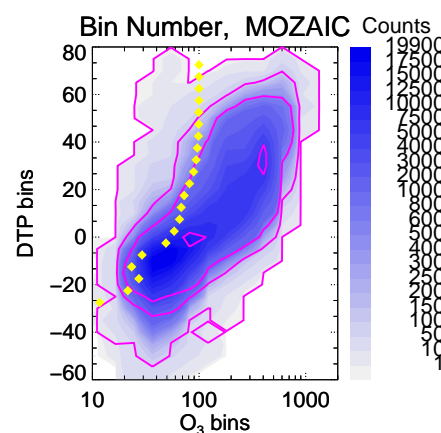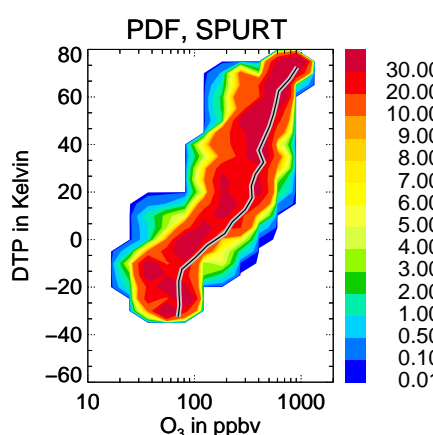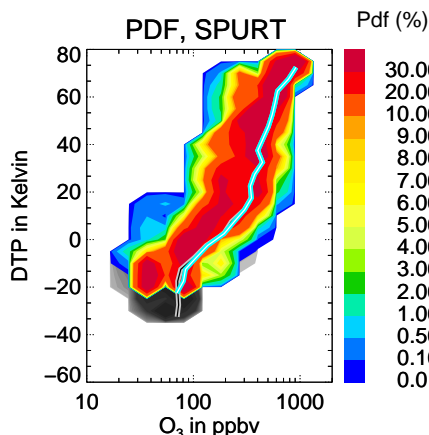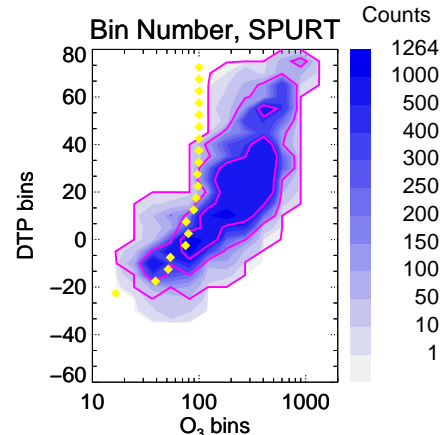

SPURT - Ozone

D: Original data   **E: Selected data**   **F: Data density**

Fig. 4. Probability distribution functions of MOZAIC (top) and SPURT (down) similar to Fig. 3, but now for O_3 mixing ratios related to the 2 PVU tropopause. The bin size for O_3 is 0.4 in the logarithmical space between 0 and 7.6, that for the DTP is again 5 K. With a very high accuracy of 5% the original trace gas distributions do not contain any accuracy limits. The right panels show the fraction of selected data relative to the original number in each DTP bin with more than 10% data as yellow diamonds.

constant around 40 ppmv in the stratosphere more than 5 K above the tropopause, whereas the SPURT mean H_2O profile decreases from 40 ppmv at the tropopause to mixing ratios lower than 10 ppmv around 60 K above the tropopause. Hereby, the 5% uncertainty of the MOZAIC sensor in the UT/LS must be accounted for. The uncertainty of $\pm 5\%$ relative humidity with respect to liquid water is shown as white dashed lines. The uncertainty range in volume mixing ratio scale is expanded in the entire stratosphere, attaining even negative values 40 K above the tropopause. The 5% RH uncertainty leads to a decreasing precision of H_2O volume mixing ratio deeper in the stratosphere. The SPURT H_2O data with a high relative accuracy of 6% of H_2O concentration do

Table 1. Selected constants for the Kolmogoroff-Smirnoff test. (Sachs and Hedderich, 2006)

α	0.20	0.15	0.10	0.05	0.01	0.001
K_α	1.07	1.14	1.22	1.36	1.63	1.95

not reveal this problem and the mean vertical mixing ratio also decreases in the stratosphere. A corresponding dashed white line is not shown in the SPURT PDF because of the small amount around the mean vertical profile.

The O_3 MOZAIC data set is stronger focused on low mixing ratios than the SPURT data set (see panels A and D of Fig. 4). There is a very high probability of MOZAIC O_3 data in the troposphere below -35 K, where the SPURT data do not contain any O_3 mixing ratios. In the UT/LS above -35 K the mean vertical O_3 profiles (grey-black lines) of SPURT and MOZAIC are very similar and the mixing ratio at the tropopause is around 150 ppbv in both cases.

The discrepancies between both data sets basically result from different instrumental characteristics or measurement strategies. Because of the different H_2O measurement techniques (see Sect. 2.2.2) the H_2O data have to be modified before a statistical comparison using the following selection criteria:

– The MOZAIC Humicap sensor has a precision of 4–7% RH, i.e. low H_2O mixing ratios are not detected and cannot be contained in the PDFs of Fig. 3. Thus the dry measurements according to RH$<$10% in particular in the stratosphere, where SPURT was focused on, cannot be included in the comparison due to sensitivity limitations of the MOZAIC sensor at low RH.

– The FISH instrument has a foreward facing inlet and measures total water, i.e. both the gas phase and the condensed phase H_2O mixing ratios. The MOZAIC Humicap sensor measures relative humidity with respect to liquid water and the mixing ratios represent only the gas phase. Therefore, only data with a relative humidity with respect to ice RH$_{ice}\leq$100% can be compared eliminating measurements in clouds and under supersaturation conditions.

– H_2O mixing ratios larger than 500 ppmv are sorted out, because the FISH instrument is calibrated for mixing ratios below this limit. At larger mixing ratios the measurement cell of FISH becomes optically dense and the FISH fluorescence method is limited on in-situ measurements above a mixing ratio of 500 ppmv. To select only data representative for the UT/LS we further choose the 250 hPa pressure level as lower limit.

– In the UT/LS the MOZAIC sensor has a response time of $\tau \approx$ 60 s and the FISH instrument of $\tau \approx$ 1 s. A running mean with a time interval of 60 seconds is therefore applied on the SPURT data for this study.

These selection criteria are applied on the MOZAIC and SPURT H_2O data. The third criterion with a data selection above the 250 hPa pressure level is also applied on the O_3 data in order to compensate for the tropospheric bias of the complete MOZAIC data set.

Panels B and E in Figs. 3 and 4 show the new H_2O and O_3 PDFs of the modified data according the selection criteria (colour coded) and the original PDFs, also shown in panels A and D, as shadowed area. The mean vertical profile of the selected H_2O data set (blue line) is shifted towards larger values in the stratosphere and towards lower values in the troposphere. As a consequence of the criterion to select data with relative humidities above RH$>$10%, H_2O mixing ratios below 10 ppmv are excluded. The most probable H_2O data in the stratosphere are now between 10 and 30 ppmv both in SPURT and MOZAIC. In the troposphere the data are removed because of the 500 ppmv, the 250 hPa and the RH$_{ice}\leq$100% criteria. According to the 250 hPa criterion there is a O_3 data loss in the troposphere, most effecting the MOZAIC data set.

The normalized frequency distributions of the H_2O (left) and O_3 mixing ratios (right) of MOZAIC (red) and SPURT (black) in Fig. 5 demonstrate an adjustment for both trace gases when the data selection is applied (solid lines=selected data; dashed lines=original data). But there are still some differences left as e.g. a high normalized H_2O frequency in SPURT at lower mixing ratios in the troposphere. A difference in sample means and medians remains. The mean O_3 mixing ratios are larger in SPURT than in MOZAIC and vice versa for H_2O (triangles), thus still reflecting the different vertical sampling range of both projects. The broadness of the SPURT and MOZAIC H_2O distribution after the selection is very similar especially in the stratosphere. The number of data points (legend of Fig. 5) demonstrates a data loss of around 65% of H_2O due to data selection both for SPURT and MOZAIC, around 12% of O_3 is lost for SPURT and 45% of O_3 data in MOZAIC.

For the following statistical analysis, the reduced data sets of H_2O and O_3 in which differences due to the different H_2O measurement techniques and sampling strategies are eliminated as far as possible, will be used.
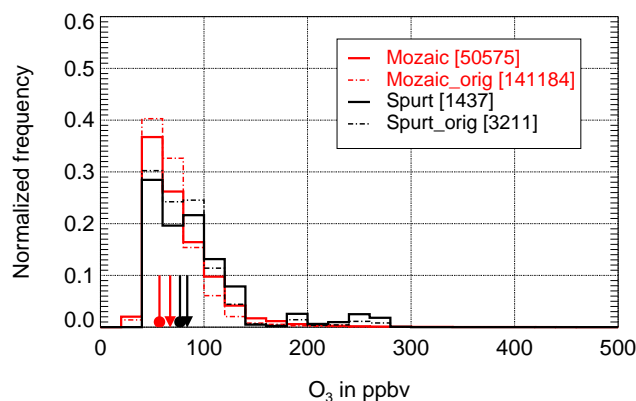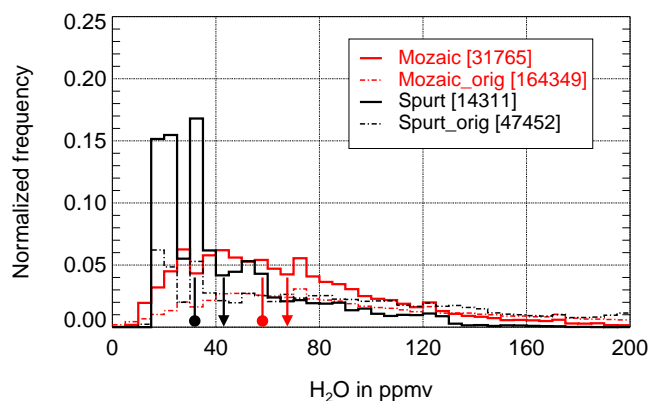
3.2 Kolmogoroff-Smirnoff test

The Kolmogoroff-Smirnoff goodness-of-fit test compares two independent random samples of measured data and examines whether they stem from the same population (Brandt, 1999; Sachs and Hedderich, 2006). Compared to other goodness-of-fit tests, e.g. the χ^2-test, the Kolmogoroff-Smirnoff test can be applied to non-normally distributed data. The test is well suited to investigate whether both random samples belong to the same population. The central tendency of the variance, the skewness and kurtosis, i.e. differences of the type of distribution and thus of the distribution functions in Fig. 5 are captured.

3.2.1 Mathematical description

The test statistic is the maximum observed difference of the ordinate between the two non overlapping cumulative frequency curves. Both statistical samples, i.e. the MOZAIC and SPURT data, are binned in an equal number of classes. The empirical cumulative distribution functions $\hat{F}_{spurt}$ and $\hat{F}_{mozaic}$ and their differences $\hat{F}_{spurt}-\hat{F}_{mozaic}$ are calculated.

Troposphere
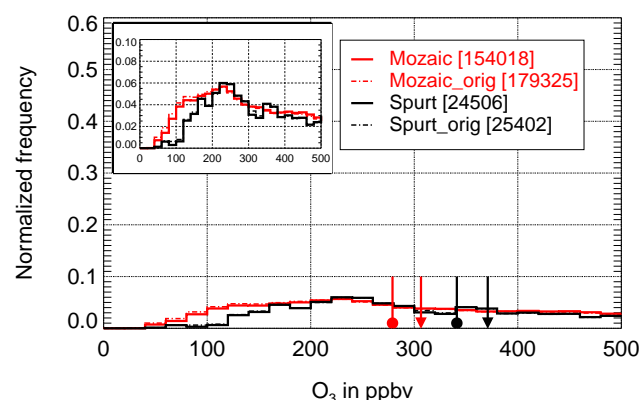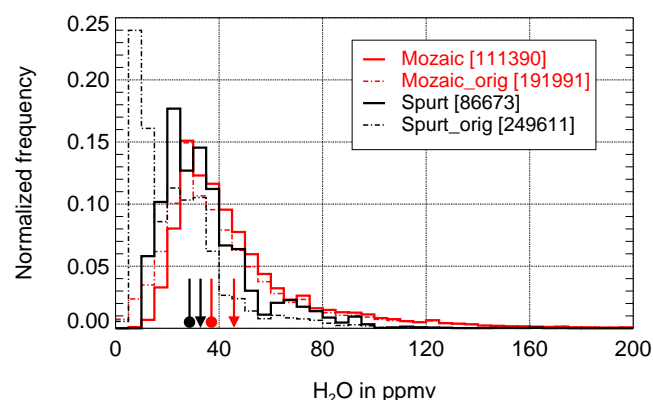

Stratosphere

Fig. 5. Frequency distributions of the H$_2$O (left) and O$_3$ mixing ratio in the troposphere (DTP<−5 K) and stratosphere (DTP>5 K), normalized by dividing the single bin frequencies in percent by the total number of data points (see legend). The frequency distributions of the data selected by the instrument criteria (see text) are represented by solid lines, those of unselected original data by dashed lines. H$_2$O is binned in 5 ppmv and O$_3$ in 10 ppbv. The means of the selected data MOZAIC and SPURT are marked by triangle symbols, the medians by circle symbols. In case of unselected data they are beyond the range of the ordinate.

The test statistic $\hat{D}$ is the maximum of the absolute value of this difference, i.e.

$$\hat{D} = \max \left| \left(\hat{F}_{\text{spurt}} - \hat{F}_{\text{mozaic}} \right) \right|. \quad (1)$$

For large sample sizes ($n_{\text{spurt}} + n_{\text{mozaic}} > 35$) the cutoff value D_α can be approximated by

$$D_\alpha = K_\alpha \cdot \sqrt{\frac{n_{\text{spurt}} + n_{\text{mozaic}}}{n_{\text{spurt}} \cdot n_{\text{mozaic}}}}, \quad (2)$$

with n_{spurt} and n_{mozaic} the number of elements of the two statistical samples and K_α the Kolmogoroff-Smirnoff constant dependent on the error probability α. Table 1 contains the corresponding values of K_α.

If the test statistic $\hat{D}$, calculated from both samples, is greater or equal to the cutoff value D_α, both distribution functions are significantly different with a selected error probability.

3.2.2 Test performance

The null-hypothesis H_0 "Both distribution functions of trace gases H$_2$O and O$_3$ in MOZAIC and SPURT are the same" is tested against the alternative hypothesis H_A "Both distribution functions are different from each other" with a confidence of α=95%. The larger the test statistic $\hat{D}$ in Eq. 1, the more the null-hypothesis has to be rejected.

Table 2 shows the values of the test statistic $\hat{D}$ and the corresponding cutoff values D_α calculated both for data within the troposphere (DTP<−5 K) and stratosphere (DTP>5 K).

The test statistic $\hat{D}$ in Table 2 is much larger than the cutoff value D_α for all cases, the null hypothesis of equal distribution functions for both the H$_2$O and the O$_3$ mixing ratio therefore can be rejected with a confidence of α=95%. The tests are also performed for different confidences varying between α=95% and α=99.9% (see also Table 2) with the same test results. Therefore, with high confidence the H$_2$O and O$_3$ mixing ratios of the MOZAIC and SPURT data
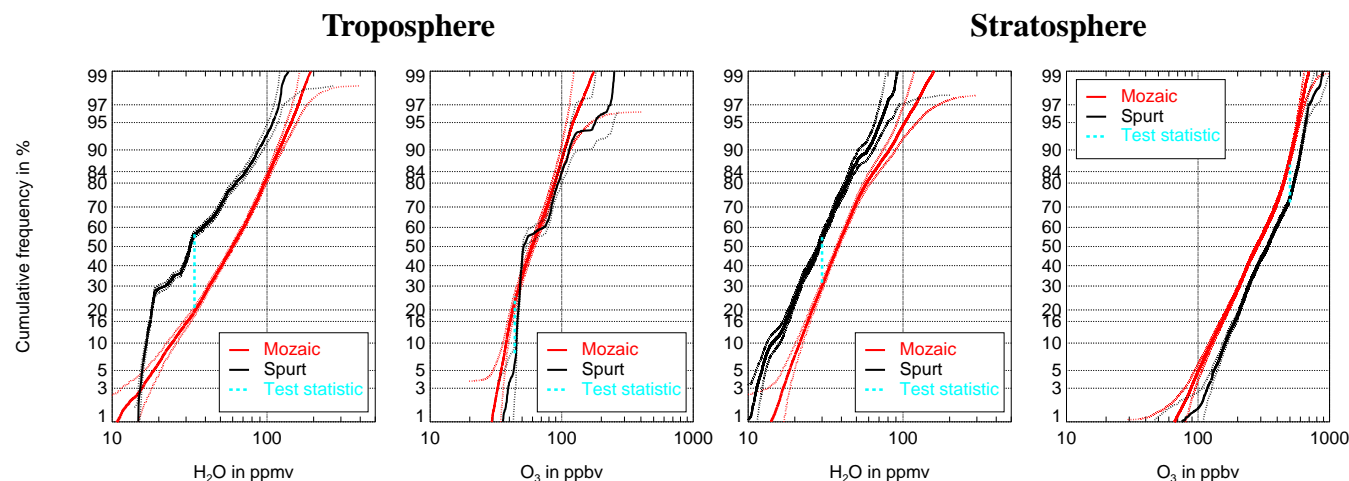
Troposphere **Stratosphere**

Fig. 6. Probability networks with empirical cumulative frequency functions in% of H_2O and O_3 mixing ratios in MOZAIC (red) and SPURT (black). Left: Troposphere (DTP<−5 K). Right: stratosphere (DTP>5 K). The cutoff value D_α is displayed as α=95% confidence region for each frequency function (dotted line) and the corresponding test statistic $\hat{D}$ (dashed cyan line).

Table 2. Kolmogoroff-Smirnoff test statistics $\hat{D}$ and cutoff values D_α (rounded for four decimal places) for two different confidences α=95% and α=99.9%

	$\hat{D}(H_2O)$	$D_{\alpha=95\%}^{H_2O}$	$D_{\alpha=99.9\%}^{H_2O}$	$\hat{D}(O_3)$	$D_{\alpha=95\%}^{O_3}$	$D_{\alpha=99.9\%}^{O_3}$
Troposphere	0.3691	0.0137	0.0164	0.1757	0.0308	0.0369
Stratosphere	0.2503	0.0061	0.0074	0.1403	0.0090	0.0107

sets differ from each other.

A graphical display of the Kolmogoroff-Smirnoff test results gives a so-called probability network (see Fig. 6). The H_2O and O_3 cumulative frequency functions $\hat{F}_{spurt}$ and $\hat{F}_{mozaic}$ are plotted logarithmically in this probability network for the troposphere (panels left) and the stratosphere (panels right). The corresponding cutoff value D_α is plotted as confidence region for each distribution function (dotted lines). If the null hypothesis H_0 of equal distribution functions is not rejected, the frequency function of $\hat{F}_{mozaic}$ lies within the confidence limit of the other distribution function $\hat{F}_{spurt}$ and vice versa. Note the distorted ordinate according to the χ^2-distribution function, which causes the different range of confidence limits although the Kolmogoroff-Smirnoff constant is equal in the whole mixing ratio range of the abszissa. The corresponding cutoff values and test statistics can be found in Table 2.

The test results in Table 2 are reflected well in these figures. In each case we find a region, where both cumulative frequency functions differ significantly from each other, i.e. where the difference between both functions is largest. The maximum difference in ordinate, corresponding to the test statistic $\hat{D}$, is always located at the middle range of mixing

ratios (see dashed cyan line in Fig. 6). The two tested cumulative distribution functions do not generally lie in the confidence limit of the other one, thus both statistical data samples are different from each other and do not belong to the same population. Although the O_3 cumulative distribution functions are very close to each other for each atmospheric region we still find a small area where the test statistic becomes larger than the cutoff value and thus there is a statistical difference between both distribution functions.

We find a difference between the cumulative distribution functions both for O_3 data based on the same measurement techniques and for H_2O data using different measurement techniques. This indicates that there are other, most likely sampling or regional causes for the differences between the trace gas data in SPURT and MOZAIC.

3.3 Variance analysis

Here, the selected data samples are examined for their variability characteristics. Each SPURT campaign consisted of typically four flights, with a flight time of around four hours each. Each season is covered by eight single flights with H_2O data. Thus these few days represent a whole

season. MOZAIC, however, provides at least two flights with H_2O data for each day. Hence the SPURT and MOZAIC data are expected to be subject to variability on different timescales. The term timescale in this context is more a matter of speech. Since the movement of the aircrafts is fast compared to the wind speed the onboard sensors encounter the spacial gradients at short timescales and the temporal gradients at long timescales. Since both aircrafts are moving with approximately the same speed the interaction of spatial and temporal gradients is comparable. The concept of a temporal statistical variance analysis is an appropriate tool to investigate trace gas variability and provides information about atmospheric and even chemical influences (Rohrer and Berresheim, 2006).

3.3.1 Test description

For a variance analysis the H_2O and O_3 data sets in MOZAIC and SPURT are binned into series of time intervals of different lengths, i.e. timescales, between several minutes and years. A mean variance is calculated for each timescale. When dividing a data set of a timescale of one year into two half year data sets, a variance is calculated of the data within the half year bins. Both resulting sample variances are averaged and the mean variance for the data set about a half year results. Then the one year data set is divided into three four month data sets, the procedure is repeated and the mean variance about a three month bin is calculated.

3.3.2 Analysis applied on CIRRUS III flight

Before doing the variance analysis on the complete SPURT and MOZAIC data as in Sect. 3.2, we introduce the analysis on the water vapour data observed during one single flight of the CIRRUS III campaign (top of Fig. 7). The motivation of the three CIRRUS campaigns between 2002 and 2006 was to investigate the formation mechanism of cirrus clouds, their radiative effects and to study the chemical or microphysical properties of the cloud particles. Both the FISH instrument and the MOZAIC sensor, already described in Sect. 2, were onboard the Learjet 35 A during the last CIRRUS III campaign in November 2006. The CIRRUS III midlatitude cirrus field experiment took also place at the Hohn military base. Six flights mainly inside and outside frontal cirrus clouds were performed in the altitude range from 7–12 km between 45–70°N. So we can perform an inflight comparison of both instruments and show the results of a variance analysis, if the data are sampled under the same spatial and temporal conditions. A good opportunity to study the importance of interaction between temporal and spatial variances on small timescales during an in-flight comparison. The selection criteria are also applied on this data set and for the resulting data (grey shaded area in Fig. 7 top) the variance analysis reveals a really good agreement between the H_2O variances as observed by FISH (black) and Mozaic H_2O (red) sensor
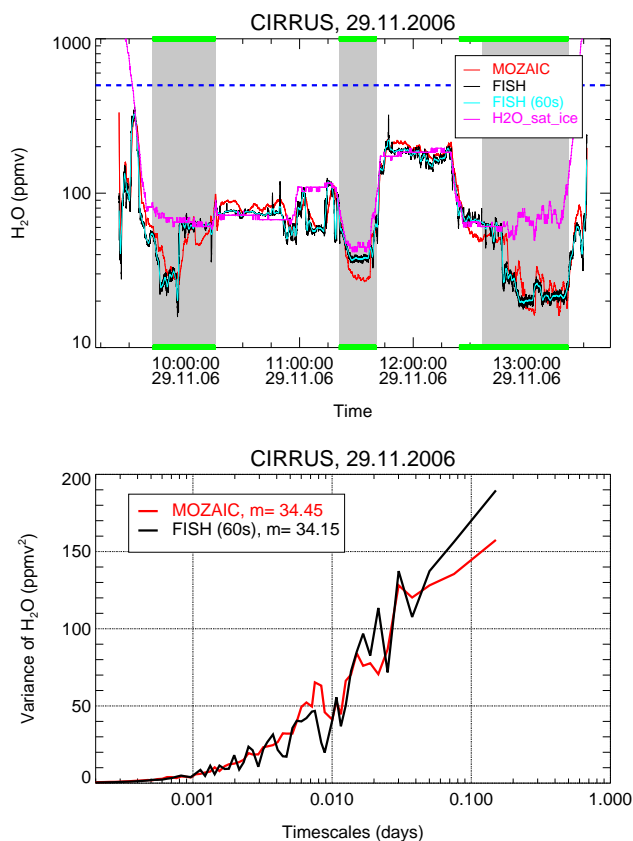

Fig. 7. Top: In-flight comparison of FISH (black) and Mozaic sensor (red) H_2O mixing ratio during one flight mission of the CIRRUS III campaign. The 60 s running mean of the FISH H_2O mixing ratio is highlighted in cyan and the saturation H_2O mixing ratio in pink. The part of the flight, which is performed above the 250 hPa pressure level is bounded by the green line. After all selection criteria are applied, data above the grey shaded area are used for the variance analysis. Down: Variance analysis of the FISH (black) and Mozaic sensor (red) H_2O mixing ratio during the CIRRUS III flight.
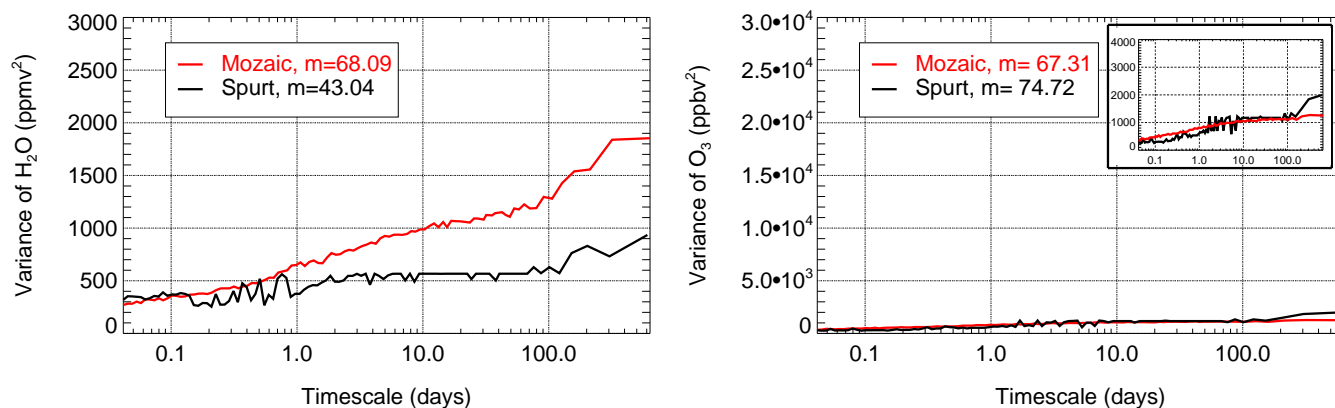
(Fig. 7, down). A similar increasing variance on a timescale of 3.5 h demonstrates that both instruments detected the same atmospheric processes and that there is no discrepancy due to the unequal measurement instruments left.

3.3.3 Analysis on MOZAIC and SPURT flights

Figure 8 shows the variance analysis of MOZAIC and SPURT H_2O (left) and O_3 data (right) for the troposphere and stratosphere.

The variance of H_2O in MOZAIC increases from short to long timescales within the troposphere (red line top of Fig. 8). There are four consecutive timescale regions, representing a different strength and change of atmospheric H_2O variability. An increasing variance on an one hour to one or two days, representing the H_2O variability on a diurnal timescale. Further an enhancement on a typical synoptical
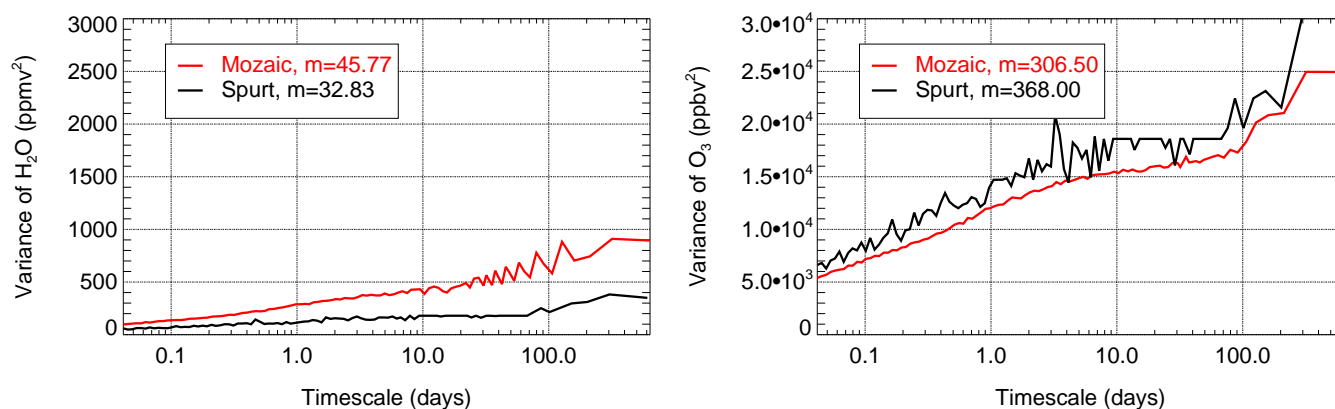
Troposphere

Stratosphere

Fig. 8. Variance analysis of the H$_2$O in ppmv^2 (left) and O$_3$ in ppbv^2 (right) mixing ratio in MOZAIC (red) and SPURT (black) in different atmospheric regions in reference to the local tropopause (PV=2). Top panels: troposphere (DTP<−5 K). Bottom panels: stratosphere (DTP>5 K). The tropospheric variance of MOZAIC O$_3$ is additionally shown enlarged top right.

ten-day timescale and on an interseasonal timescale between 10 and 90 days is observed. On both timescales the variance enhancement is not as sharp as on the diurnal timescale. At least there is an extreme increase of variance of H$_2$O data on the 90 to 300 days timescale, representing a seasonal variability of H$_2$O mixing ratio in MOZAIC.

The tropospheric H$_2$O variance in SPURT (black line) coincides with that of MOZAIC on a timescale of 0.15 days, i.e. around four hours. This variance is not only representing the temporal but also the spatial variance. A typical duration of a SPURT flight and those of the MOZAIC flight within Europe was around four hours. The aircrafts velocity of both projects is nearly the same and both measurement systems are comparable on a short timescale of some hours as shown in Fig. 7. There is still a good agreement on a timescale of 1 day, but on longer timescales both variances diverge more and more resulting in a much lower variance

of H$_2$O in SPURT than in MOZAIC. An increasing variance of SPURT H$_2$O can be observed on a three day timescale, the typical timescale of the mission days during each aircraft campaign.

On longer periods till 90 days the variance remains approximately constant, fluctuating around a statistical mean on a three to ten day timescale. This fluctuation reduces on longer timescales. On a seasonal timescale we find again a variance of SPURT H$_2$O data.

When dividing the SPURT data set into different time series of non-regular timescales, most of the bins do not contain measurement data. On an inter-seasonal timescale there are SPURT data available on two or three consecutive days. As a consequence when calculating the variance on a 100 day timescale the variance will remain constant until reaching the prescribed bin, which contains the measurement data. This bin includes a timescale of one or two days. As consequence

we do not find any H$_2$O variability on an inter-seasonal timescale in SPURT. The variance on a seasonal timescale bases on single flights on two or three consecutive days each season during the two years.

The H$_2$O variances decrease in the stratosphere (bottom panels in Fig. 8), representing the smaller H$_2$O variability in the upper atmosphere. The difference of the variance between MOZAIC and SPURT reduces in the stratosphere, but a discrepance remains.

For SPURT, the stratospheric tracer O$_3$ reveals an enhancing variance on a ten-day timescale as for MOZAIC (see Fig. 8 right panels). There is no enhancement of SPURT O$_3$ variance on an interseasonal timescale till 90 days, but also for MOZAIC the O$_3$ variance increases only marginally. On a seasonal timescale till 300 days there is an increasing variance both for SPURT and MOZAIC O$_3$. Compared to the troposphere the O$_3$ variance increases in the stratosphere. The slope of the O$_3$ variance of SPURT is similar to that of MOZAIC and there is no considerable difference between SPURT and MOZAIC O$_3$ variance as observed for the tropospheric tracer H$_2$O.

4   Discussion

The data selection in Sect. 3.1 is essential to achieve a sufficient agreement of the frequency distribution functions for both trace gases and projects (see Fig. 5) with some differences left to allow for a statistical comparison of both data sets.

The Kolmogoroff-Smirnoff test reveals a statistical difference between the respective H$_2$O and O$_3$ mixing ratios observed during both projects. The H$_2$O cumulative distribution function for MOZAIC is larger than that for SPURT both in the UT and LS, and vice versa for O$_3$. There are still different sample means with higher SPURT H$_2$O means and lower O$_3$ means than in MOZAIC. Especially in the stratosphere this must be due to the different campaign performance, with the Learjet in SPURT flying deeper into the stratosphere and thus sampling a higher O$_3$ and lower H$_2$O mixing ratio to average (Fig. 2).

The causes for the statistical difference in the H$_2$O and O$_3$ data sets become apparent by a variance analysis (Fig. 8). The H$_2$O data observed during the SPURT campaigns contain atmospheric processes, which take place on a diurnal timescale. There is a fluctuating variance between several minutes and two to three days.

The SPURT data set does not contain information about any processes on longer inter-seasonal timescales, but on a seasonal timescale between 90 and 300 days. Thus SPURT contains on the one hand processes playing a role on the typical campaign timescale (one till three days). Further the seasonal variability is based on the equally time-spaced performed campaigns, each season is covered by two campaigns. Thus the trace gas variability within a season (10

till 90 days) is not included which is about 50% of the total variance of H$_2$O.

The MOZAIC H$_2$O measurements are influenced by synoptic scale processes on a ten-day timescale and by processes on an inter-seasonal timescale. The variance enhancing on a ten-day timescale represents a variability which is typical for synoptic weather systems influencing the air mass composition in a specific region as low or high pressure systems. There is further a variance between 10 and 90 days, representing processes varying on an inter-seasonal timescale up to three months and a variability on a seasonal timescale. Contrary to SPURT the MOZAIC data set thus gives information about processes on each timescale.
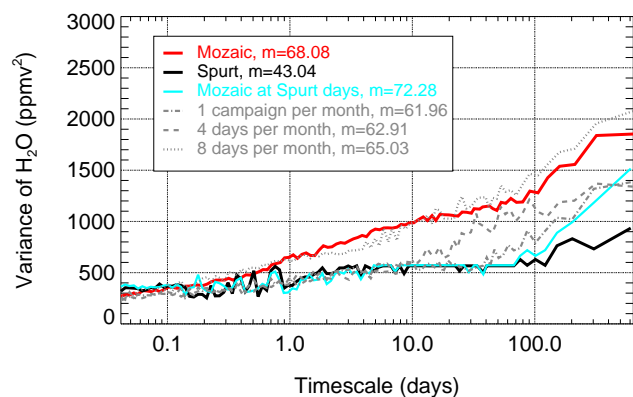
Especially MOZAIC contains information about processes which are representative for the different seasons. The SPURT H$_2$O variance is not representative for the seasonal timescale and rather gives an instantaneous picture of the atmosphere on the single flight days. SPURT is rather dominated by short scale fluctuating processes. These different processes in both data sets are the reason for differences in the frequency distribution functions (Fig. 5).

On long timescales the H$_2$O variances of MOZAIC and SPURT differ more and more due to the different measurement frequency. The difference is largest in the troposphere, the full atmospheric H$_2$O variance in the UT is not captured by the SPURT campaigns. Large scale atmospheric processes and turbulent systems playing a role in the UT on a longer than diurnal timescale and influencing the variability of the tropospheric tracer H$_2$O are not contained in the SPURT data and account for this difference. The H$_2$O difference in variance lessens in the stratosphere, but still remains.

The stratospheric tracer O$_3$ does not reveal the differences in variance in SPURT and MOZAIC as observed for H$_2$O. SPURT O$_3$ data represent the atmospheric processes influencing the O$_3$ distribution in the UT/LS on each timescale despite the inter-seasonal timescale between 10 and 90 days as expected. But the full atmospheric O$_3$ variance as MOZAIC shows is achieved in the UT/LS on every timescale thus demonstrating that contrary to the tropospheric tracer H$_2$O the amount of SPURT data is sufficient to represent the full O$_3$ variability even on seasonal timescales. This demonstrates the different variability behaviour of the stratospheric tracer O$_3$ independent of short scale fluctuating processes and acting on longer timescales.

The variance analysis is further performed for different subsamples of MOZAIC data (Fig. 9). The variance of the full MOZAIC H$_2$O data between November 2001 and July 2003 (red line) is compared with that of the MOZAIC data on the single Spurt mission days (cyan line), which has a very similar shape as that of SPURT H$_2$O (black line). The difference to the variance of the full MOZAIC data (red) reduces marginally on the timescale between 40 to 300 days in the troposphere and between 40 to 150 days in the stratosphere, if there would be one campaign each month (dashed-dotted line). To capture the full atmospheric H$_2$O variance
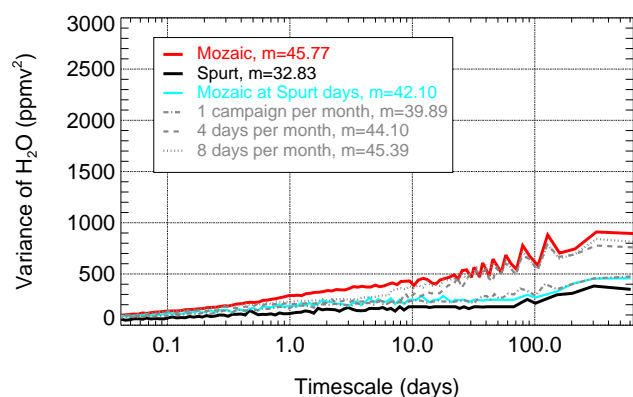
Troposphere

Stratosphere

Fig. 9. Variance analysis of different H_2O subsamples from the MOZAIC data set in the troposphere (top) and stratosphere (bottom). The H_2O variances of the SPURT (black) and full MOZAIC data (red) from Fig. 8 are additionally shown with the variance of MOZAIC data sampled on the single SPURT flight days (cyan). Further variances are calculated corresponding to one campaign per month (dashed-dotted), i.e. flights on two consecutive days each month between November 2001 and July 2003. Four flight days per month (dashed), i.e. flights on every sixth day. Eight flight days according to flights each fourth day (dotted).

as MOZAIC shows there have to be each fourth day measurement flights in the troposphere (dotted line), while in the stratosphere measurements each sixth day are sufficient especially on inter-seasonal and seasonal timescales (dashed line). That means that the Learjet would have to fly on around eight days per month in the troposphere and around four days per month in the stratosphere to capture the full climatological variability of MOZAIC H_2O.

5   Conclusions

The statistical analysis shows that the SPURT data set, despite its much larger temporal and spatial coverage as compared to other campaigns with research aircraft, does not represent the full variability of atmospheric H_2O in the tropopause region and can only be used for limited climatological investigations. The single flights of SPURT cannot replace the large number of MOZAIC flights when analysing the H_2O distribution in a climatological manner. The SPURT observations rather give an instantaneous picture of one day variability especially of the upper tropospheric H_2O mixing ratio observed during the limited number of flight hours of the single flights. Information about large scale processes varying on a seasonal timescale are less representative, as a variance analysis reveals. For O_3 the number of SPURT flights is almost sufficient. SPURT delivers the atmospheric variability of O_3 on each timescale except of the interseasonal one, which however is weak as the MOZAIC data show. SPURT O_3 can therefore be used even for climatological investigations. The MOZAIC trace gas data are not limited in the variance characteristics. These data represent atmospheric processes varying on longer timescales like synoptical weather systems. They are ideal for seasonal and annual investigations of H_2O and O_3 mixing ratios.

However, the statistical comparison reveals the known limitation of the MOZAIC RH sensor in the LS. Small scale fluctuations in the UT/LS cannot be observed by this capacitive sensor, while the FISH instrument in SPURT is well suited for studies with attention to fast processes in the UT/LS, as mixing and transport processes.

We have introduced a convenient statistical procedure to compare trace gas data sets of different projects even if they do not coincide in space and time. It would be interesting to adapt these tests on other observational data sets. The tests are further suited for an evaluation and comparison with results from atmospheric models as the Chemical Lagrangian Model of the Stratosphere (CLaMS) (McKenna et al., 2002) and MOZART, the Model of Ozone and Related Chemical Tracers (ECHAM5-MOZ).

Edited by: M. Dameris

References

Brandt, S.: Data Analysis – Statistical and Computational Methods for Scientests and Engineers, Springer Berlin, 686 pp,. 1999.

Engel, A., Bönisch, H., Brunner, D., Fischer, H., Franke, H., Günther, G., Gurk, C., Hegglin, M., Hoor, P., Königstedt, R., Krebsbach, M., Maser, R., Parchatka, U., Peter, T., Schell, D., Schiller, C., Schmidt, U., Spelten, N., Szabo, T., Weers, U., Wernli, H., Wetter, T., and Wirth, V.: Highly resolved observations of trace gases in the lowermost stratosphere and upper troposphere from the SPURT project: an overview, Atmos. Chem. Phys., 6, 283–310, 2006,
http://www.atmos-chem-phys.net/6/283/2006/.

Hegglin, M. I., Brunner, D., Peter, T., Hoor, P., Fischer, H., Stähelin, J., Krebsbach, M., Schiller, C., Parchatka, U., and Weers, U.: Measurements of NO, N$_y$, N$_2$O, and O$_3$ during SPURT: implications for transport and chemistry in the lowermost stratosphere, Atmos. Chem. Phys, 6, 1331–1350, 2006.

Helten, M., Smit, H. G. J., Sträter, W., Kley, D., Nedelec, P., Zöger, M., and Busen, R.: Calibration and performance of automatic compact instrumentation for the measurement of relative humidity from passenger aircraft, J. Geophys. Res., 103, 25 643–25 652, 1998.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.: Stratosphere-troposphere exchange, Rev. Geophys., 33, 403–440, 1995.

Hoor, P., Fischer, H., Lange, L., Lelieveld, J., and Brunner, D.: Seasonal variations of a mixing layer in the lowermost stratosphere as identified by the CO-O$_3$ correlation from in-situ measurements, J. Geophys. Res., 107(D5), 4044, doi:10.1029/2000JD000289, 2002.

Hoor, P., Gurk, C., Brunner, D., Hegglin, M. I., Wernli, H., and Fischer, H.: Seasonality and extent of extratropical TST derived from in-situ CO measurements during SPURT, Atmos. Chem. Phys., 4, 1427–1442, 2004,
http://www.atmos-chem-phys.net/4/1427/2004/.

Krebsbach, M., Schiller, C., Brunner, D., Günther, G., Hegglin, M. I., Mottaghy, D., Riese, M., Spelten, N., and Wernli, H.: Seasonal cycles and variability of O$_3$ and H$_2$O in the UT/LMS during SPURT, Atmos. Chem. Phys., 6, 109–125, 2006,
http://www.atmos-chem-phys.net/6/109/2006/.

Law, K. S., Plantevin, P. H., Shallcross, D. E., Rogers, H. I., and Pyle, J. A.: Evaluation of modeled O$_3$ using Measurement of Ozone by Airbus In-Service Aircraft (MOZAIC) data, J. Geophys. Res., 103, 25 721–25 737, 1998.

Marenco, A., Thouret, V., Nedelec, P., Smit, H., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C., and Cook, T.: Measurement of ozone and water vapour by Airbus in-service aircraft: The MOZAIC airborne program, An overview, J. Geophys. Res., 103, 25 631–25 642, 1998.

Mottaghy, D.: Ozonmessungen in der unteren Stratosphäre, Master's thesis, Rheinisch-Westfälische Technische Hochschule Aachen, in cooperation with the Institute for Chemistry and Dynamics of the Geosphere, ICG-1: Stratosphere, 2001.

McKenna, D., S., Konopka, P., Grooš, J., U., Günther, G., and Müller, R.: A new Chemical Lagrangian Model of the Stratosphere (CLaMS): 1. Formulation of advection and mixing, J. Geophys. Res., 107(D16), 4309, doi:10.1029/2000JD000114, 2002.

Rohrer, F. and Berresheim, H.: Strong correlation between levels of tropospheric hydroxyl radicals and solar ultraviolet radiation, Nature, 442, 184–187, 2006.

Sachs, L. and Hedderich, J.: Angewandte Statistik – Methodensammlung mit R, Springer Berlin Heidelberg New York, 702 pp., 2006.

Stohl, A., Bonasoni, P., Cristofanelli, P., Collins, W., Feichter, J., Frank, A., Forster, C., Gerasopoulos, E., Gäggeler, H., James, P., Kentarchos, T., Kromp-Kolb, H., Krüger, B., Land, C., Meloen, J., Papayannis, A., Priller, A., Seibert, P., Sprenger, M., Roelofs, G. J., Scheell, H., E. Schnabel, C., Siegmund, P., Tobler, L., Trickl, T., Wernli, H., Wirth, V., Zanis, P., and Zerefos, C.: Stratosphere-troposphere exchange: A review, and what we have learned from STACCATO, J. Geophys. Res., 108(D12), 8516, doi:10.1029/2002JD002490, 2003.

Thouret, V., Marenco, A., Logan, J., Nedelec, P., and Grouhel, C.: Comparisons of ozone measurements from the MOZAIC airborne program and the ozone sounding network at eight locations, J. Geophys. Res., 103(D19), 25 695–25 720, 1998.

Thouret, V., Cammas, J.-P., Sauvage, B., Athier, G., Zbinden, R., Nedelec, R., Simon, P., and Karcher, F.: Tropopause referenced ozone climatology and inter-annual variability (1994–2003) from the MOZAIC programme, Atmos. Chem. Phys., 6, 1033–1051, 2006,
http://www.atmos-chem-phys.net/6/1033/2006/.

Zöger, M., Afchine, A., Eicke, N., Gerhards, M.-T., Klein, E., McKenna, D. S., Mörschel, U., Schmidt, U., Tan, V., Tuitjer, F., Woyke, T., and Schiller, C.: Fast in-situ stratospheric hygrometers: A new family of balloon-borne and airborne Lyman α photofragment fluorescence hygrometers, J. Geophys. Res., 104, 1807–1816, 1999.