SPARC Data Initiative: climatology uncertainty assessment

Matthew Toohey (1), Thomas von Clarmann (2), Michaela Hegglin (3), Susann Tegtmeier (1), and the SPARC Data Initiative team

(1) IFM-GEOMAR, Kiel, Germany. (2) Karlsruhe Institute of Technology, Karlsruhe, Germany. (3) University of Toronto, Toronto, Canada.

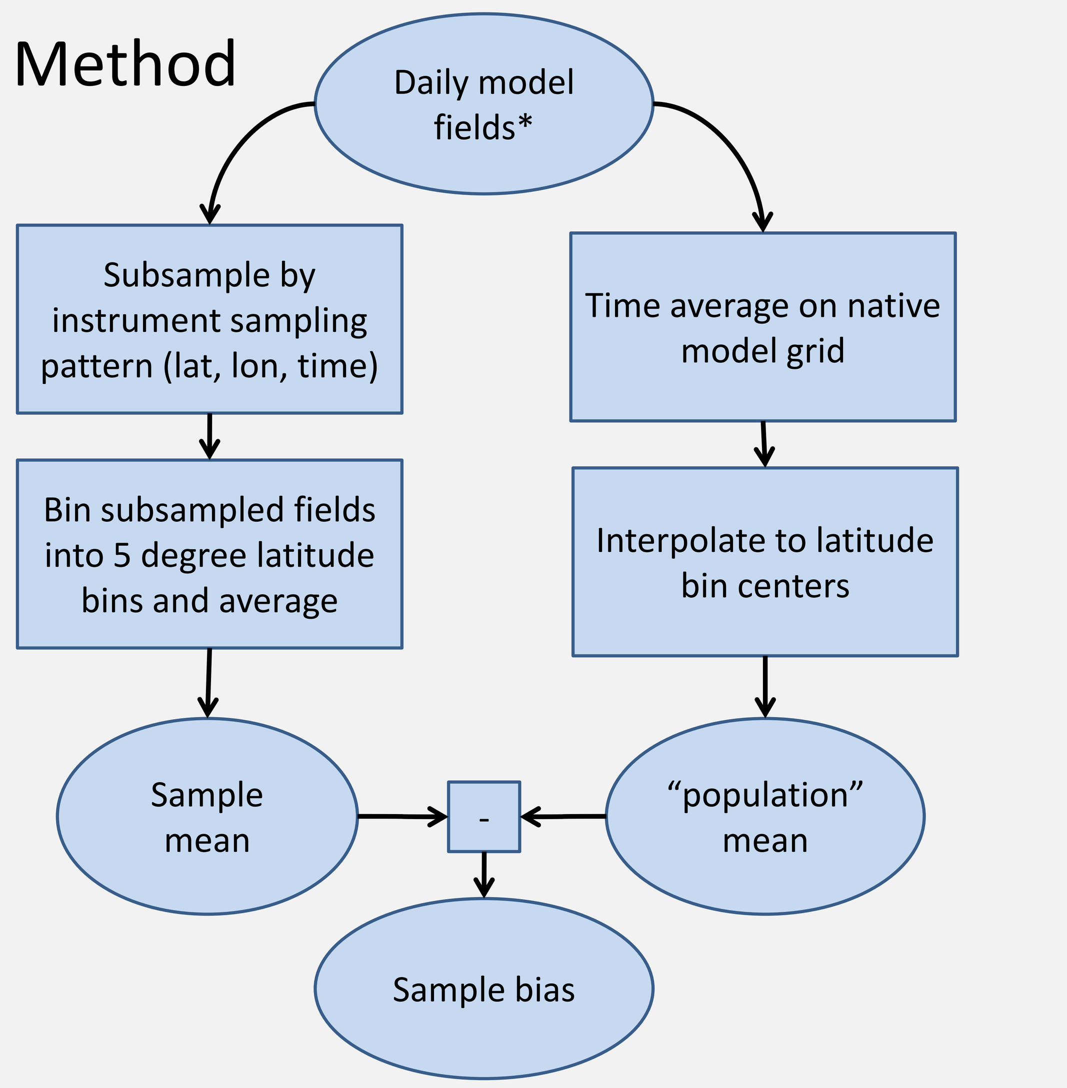
contact: mtoohey@ifm-geomar.de

1. Abstract

The SPARC Data Initiative aims to produce trace gas climatologies for a number of species from a number of instruments. In order to properly compare these climatologies, and interpret differences between them, it is necessary to know the uncertainty in each calculated climatological mean field. The inhomogeneous and finite temporal-spatial sampling pattern of each instrument can lead to biases and uncertainties in the mean climatologies. Sampling which is unevenly weighted in time and space leads to biases between a data set's climatology and the truth. Furthermore, the systematic sampling patterns of some instruments may mean that uncertainties in mean fields calculated through traditional methods that assume random sampling may be inappropriate. We aim to address these issues through an exercise wherein high resolution chemical fields from a coupled Chemistry Climate Model are sub-sampled based on the sampling pattern of each instrument. Climatologies based on the sub-sampled data can be compared to those calculated with the full data set, in order to assess sampling biases. Furthermore, investigating the ensemble variability of climatologies based on sub-sampled fields will allow us to assess the proper methodology for estimating the uncertainty in climatological mean fields.

2. How does instrument sampling affect the sample mean?

Method



*We use here daily mean O_3 fields from a WACCM version 3 simulation , with 1.9° x 2.5° horizontal resolution.

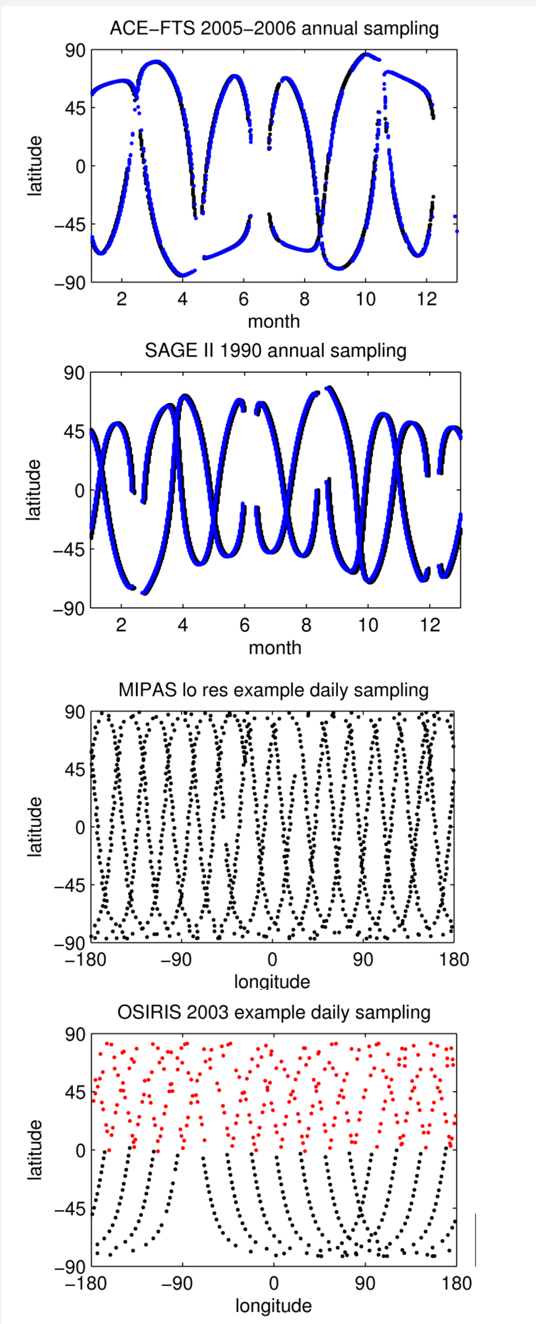


Figure 1: Annual (ACE-FTS, SAGE II) and daily (MIPAS, OSIRIS) spatial sampling patterns for select instruments in the SPARC DI.

Case study: April sampling

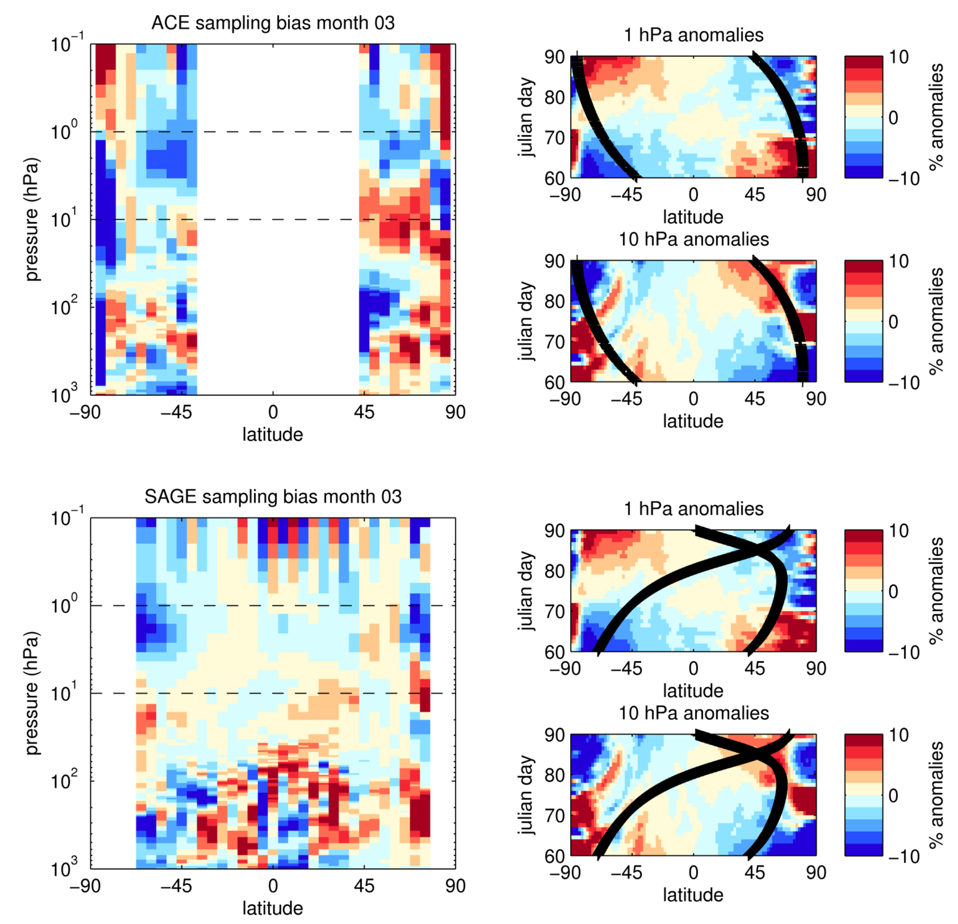


Figure 2: April sampling bias estimates are shown for solar occultation instruments ACE-FTS and SAGE II in terms of percent difference (left). Percent anomalies from monthly mean zonal mean are shown as a function of Julian day for the 1 and 10 hPa surfaces, along with the locations of all measurements for the instrument within the month (right).

- Occultation instruments often sample specific latitudes only at one time of month, leading to bias when the sampled field exhibits sizeable variation over the month.
- Similar sampling patterns can lead to similar sampling biases. As a result, climatological means from these two instruments may agree with each other, but may differ from other instruments with uniform sampling.

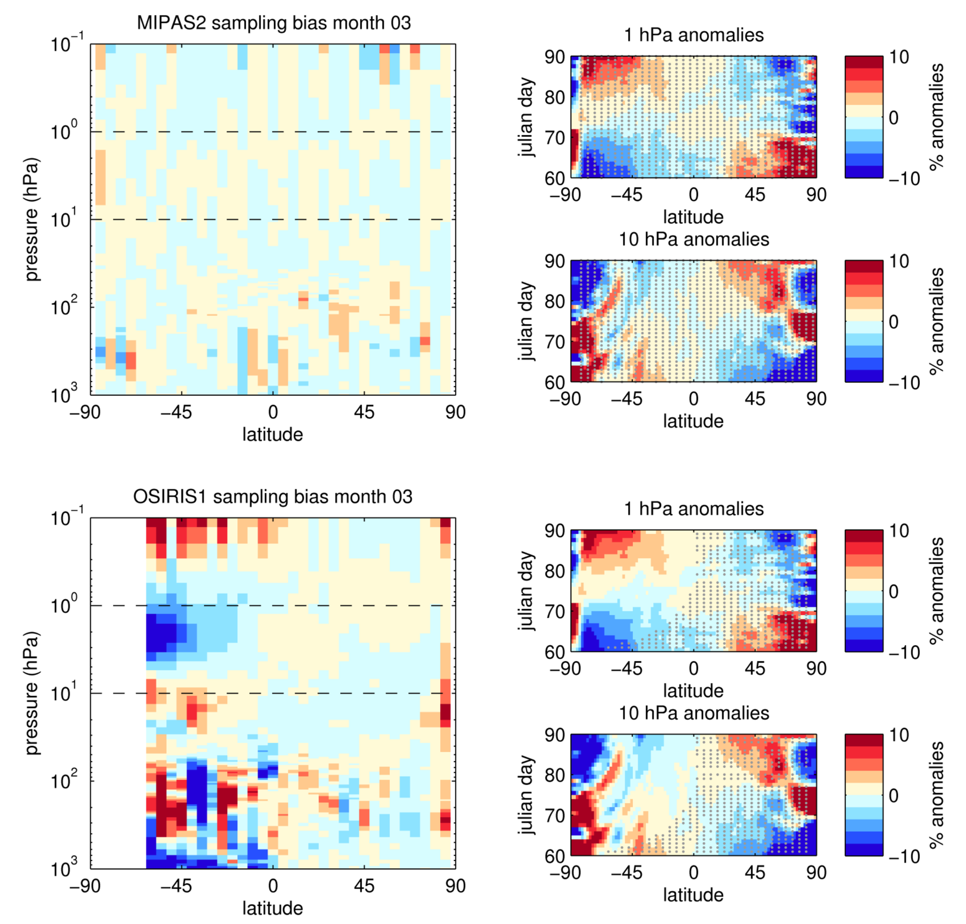


Figure 3: April sampling bias estimates are shown for MIPAS and OSIRIS in terms of percent difference (left). Percent anomalies from monthly mean zonal mean are shown as a function of Julian day for the 1 and 10 hPa surfaces, along with the locations of where at least one measurement was collected within the month (right).

- The relatively uniform sampling of MIPAS leads to very little sampling bias, with values between -1 and +1%.
- The sampling of OSIRIS varies with the season, with dense sampling on the summer hemisphere. In transition months such as April, some latitudes care only sampled at the beginning or end of a month, leading to biases similar to those of the occultation instruments.

3. How does instrument sampling affect the standard error of the mean?

When one does not assume that measurements are independent and uncorrelated, the SEM can be written

$$SEM = \frac{\hat{\sigma}}{\sqrt{N}} k \qquad \text{where} \qquad k = \sqrt{1+(N-1)\bar{r}}$$

and $\bar{r}$ is the average correlation coefficient between the measurements of the sample (Jones et al., 1997). When $\bar{r} = 0$, the measurements are independent and uncorrelated, k = 1, and we get the familiar estimator for the SEM. The factor k thus describes the quality of the standard SEM expression, and that quality depends upon the degree of correlation between the measurements.

By extending the sampling exercise described above, we assess the impact of orbital sampling patterns on the SEM of climatologies built from satellite based atmospheric measurements. Specifically, we estimate k for each latitude height bin of our grid using two methods.

1. Based on the definition of the SEM, we estimate the SEM by sampling the model data with an ensemble of 'equivalent' sampling patterns, which are identical in latitude and time of day but randomly shifted in longitude. The SD of the sample means gives the SEM, and k is calculated as the ratio of the SEM and $\sigma/\sqrt{N}$.
2. k can be calculated explicitly by assessing the correlation between measurements. Using the model data, we calculate the correlation coefficient in time between the model anomalies from zonal mean at each location.

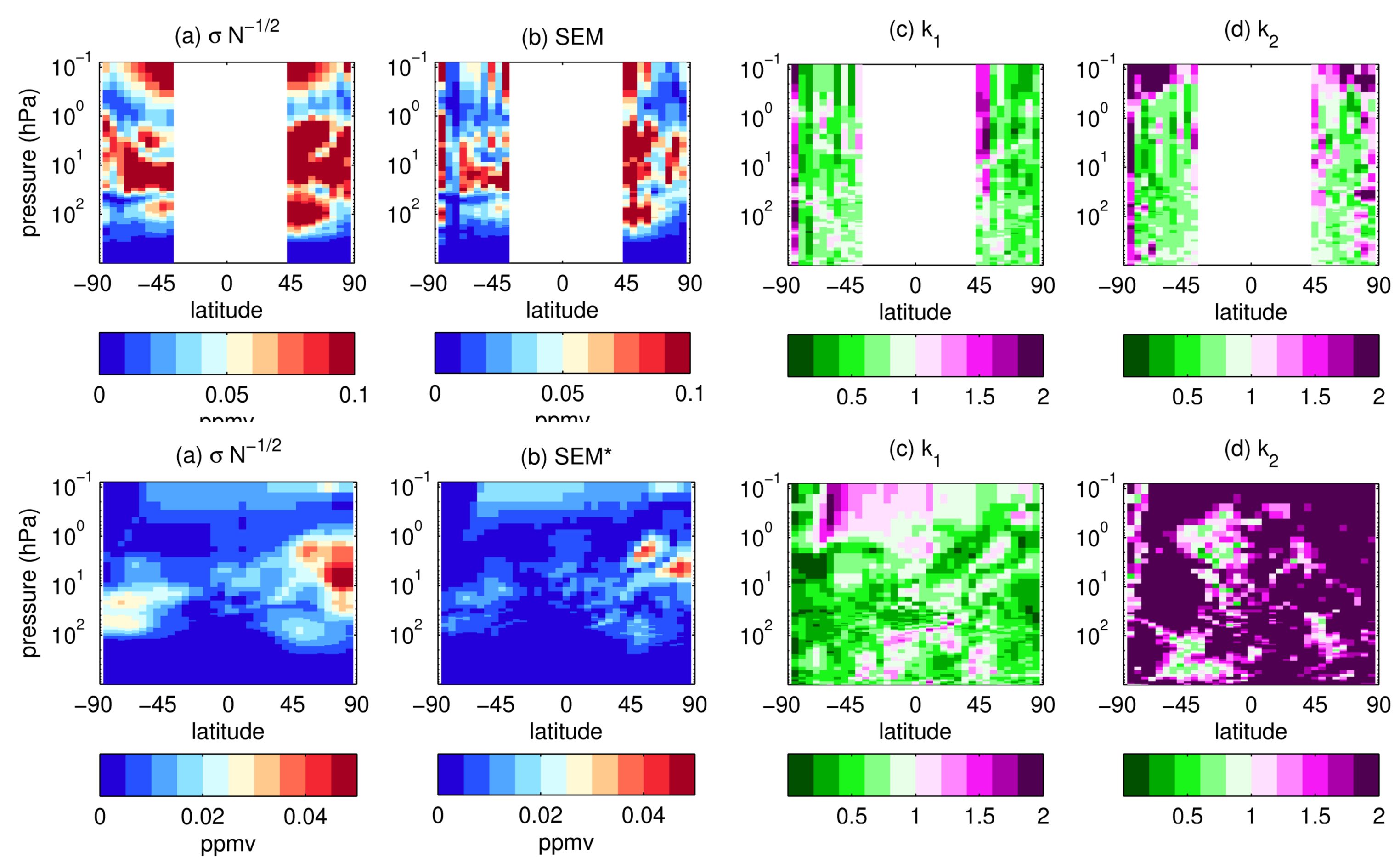


Figure 4: Results of SEM sampling exercise for ACE-FTS sampling in the month of April, and MIPAS sampling for the month of January. Quantities shown : "classic" SEM (a), SEM estimated by ensemble method (b), k calculated as ratio of a and b (c) and k calculated based on correlations of model data (d).

For both ACE-FTS and MIPAS sampling, the SEM estimated through the ensemble sampling experiment is less than the SEM estimate σ/sqrt(N), at most latitudes and heights, leading to k values less than one. Exceptions to this occur at the edges of the latitudinal sampling extent of ACE-FTS, and at heights above 1 hPa for MIPAS. k<1 implies a negative mean correlation coefficient between measurements, which may be possible when a sampling pattern systematically samples opposite sides of the globe, and variability is dominated by large-scale, symmetric wave-like structures. k values obtained through the ensemble sampling exercise are confirmed by analysis of the correlation of model data for ACE-FTS sampling, but not for MIPAS.

4. Conclusions

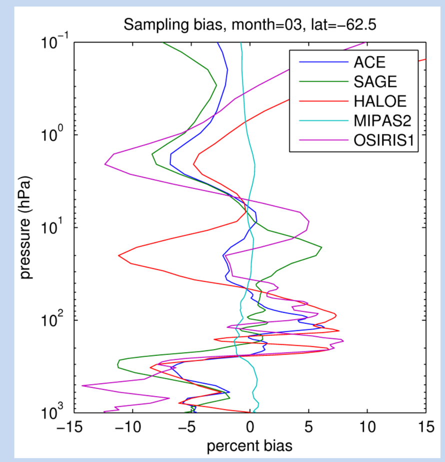


Figure 5: Estimated sampling bias for O_3 at 62.5°S for a selection of instruments.

- Sampling bias can easily affect the climatological means calculated from satellite instruments, especially those whose sampling pattern is irregular in time.
- Sample biases for O_3 are estimated to reach values of 10% in some regions of the stratosphere.
- Similarities in the sampling patterns of instruments can lead to similar biases.
- Extending our sampling exercise to examine the standard error of the mean, we find the SEM of climatological means is often smaller than the one calculated by the classic equation.
- The uniform nature of sampling, along with large-scale variability in the stratosphere, can lead to negative correlations between measurements.
- Therefore, use of the classic SEM equation should generally produce a conservative estimate of the SEM.

The SPARC Data Initiative team thanks WCRP, the Toronto SPARC office, and the International Space Science Institute (ISSI) in Bern, for hosting team meetings and providing generous travel support.