Atmos. Meas. Tech. Discuss., 5, 8241–8269, 2012
www.atmos-meas-tech-discuss.net/5/8241/2012/
doi:10.5194/amtd-5-8241-2012

Atmospheric Measurement Techniques Discussions

Climatologies from satellite measurements: the impact of orbital sampling on the standard error of the mean

M. Toohey[1] **and T. von Clarmann**[2]

[1]GEOMAR Helmholtz-Centre for Ocean Research Kiel, Kiel, Germany
[2]Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Karlsruhe, Germany

Received: 31 October 2012 – Accepted: 1 November 2012 – Published: 9 November 2012

Correspondence to: M. Toohey (mtoohey@geomar.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Climatologies of atmospheric observations are often produced by binning measurements according to latitude, and calculating zonal means. The uncertainty in these climatological means is characterized by the standard error of the mean (SEM). However, the usual estimator of the SEM, i.e. the sample standard deviation divided by the square root of the sample size, holds only for uncorrelated randomly sampled measurements. Measurements of the atmospheric state along a satellite orbit cannot always be considered as independent because (a) the time-space interval between two nearest observations is often smaller than the typical scale of variations in the atmospheric state, and (b) the regular time-space sampling pattern of a satellite instrument strongly deviates from random sampling. We have developed an experiment where global chemical fields from a chemistry climate model are sampled according to real sampling patterns of satellite-borne instruments. As case studies, sampling patterns of the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) and Atmospheric Chemistry Experiment Fourier-Transform Spectrometer (ACE-FTS) satellite instruments are used to iteratively subsample the model O_3 fields and produce empirical estimates of the standard error of monthly mean zonal mean model O_3 in 5° latitude bins. We find that generally the classic SEM estimator is a conservative estimate of the SEM, i.e. the empirical SEM is often less than the classic estimate. Exceptions occur in instances where the zonal sampling distribution shows non-uniformity with a similar zonal structure as variations in the sampled field, leading to maximum sensitivity to arbitrary phase shifts between the sample distribution and sampled field. The occurrence of such instances is thus very sensitive to slight changes in the sampling distribution, and to the variations in the measured field. This study highlights the need for caution in the interpretation of the oft-used classically computed SEM, and outlines a relatively simple methodology that can be used to assess one component of the uncertainty in monthly mean zonal mean climatologies produced from measurements from satellite-borne instruments.

AMTD
5, 8241–8269, 2012

1 Introduction

Atmospheric observations are often averaged within time-space intervals, such as calendar months and latitude bands, producing so-called "climatologies" (e.g. Grooß and Russell III, 2005; Hegglin and Tegtmeier, 2011). While the motives behind the construction of such climatologies can be simply pragmatic – for instance to simplify comparison with similarly averaged model fields – averaging does have the advantageous effect of reducing the impact of random variations present in individual measurements due to measurement errors and natural variability. The standard error of the mean (SEM) is a statistical quantity which quantifies the the random error in the calculated mean value.

In general terms, the standard error describes the random error of an estimate based on limited sampling of a population. For example, the SEM describes the potential variation of a sample mean of n samples if other, equally probably sets of n samples were drawn instead. The "classic" and oft-used SEM estimator is given by the standard deviation of the sample divided by the square root of the sample size, however, this estimator is only valid when the measurements are uncorrelated. Consideration of correlations of measured data is a standard in various applications of statistical estimators inferred from atmospheric measurements, e.g. Jones et al. (1997) consider inter-site correlations in the estimation of global mean temperatures; Weatherhead et al. (1998) present a scheme to consider autocorrelations in estimating uncertainties in trends; and von Clarmann et al. (2010) propose a generic approach to consider arbitrary correlations in trend estimation. For the SEM, one of the most fundamental estimators of a finite sample of atmospheric data, little literature is available.

Correlations in atmospheric measurement sets depend upon the underlying time-space correlations of the atmosphere, and the time-space sampling patterns of the measurements themselves. Observational data sets from satellite instruments have distinct sampling patterns which depend on the orbit and measurement technique of the instrument. Different sampling patterns can lead to differences in the means of two data sets: in this case, the difference is referred to as a sampling bias. For example,

AMTD
5, 8241–8269, 2012

Aghedo et al. (2011) have examined the role of sampling in biasing monthly mean values of satellite-based measurements of tropospheric chemical species and temperature. However, the potential impact that sampling may have on the SEM of atmospheric climatologies has not to our knowledge been formally addressed.

We present an experiment designed to assess the impact of time-space sampling patterns on the SEM of climatologies built from satellite-based atmospheric measurement sets. This exercise makes use of model fields from a coupled chemistry climate model, and subsamples the model data based on the sampling pattern of a satellite instrument. Assuming that the model accurately reproduces, in a statistical sense, the correlations of the true atmosphere on scales larger than the horizontal footprint of the satellite measurements, results from this experiment can be used to draw some general statements about the quality of the SEM estimates usually produced from measurements.

2 Theory and methodology

Given a set of N measurements x_n, the sample mean $\bar{x}$ and sample standard deviation $\hat{\sigma}$ are calculated as

$$\bar{x} = \frac{\sum_{n=1}^{N} x_n}{N} \tag{1}$$

and

$$\hat{\sigma} = \sqrt{\frac{\sum_{n=1}^{N} (x_i - \bar{x})^2}{(N-1)}} \tag{2}$$

respectively. The sample mean is an estimate of the population mean, while the sample standard deviation characterizes the scatter of the measured data and thus includes both the natural variability within the population and the measurement error.

AMTD
5, 8241–8269, 2012

The sample mean is intentionally calculated without consideration of any predicted measurement error for weighting purposes, since measurement error can be a function of geolocation and thus could bias the mean: e.g. measurement errors of gases measured by infrared emission are usually smaller when the atmosphere is warmer, and inverse weighting of the measurements by the measurement error variance would bias the sample mean such that it is more representative for the warmer parts of the atmosphere.

Each single measurement differs from the sample mean by some amount due to natural variability and measurement error. Treating each such deviation as an "error", and the sample standard deviation as an estimate of the average "error", the calculation of the SEM follows directly from generalized Gaussian error propagation:

$$\mathrm{SEM}^2 = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right) \begin{pmatrix} \hat{\sigma}^2, & \ldots, & r_{1,N}\hat{\sigma}^2 \\ \vdots & \ddots, & \vdots \\ r_{N,1}\hat{\sigma}^2, & \ldots, & \hat{\sigma}^2 \end{pmatrix} \begin{pmatrix} \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{pmatrix} \tag{3}$$

$$= \hat{\sigma}^2 \left(\frac{1 + (N-1)\bar{r}}{N}\right), \tag{4}$$

where $\bar{r}$ is the average correlation coefficient between the measurements of the sample (cf. Jones et al., 1997). Defining

$$k = \sqrt{1 + (N-1)\bar{r}}, \tag{5}$$

the SEM can be written as

$$\mathrm{SEM} = \frac{\hat{\sigma}}{\sqrt{N}} k. \tag{6}$$

With $\bar{r} = 0$, i.e. independent uncorrelated measurements (both in terms of measurement error and natural variability), $k = 1$ and the expression for the SEM simplifies to

AMTD
5, 8241–8269, 2012

its common estimator $\hat{\sigma}/\sqrt{N}$. When the average correlation between measurements is positive, $k > 1$ and the SEM is greater than $\hat{\sigma}/\sqrt{N}$, i.e. the usual estimator which assumes independent measurements can be seen to underestimate the true SEM. Likewise, when the average correlation between measurements is negative (but no less than $-1/(N-1)$), $0 < k < 1$ and the usual estimator, $\hat{\sigma}/\sqrt{N}$, is an overestimate of the true SEM.

In this paper we indirectly assess the role of the mean correlation coefficient $\hat{r}$ between measurements for example satellite-borne instrument sampling patterns. We do so by producing an empirical estimate of the SEM. Here we take advantage of the fact that the SEM can be defined as the standard deviation of all possible sample means (of a given size) drawn from the population. Firstly, we subsample model fields based on the sampling pattern of a satellite instrument. Sample means for each latitude bin are calculated from the subsampled model data for this sampling pattern. Then, we produce an "equivalent" sampling pattern, which reproduces the most important features of the sampling (latitude and local solar time) but is randomly shifted in longitude (and universal time such that local solar time is held constant). Each such equivalent sampling pattern can be thought of as resulting from a satellite instrument which has the exact same orbit as the original, except with a different position along the orbit at any point in time. Performing the equivalent sampling a number of times (J), we produce an ensemble of equally probable sample means for each bin ($\bar{x}_j$). The expected value of the sample mean $\langle \bar{x}_j \rangle$ is taken to be the ensemble mean of the sample means, the error of each ensemble mean is thus $\bar{x}_j - \langle \bar{x}_j \rangle$ and the SEM is:

$$\mathrm{SEM}^2 = \sum_{j=1}^{J} \frac{(\bar{x}_j - \langle \bar{x}_j \rangle)^2}{J-1}. \quad (7)$$

Another option would be to replace the expectation value $< \bar{x} >$ in Eq. (7) by the true average of all modeled values in the latitude/time bin under consideration. In this case,

AMTD
5, 8241–8269, 2012

the resulting standard error of the mean would also include any potential sampling bias, while our analysis aims at the assessment of the random error.

3 Case Studies

3.1 Sampling patterns

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) is a mid-infrared Fourier transform limb emission spectrometer designed and operated for measurement of atmospheric trace species from space (Fischer et al., 2008). MIPAS passes the equator in a southerly direction at 10:00 a.m. local time 14.3 times a day, observing the atmosphere during day and night with global coverage from pole to pole.

Two different sampling patterns from MIPAS mission are used here. From July 2002 to March 2004, MIPAS operated at full spectral resolution, and recorded profiles of limb spectra every 90 s, corresponding to an along-track sampling of approximately 500 km and providing about 1000 vertical profiles per day in its standard observation mode. The latitudes of each MIPAS profile spectra were nominally fixed, i.e. for the majority of orbits, profiles were repeatedly measured at specific latitudes. The sampling pattern associated with this period of high spectral resolution MIPAS measurements is referred to as MIPAS-HR.

Due to problems with the interferometer mirror slide system, MIPAS performed few operations from April to December 2004. In January 2005 regular observations resumed, but with reduced duty cycle and a reduced spectral resolution. Lower spectral resolution measurements take less time to perform, and as a result, vertical and horizontal measurement frequency was increased during this time period compared to the former "high"-spectral resolution period, with horizontal measurement density increasing by about 20 %. The latitudes of measured profiles were not fixed during this period. The sampling pattern for this period of low spectral resolution measurements is referred to as MIPAS-LR. (Note that the identifiers HR and LR refer to high and low

AMTD
5, 8241–8269, 2012

spectral resolution, respectively, which correspond in contrast to low and high time-space sample resolution, respectively.)

We have taken actual measurement locations for MIPAS measurements from scan pattern cycles 13 and 75, corresponding to 13 January–17 February 2003 and 22 December 2008–26 January 2009 as example sample patterns for MIPAS-HR and MIPAS-LR periods respectively. Since there are no drastic differences between the month-to-month sampling patterns of MIPAS (aside from the change between the HR and LR sampling patterns), the sampling patterns from these scan pattern cycles have been used to define the example sampling patterns for all calendar months for the two periods. It should be noted that since these are actual measurement locations, there are some deviations from the nominal sampling patterns, e.g. sample locations removed because of poor data quality or retrieval problems.

Sampling characteristics for the MIPAS sampling patterns are shown in Figs. 1 and 2. Example daily sample locations for MIPAS-HR (Fig. 1a) span 87.10° S to 89.25° N latitude, with approximately 1000 sample locations per day. Over a full month of sampling, around 800 samples are collected within 5° latitude bins (Fig. 1b). Since the latitudinal spacing of consecutive measurements is approximately 5° (median difference of 4.77°), consecutive measurements are generally not grouped into the same latitude bin (exceptions occuring within the 5–10° N and 85–90° N bins), therefore consecutive measurements within one latitude bin occur during ascending and descending portions of a single orbit, and as a result, the longitude spacing between consecutive measurements within one latitude bin is approximately 180° longitude (median difference of 177°). Taking an example latitude bin, 55–60° N, Fig. 1c shows the time (decimal Julian day) versus longitude sampling pattern, which shows remarkable uniformity. A histogram of sample count per 30° longitude bin for the 55–60° N latitude bin shows a uniform distribution of longitudinal sampling (Fig. 1d), with between 60 and 70 samples per 30° longitude bin. A similar uniforminty of sampling pattern is found in the example SH latitude bin of 55–60° S (Fig. 1e, f).

AMTD
5, 8241–8269, 2012

For MIPAS-LR sampling, samples are more closely spaced in time and space, and no longer fixed to certain latitudes. The number of monthly samples within 5° latitude bins is around 1000 (Fig. 2b), with some variation from bin to bin due to the more random nature of the latitudinal sampling. Since the time-space distance between samples is less than for the MIPAS-HR sampling case (3.6° latitude median difference between consecutive samples), consecutive measurements from the MIPAS orbit are occasionally binned within a single 5° latitude bin. For example, approximately 1/4 of samples within the 55–60° N latitude bin are "double samples", i.e. two samples with small differences in time and space between them. In some cases, double samples within a latitude bin occur preferentially within certain longitude ranges, leading to non-uniformity in the monthly zonal sample distribution. This is the case for example in the 55–60° S bin (Fig. 2e, f), with a notable excess of samples in the Eastern Hemisphere (0–180° longitude). Within the 55–60° N bin, however, the double samples are more randomly (and uniformly) distributed (Fig. 2c), and the resulting zonal sample distribution (Fig. 2d) is of similar uniformity as that of MIPAS-HR.

The Atmospheric Chemistry Experiment-Fourier Transform Spectrometer (ACE-FTS), on board the SCISAT-1 satellite, uses mid-infrared solar occultation to investigate the chemical composition of the atmosphere (Bernath et al., 2005). The SCISAT-1 satellite was launched on 12 August 2003 and began routine measurements on 21 February 2004. The ACE-FTS measures approximately 15 sunrise and 15 sunset occultations per day. A high inclination (74°), circular low- earth orbit (650 km) leads to global coverage of ACE-FTS measurements, but with almost 50 % of the occultation measurements made by the ACE-FTS are at latitudes of 60° and higher. The latitudes of the ACE-FTS sunrise and sunset samples vary with time: global latitude coverage is achieved over a period of approximately three months.

For the sampling exercise presented here, we use the ACE-FTS sampling locations from the year 2005, and examine the months March and April as example cases – March gives reasonable coverage of both the southern and northern mid- and high-latitudes, while April sampling covers the tropics and subtropics. Characteristics of the

AMTD
5, 8241–8269, 2012

sample patterns for ACE-FTS in March and April are shown in Figs. 3 and 4. For March, highest sample density is found in the high latitudes, which is a product of design, the ACE mission being focused on obtaining measurements during the polar winters and springs when O_3 destruction chemical processes are underway (Bernath et al., 2005). Sample counts within 5° latitude bins for March range between ∼10 to ∼80 depending on latitude within this month. For the example latitude bin of 55–60° N, ACE-FTS samples are collected over a few days near the end of the month (Fig. 3c) while in the 55–60° S bin, ACE-FTS samples are collected over a few days at the beginning of the month (Fig. 3e). Within this time frame, the samples circle the Earth in terms of longitude, and the distribution of longitudes sampled is thus relatively uniform given such small sample size, (Fig. 3d, f), with non-uniformity occurring because of missing measurements or overlap of longitudinal sampling cycles. At latitudes higher than the 55–60° bands shown here, sampling density increases substantially (Fig. 3b), and the zonal distributions become more uniform, while at lower latitudes, the opposite is true.

In April, the ACE-FTS sunrise and sunset sampling patterns cross through the tropics (Fig. 4a). Taking 15–20° N as an example latitude bin in the tropics, we see that samples for this bin are composed of measurements at the beginning and end of the month (Fig. 4c). The longitude spacing between consecutive measurements is nominally 24.5°, and samples are collected in this bin over ∼2 days, long enough for the samples to cover the full zonal band, leading to 2 or more samples within 9 out of the 12 30° longitude bins shown in Fig. 4d. There are also a handful of samples within April in the SH high latitudes. Taking the example band of 70–75° S, these samples are collected near the end of April (Fig. 4e), and are notably non-uniform in their longitudinal distribution (Fig. 4f).

3.2 Model description

The CMAM is an extended version of the Canadian Centre for Climate Modelling and Analysis spectral general circulation model. The dynamical core and chemistry scheme are described by Beagley et al. (1997) and de Grandpré et al. (1997) respectively.

AMTD
5, 8241–8269, 2012

Simulated chemical fields from the last ten years (1995–2004) of the CMAM REF1 simulation described by Eyring et al. (2006) are used here. The chemical fields are available for every model gridpoint, in intervals of 18 h.

The distributions of chemical species in the CMAM have been seen to generally compare well with observations (e.g. de Grandpré et al., 2000; Farahani et al., 2007; Hegglin and Shepherd, 2007; Jin et al., 2005, 2009; Melo et al., 2008). While this version of CMAM does not simulate the quasi-biennial oscillation and thus underestimates interannual variability in the tropics, the intra-month variability appears to be of realistic magnitude (see Chapter 7 of SPARC CCMVal, 2010; Toohey et al., 2010). The persistence (i.e. autocorrelation) of zonal mean O_3 anomalies in CMAM agrees extremely well with observations, with interannual anomalies established through winter and spring persisting with very high correlation coefficients through summer until early autumn (Tegtmeier and Shepherd, 2007).

Figure 5 shows the monthly mean zonal mean O_3 distribution for March, as well as the standard deviation (SD) of all data for each latitude and height. Maxima in short-term (intra-monthly) O_3 variability are found generally where spatial gradients in O_3 are strong. Variability is generally weak during summer months, therefore examining variability around the equinoxes allows for a case when there is appreciable variability in both hemispheres.

3.3 Results

3.3.1 MIPAS

Figure 6a–c shows results of the resampling exercise for MIPAS-HR sampling of O_3 over the month of March. The classic SEM estimate (Fig. 6a) follows the natural variability of the O_3 field, with maximum values in the mid to high latitudes of the middle stratosphere. Maximum SEM values of ~ 0.03 ppmv correspond roughly to the ratio of maximum model variability (~ 1 ppmv, see Fig. 5) to the square root of the sample size (~ 800).

AMTD
5, 8241–8269, 2012

The SEM estimated through the ensemble resampling technique (Fig. 6b) is notably smaller than the classic SEM estimator for almost all latitudes and heights. Given the rather uniform sampling pattern of MIPAS-HR, the sample mean is quite insensitive to shifts in the longitudinal distribution. As a result, k values (Fig. 6c) are less than 1 throughout almost all of the stratosphere. This means that the classic SEM estimator almost always overestimates the true SEM given the MIPAS-HR sampling, in other words, in this case the classic SEM is a conservative estimate of the true SEM.

Figure 6d–f shows results of the resampling exercise for MIPAS-LR sampling of O_3 over March. Due to its larger sample sizes, the classic SEM estimate (Fig. 6d) for MIPAS-LR sampling is slightly smaller than that for MIPAS-HR. As for MIPAS-HR sampling, the SEM estimated through the ensemble resampling technique for MIPAS-LR sampling (Fig. 6e) is in general less than the classic estimator, with k values less than 1 (Fig. 6f). However, the results for MIPAS-LR sampling are also generally in closer agreement with the classic estimator for MIPAS-LR sampling than for MIPAS-HR, with approximately 26 % of the points shown in Fig. 6f having k values between 0.8 and 1.2, compared to only 3 % for MIPAS-HR (Fig. 6c). We interpret this result as a consequence of differences in the sampling patterns of the two MIPAS periods. With sampled latitudes within each bin varying from orbit to orbit, and the closer latitude spacing leading to occasional "double samples", MIPAS-LR sampling is a closer approximation of the random sampling assumed by the classic SEM estimator, therefore it stands to reason that the SEM estimated through the resampling exercise for MIPAS-LR are in closer agreement with the classic SEM estimator than for MIPAS-HR.

For MIPAS-LR, there are also occasions when the classic SEM estimator underestimates the SEM computed through the resampling exercise, leading to k values greater than one, e.g. at isolated latitude bins in the mid-to-high latitudes of both hemispheres, and in the lower tropical stratosphere. In these cases the sample mean is quite sensitive to longitudinal shifts in the samping pattern, and as a result the SEM computed from any one sample set would underestimate the true SEM.

AMTD
5, 8241–8269, 2012

3.3.2 ACE-FTS

Figure 7 shows results of the ensemble resampling exercise for ACE-FTS sampling of O_3 over month of March and April. With much lower sample sizes than for MIPAS, the classic SEM estimator (7a and d) gives notably larger values than for MIPAS sampling (note different color scale beween Figs. 6 and 7).

For March sampling, the SEM estimated through the ensemble resampling technique (Fig. 7b) is notably smaller than the classic SEM estimator for almost all latitudes and heights. As a result, k values (Fig. 7c) are less than one throughout much of the stratosphere. As for MIPAS-LR sampling, many k values are relatively close to one, with 30 % of k values between 0.8 and 1.2. There also exist a few isolated instances of positive k values, where the classic SEM estimator is seen to underestimate the SEM estimated through the resampling technique.

For April sampling, the SEM estimated through the ensemble resampling technique (Fig. 7e) is generally close in value or slightly less than the classic estimator, leading to k values approximately equal to or less than one, with 36 % of k values between 0.8 and 1.2. Instances of positive k, where the classic SEM estimator is seen to underestimate the SEM estimated through the resampling technique, are more prevalent than found for March sampling, with large k values found occur in the SH high latitudes and SH subtropics.

3.4 Discussion

Ignoring variations in sampling distribution with time, the resampling exercise used to produce the results in Figs. 6 and 7 can be simplified into the following: for each latitude bin, each resampled monthly mean value can be thought of as a weighted mean of the model zonal O_3 field, where the weights are defined by the zonal distribution of the monthly sample number. Each ensemble member of the resampling exercise is then produced by randomly shifting the zonal sample distribution pattern with respect to the monthly mean O_3 field. Variations in the monthly mean zonal mean sample

AMTD
5, 8241–8269, 2012

means occur based on the relationship between the zonal structures of the O_3 field and the sampling distribution – if either is completely uniform, then shifts in the relative zonal structure will have no effect on the sample mean. Furthermore, the degree to which the ensemble of sample means vary will depend on the similarity between the two distributions: maximum variation between ensemble member means should result when the mean O_3 field and sample distribution have the same zonal structure.

This mechanism is explored in Fig. 8 for MIPAS-LR sampling. In Fig. 6f, it was seen that k values greater than one were found in the NH mid to high latitudes, meaning that the SEM estimated though the resampling exercise was found to be larger than that estimated through the classical estimator. As an example case of large k values, we examine the 50–55° N latitude bin at 60 hPa, a location of a local maximum in the k values shown in Fig. 6f. Figure 8 shows a histogram of the sample distribution for the MIPAS-LR sampling pattern for the 50–55° N latitude bin separated into 30° longitude bins. At this latitude, the MIPAS-LR sampling pattern has notable zonal structure, with a maximum and minimum in sample density separated by approximately 120°. Also shown is the monthly mean zonal O_3 anomaly field for the model latitude of 52° at 60 hPa. The O_3 field has been shifted in longitude to produce maximum (solid) and minimum (dashed) values of a weighted mean of the O_3 field calculated by using the sample distribution as weights. There is a clear correspondence between the structures of the sampling distribution and the O_3 field, and it follows that the sensitivity of the sample mean to the phasing of sample distribution is related to the similarity between the two distributions.

In this way, the results of the sampling exercise can be seen to be related to the correlations between the zonal structures of the sample distribution and the measured field. The results of the sampling exercise can also be interpreted in terms of correlations between individual measurements. Equation (5) shows the relationship between k and the mean correlation coefficient between all pairs of measurements. k values less than 1 imply a negative mean correlation coefficient. For measured fields with a periodic structure in longitude, sampled with very uniform sampling, we might expect a negative

AMTD
5, 8241–8269, 2012

correlation since every measurement is balanced by a corresponding measurement on the other side of the Earth. For non-uniform sampling, e.g. when some measurements are clustered around a certain longitude, then the similarity of the O_3 field measured around this cluster leads to an increase in the mean correlation, leading to mean correlation coefficients approaching zero or reaching positive values, and correspondingly k values of 1 or greater, as was found for the MIPAS-LR sampling at certain locations.

Figure 9 shows a similar explanatory plot for ACE-FTS. Maximum k values for ACE-FTS sampling of April O_3 were found in the high SH latitudes (Fig. 7f). At these latitudes, the very large k values are the result of a highly non-uniform sampling pattern with respect to longitude (as shown in Fig. 4), with most samples clustered within 120° longitude of each other. As a result of this sampling pattern, any non-uniformity in the measured O_3 field will lead to variations in the monthly mean zonal mean values produced by each realization of the ensemble resampling, and as a result the SEM estimated by the resampling technique is large. Figure 9 shows that the ACE-FTS sampling distribution for the 70–75° S latitude band, as well as the O_3 anomaly field for this latitude at 10 hPa over the days of ACE-FTS sampling of this latitude. O_3 anomalies show a clear zonally periodic variability at this latitude and height, and as a result the sample mean is sensitive to the phase shift of the non-uniform sample distribution.

4 Conclusions

The usual way to calculate the standard error of the mean by division of the sample standard deviation by the square root of the sample size is exact only if the elements within the sample are uncorrelated. Satellite measurement data sets, however, are not random samples because measurement locations are the result of such factors as the regular satellite orbit and and limitations of measurement frequency. By subsampling model data according to the real sampling patterns of two modern satellite-borne instruments, this effect has been assessed for a number of test cases. Two competing mechanisms have been discovered: in some cases the true standard error is smaller

AMTD
5, 8241–8269, 2012

than that estimated by the classical estimator. This effect is most pronounced when the sample distribution is quite uniform with respect to longitude. Since variations of stratospheric trace gases such as O_3 typically follow wave like patterns along zonal bands, within a zonal band, uniform sampling leads to negative mean correlation coefficients since each too-low measurement is compensated by a too-high measurement. As a result, the classical SEM estimator, which assumes random sampling and not the highly uniform sampling of the satellite instrument overestimates the true standard error of the mean.

In other cases the true standard error is larger than that estimated by the classical SEM estimator. This applies particularly to cases where the zonal sampling distribution is non-uniform. If the non-uniformity of the sampling pattern is of similar zonal structure to variations in the measured field, then the measured zonal mean is sensitive to arbitrary phase shifts between the sampling pattern and the measured field. As a result, the SEM is larger than that estimated by the classic estimator. In this case, the similar zonal structure of the sampling distribution and the measured field can be understood to result in positive mean correlation between samples, which we suggest is an equivalent explanation for the fact that the classical SEM underestimates the true SEM.

For satellite instruments with high sample density, such as MIPAS, isolated intances where the SEM calculated through the classic estimator is on the order of a factor of 2 too small may have very little practical importance. With such large sample sizes, the standard error of climatological means is practically so small that any differences between two such instruments is very likely dominated by systematic errors rather than random errors. However, for instruments with much smaller sample sizes, such as the solar occulation instrument ACE-FTS, proper interpretation of inter-instrument or instrument-model comparisons may rely more heavily on the calculation of a appropriate SEM. In such cases, the results of this study suggest that in order for the classic SEM estimator to be used, a climatology producer should be encouraged to require some degree of zonal uniformity in the sample distribution of measurements used to

AMTD
5, 8241–8269, 2012

calculate a zonal mean. In fact, we find that the classic SEM may be still valid (or even a conservative estimate) for quite small sample sizes (e.g. under 10), as long as the zonal sample distribution is relatively uniform.

Acknowledgements. The authors thank the instrument teams of MIPAS and ACE-FTS for providing sample patterns for use in this study, as well as Andreas I. Jonsson and Theodore G. Shepherd for providing the CMAM chemical fields. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT-1, is a Canadian-led mission mainly supported by the Canadian Space Agency (CSA) and the Natural Sciences and Engineering Research Council (NSERC) of Canada. The CMAM is supported by the Canadian Foundation for Climate and Atmospheric Sciences and the CSA. M. T. acknowledges partial support through the EU project SHIVA (grant number 226224). This work was motivated by discussions within the SPARC Data initiative: the authors thank co-leaders Michaela Hegglin and Susann Tegtmeier, and all members of the SPARC DI team. Also thanks to Kaley Walker for helpful comments on a draft manuscript.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

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. 8244

Beagley, S. R., de Grandpré, J., Koshyk, J. N., McFarlane, N. A., and Shepherd, T. G.: Radiative-dynamical climatology of the first-generation Canadian Middle Atmosphere Model, Atmos. Ocean, 35, 293–331, 1997. 8250

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMazière, M., Drummond, J. R., Dufour, D., Evans, W. F. J., Fast, H., Fussen, D., Gilbert, K., Jennings, D. E., Llewellyn, E. J., Lowe, R. P., Mahieu, E., McConnell, J. C., McHugh, M., McLeod, S. D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C. P., Rochon, Y. J., Rowlands, N.,

AMTD
5, 8241–8269, 2012

Semeniuk, K., Simon, P., Skelton, R., Sloan, J. J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K. A., Walkty, I., Wardle, D. A., Wehrle, V., Zander, R., and Zou, J.: Atmospheric Chemistry Experiment (ACE): mission overview, Geophys. Res. Lett., 32, L15S01, doi:10.1029/2005GL022386, 2005. 8249, 8250

de Grandpré, J., Sandilands, J. W., Mcconnell, J. C., Beagley, S. R., Croteau, P. C., and Danilin, M. Y.: Canadian middle atmosphere model: preliminary results from the chemical transport module, Atmos. Ocean, 35, 385–431, 1997. 8250

de Grandpré, J., Beagley, S. R., Fomichev, V. I., Griffioen, E., McConnell, J. C., Medvedev, A. S., and Shepherd, T. G.: Ozone climatology using interactive chemistry: results from the Canadian Middle Atmosphere Model, J. Geophys. Res., 105, 26475–26492, doi:10.1029/2000JD900427, 2000. 8251

Eyring, V., Butchart, N., Waugh, D. W., Akiyoshi, H., Austin, J., Bekki, S., Bodeker, G. E., Boville, B. A., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deushi, M., Fioletov, V. E., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Grewe, V., Jourdain, L., Kinnison, D. E., Mancini, E., Manzini, E., Marchand, M., Marsh, D. R., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Shepherd, T. G., Shibata, K., Stolarski, R. S., Struthers, H., Tian, W., and Yoshiki, M.: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, J. Geophys. Res., 111, D22308, doi:10.1029/2006JD007327, 2006. 8251

Farahani, E., Fast, H., Mittermeier, R. L., Makino, Y., Strong, K., McLandress, C., Shepherd, T. G., Chipperfield, M. P., Hannigan, J. W., Coffey, M. T., Mikuteit, S., Hase, F., Blumenstock, T., and Raffalski, U.: Nitric acid measurements at Eureka obtained in winter 2001–2002 using solar and lunar Fourier transform infrared absorption spectroscopy: comparisons with observations at Thule and Kiruna and with results from three-dimensional models, J. Geophys. Res., 112, D01305, doi:10.1029/2006JD007096, 2007. 8251

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopman, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmos. Chem. Phys., 8, 2151–2188, doi:10.5194/acp-8-2151-2008, 2008. 8247

Grooß, J.-U. and Russell III, J. M.: Technical note: A stratospheric climatology for O_3, H_2O, CH_4, NO_x, HCl and HF derived from HALOE measurements, Atmos. Chem. Phys., 5, 2797–2807, doi:10.5194/acp-5-2797-2005, 2005. 8243

AMTD
5, 8241–8269, 2012

Hegglin, M. I. and Shepherd, T. G.: O_3–N_2O correlations from the atmospheric chemistry experiment: revisiting a diagnostic of transport and chemistry in the stratosphere, J. Geophys. Res., 112, 19301, doi:10.1029/2006JD008281, 2007. 8251

Hegglin, M. I. and Tegtmeier, S.: The SPARC data initiative, SPARC Newsletter, 36, 22–23, 2011. 8243

Jin, J. J., Semeniuk, K., Jonsson, A. I., Beagley, S. R., McConnell, J. C., Boone, C. D., Walker, K. A., Bernath, P. F., Rinsland, C. P., Dupuy, E., Ricaud, P., De La Noë, J., Urban, J., and Murtagh, D.: Co-located ACE-FTS and Odin/SMR stratospheric-mesospheric CO 2004 measurements and comparison with a GCM, Geophys. Res. Lett., 32, L15S03, doi:10.1029/2005GL022433, 2005. 8251

Jin, J. J., Semeniuk, K., Beagley, S. R., Fomichev, V. I., Jonsson, A. I., McConnell, J. C., Urban, J., Murtagh, D., Manney, G. L., Boone, C. D., Bernath, P. F., Walker, K. A., Barret, B., Ricaud, P., and Dupuy, E.: Comparison of CMAM simulations of carbon monoxide (CO), nitrous oxide (N_2O), and methane (CH_4) with observations from Odin/SMR, ACE-FTS, and Aura/MLS, Atmos. Chem. Phys., 9, 3233–3252, doi:10.5194/acp-9-3233-2009, 2009. 8251

Jones, P. D., Osborn, T. J., and Briffa, K. R.: Estimating sampling errors in large-scale temperature averages, J. Climate, 10, 2548–2568, 1997. 8243, 8245

Melo, S. M. L., Blatherwick, R., Davies, J., Fogal, P., de Grandpré, J., McConnell, J., McElroy, C. T., McLandress, C., Murcray, F. J., Olson, J. R., Semeniuk, K., Shepherd, T. G., Strong, K., Tarasick, D., and Williams-Rioux, B. J.: Summertime stratospheric processes at northern mid-latitudes: comparisons between MANTRA balloon measurements and the Canadian Middle Atmosphere Model, Atmos. Chem. Phys., 8, 2057–2071, doi:10.5194/acp-8-2057-2008, 2008. 8251

SPARC CCMVal: SPARC Report on the Evaluation of Chemistry-Climate Models, edited by: Eyring, V., Shepherd, T. G., Waugh, D. W., SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, available at: http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/ (last access: 20 October 2012), 2010. 8251

Tegtmeier, S. and Shepherd, T. G.: Persistence and photochemical decay of springtime total ozone anomalies in the Canadian Middle Atmosphere Model, Atmos. Chem. Phys., 7, 485–493, doi:10.5194/acp-7-485-2007, 2007. 8251

Toohey, M., Strong, K., Bernath, P. F., Boone, C. D., Walker, K. A., Jonsson, A. I., and Shepherd, T. G.: Validating the reported random errors of ACE-FTS measurements, J. Geophys. Res., 115, D20304, doi:10.1029/2010JD014185, 2010. 8251

AMTD
5, 8241–8269, 2012

von Clarmann, T., Stiller, G., Grabowski, U., Eckert, E., and Orphal, J.: Technical Note: Trend estimation from irregularly sampled, correlated data, Atmos. Chem. Phys., 10, 6737–6747, doi:10.5194/acp-10-6737-2010, 2010. 8243

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, F. E.: Factors affecting the detection of trends: statistical considerations and applications to environmental data, J. Geophys. Res., 103, 17149–17161, 1998. 8243

AMTD

5, 8241–8269, 2012

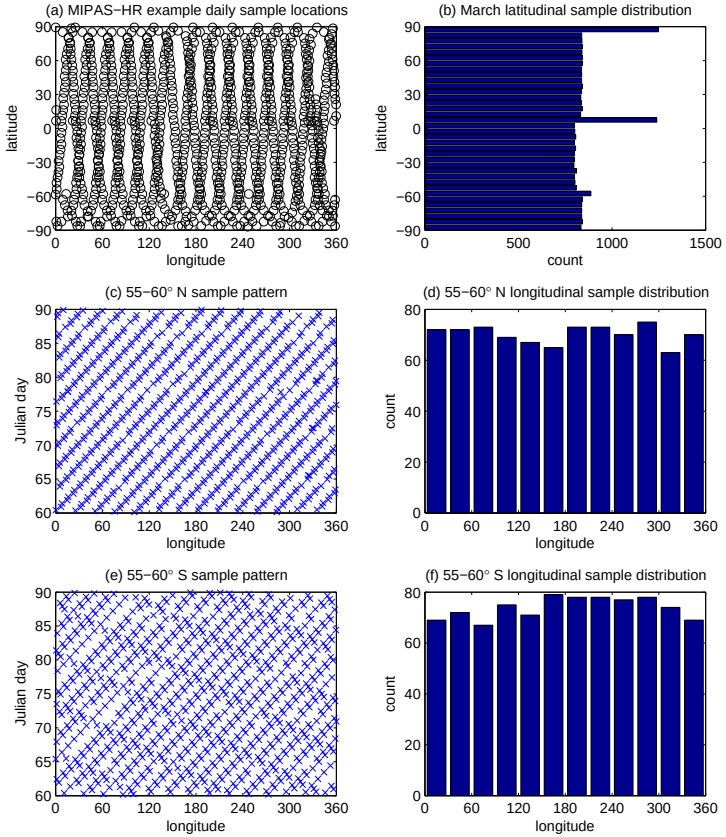

Fig. 1. MIPAS-HR March sampling, approximating MIPAS sampling over the time interval July 2002 to March 2004. **(a)** Example daily sampling spatial pattern, **(b)** monthly sample counts per 5° latitude bin, **(c)** time, longitude pattern of samples and **(d)** zonal sample distribution for the 55–60° N latitude bin, **(e)** time, longitude pattern of samples and **(f)** zonal sample distribution for the 55–60° S latitude bin.

AMTD
5, 8241–8269, 2012

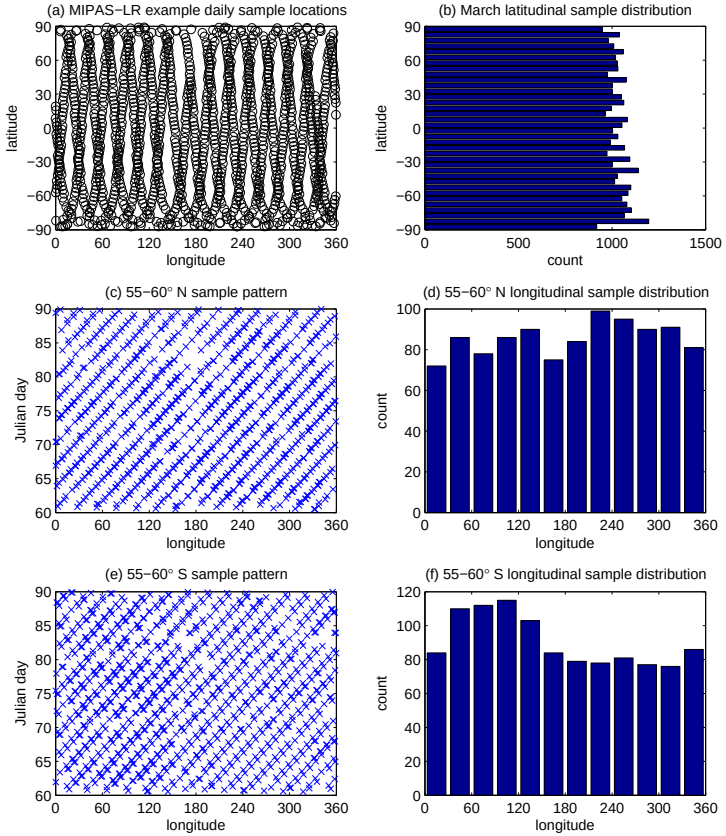

Fig. 2. MIPAS-LR March sampling, appromximating MIPAS sampling from January 2005 until mission conclusion. **(a)** Example daily sampling spatial pattern, **(b)** monthly sample counts per 5° latitude bin, **(c)** time, longitude pattern of samples and **(d)** zonal sample distribution for the 55–60° N latitude bin, **(e)** time, longitude pattern of samples and **(f)** zonal sample distribution for the 55–60° S latitude bin.

AMTD
5, 8241–8269, 2012

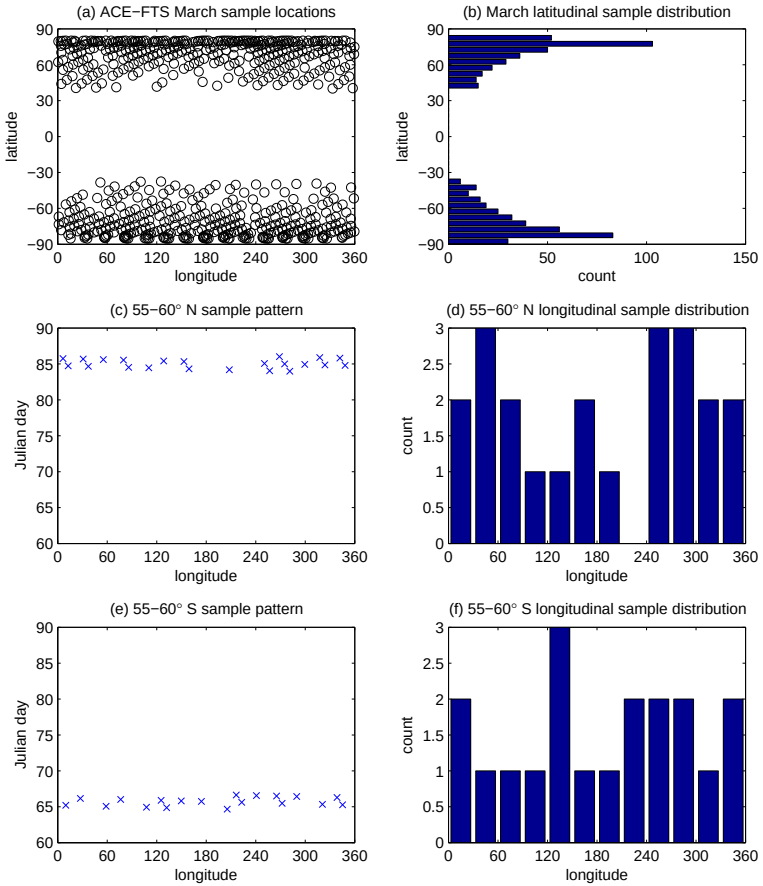

Fig. 3. ACE-FTS sampling in March 2005. **(a)** Full monthly sampling spatial pattern, **(b)** monthly sample counts per 5˚ latitude bin, **(c)** time, longitude pattern of samples and **(d)** zonal sample distribution for the 55–60˚ N latitude bin, **(e)** time, longitude pattern of samples and **(f)** zonal sample distribution for the 55–60˚ S latitude bin.

AMTD
5, 8241–8269, 2012

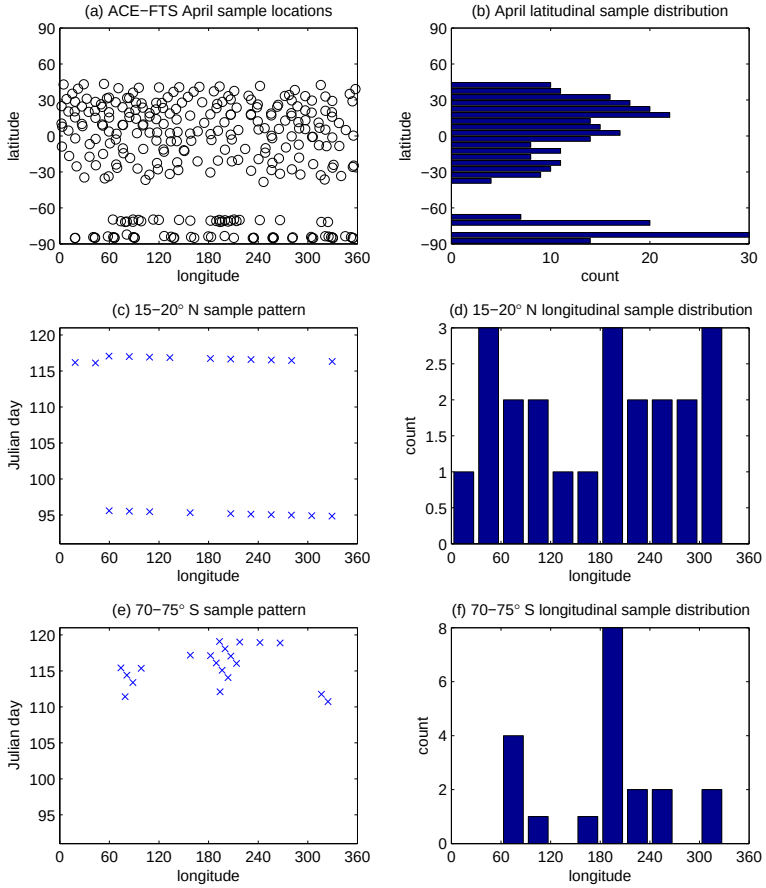

Fig. 4. ACE-FTS sampling from April 2005. **(a)** Full monthly sampling spatial pattern, **(b)** monthly sample counts per 5° latitude bin, **(c)** time, longitude pattern of samples and **(d)** zonal sample distribution for the 15–20° N latitude bin, **(e)** time, longitude pattern of samples and **(f)** zonal sample distribution for the 70–75° S latitude bin.

AMTD
5, 8241–8269, 2012

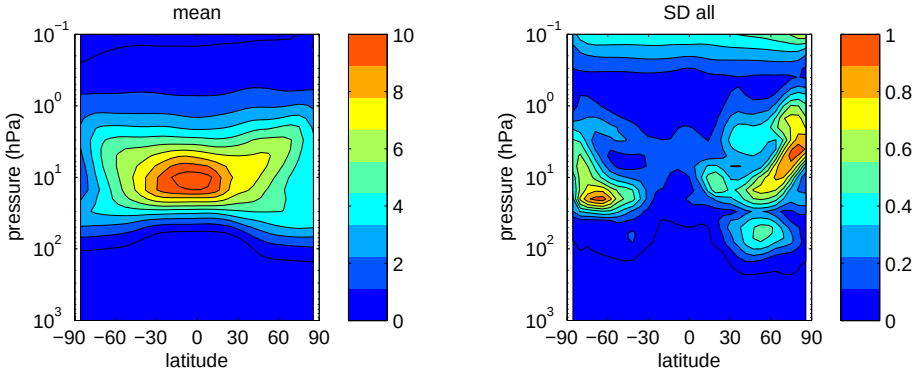

Fig. 5. CMAM March O_3: (left) monthly mean zonal mean O_3 and (right) standard deviation of all monthly values as function of latitude and pressure.

AMTD
5, 8241–8269, 2012

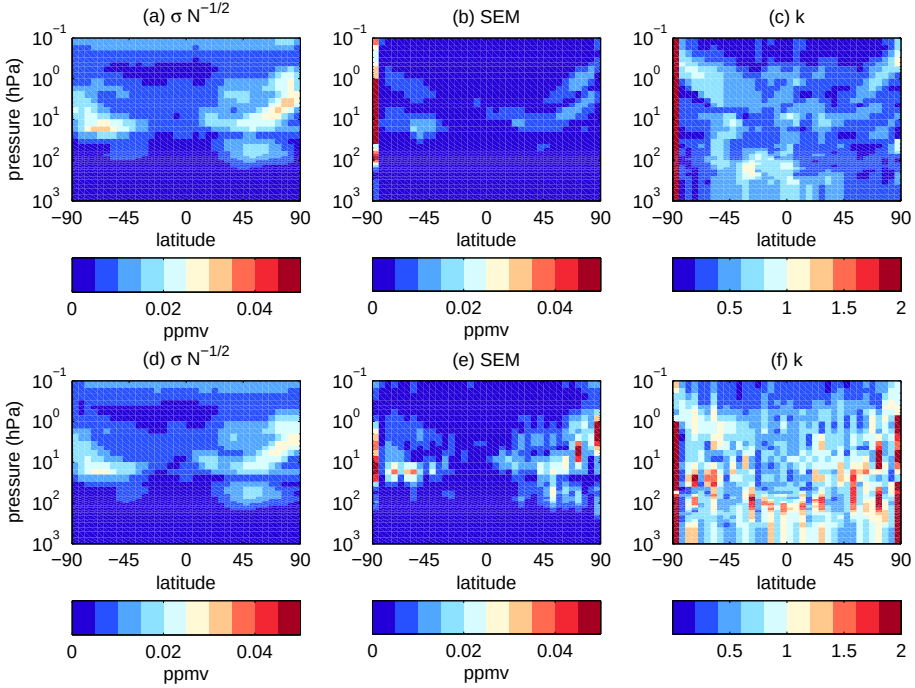

Fig. 6. Pressure latitude sections of zonal mean values of the classic SEM estimator **(a)**, the empirically derived SEM **(b)** and *k* **(c)** based on MIPAS-HR (top) and MIPAS-LR (bottom) sampling of CMAM March O_3.

AMTD
5, 8241–8269, 2012

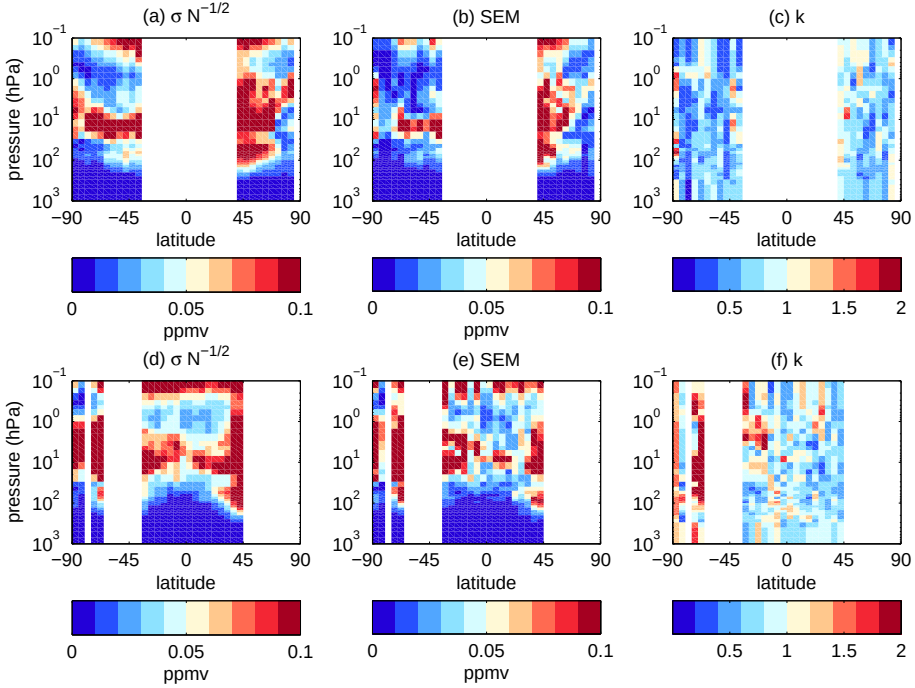

Fig. 7. Pressure latitude sections of zonal mean values of the classic SEM estimator **(a)**, the empirically derived SEM **(b)** and *k* **(c)** based on ACE-FTS sampling of CMAM March (top) and April (bottom) O_3.

AMTD
5, 8241–8269, 2012

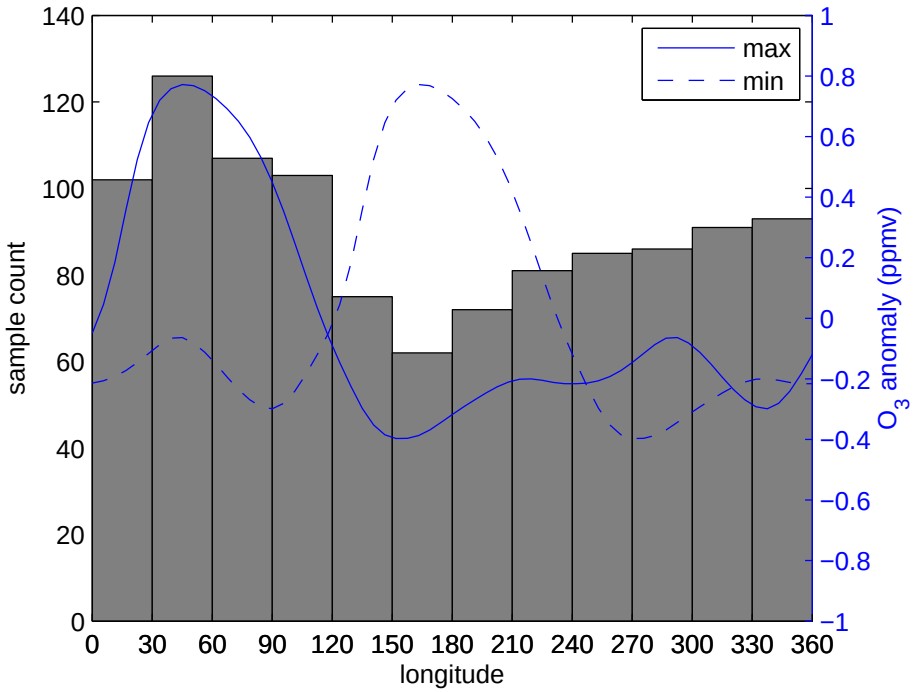

Fig. 8. Zonal MIPAS-LR sampling distribution for March within the 50–55° N latitude bin (gray bars). Also shown is the monthly mean zonal O_3 anomaly field at 52.4° N, at 20 hPa. The O_3 field has been shifted in longitude to produce maximum (solid) and minimum (dashed) values of a weighted mean of the O_3 field using the sample distribution as weights.

AMTD
5, 8241–8269, 2012

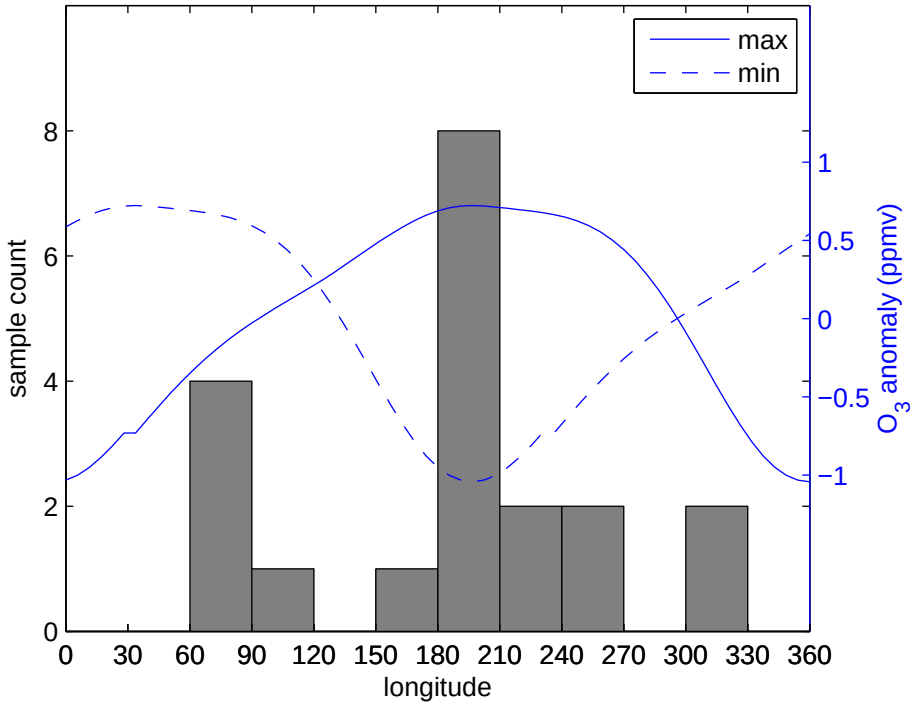

Fig. 9. Zonal ACE-FTS sampling distribution for March within the 70–75° S latitude bin (gray bars). Also shown is the monthly mean zonal O_3 anomaly field at 74.7° N, at 10 hPa. The O_3 field has been shifted in longitude to produce maximum (solid) and minimum (dashed) values of a weighted mean of the O_3 field using the sample distribution as weights.

AMTD
5, 8241–8269, 2012