Characterizing sampling-related uncertainties in stratospheric trace gas climatologies

Matthew Toohey, Thomas von Clarmann,

Michaela Hegglin, Susann Tegtmeier,

and the SPARC Data Initiative team

7th Atmospheric Limb Conference

18.06.2013

GEOMAR

Introduction

- Monthly zonal mean “climatologies” of stratospheric trace gases can be produced by binning measurements from satellite instruments into monthly, 5° latitude bins (e.g., SPARC Data Initiative).
- Instruments have a wide range of sampling patterns
  - Global daily sample counts from ~10-1000 per day
  - Latitudinal coverage that may vary with time
  - Possibly incomplete coverage of bin space (month, latitude, longitude)

GEOMAR

Introduction

1. Sampling biases in stratospheric climatologies

   → *How does sampling impact monthly zonal means?*

2. Sampling and the standard error of the mean

   → *Is* $\hat{\sigma}_x/\sqrt{N}$ *a good estimate of the random error of monthly zonal means?*

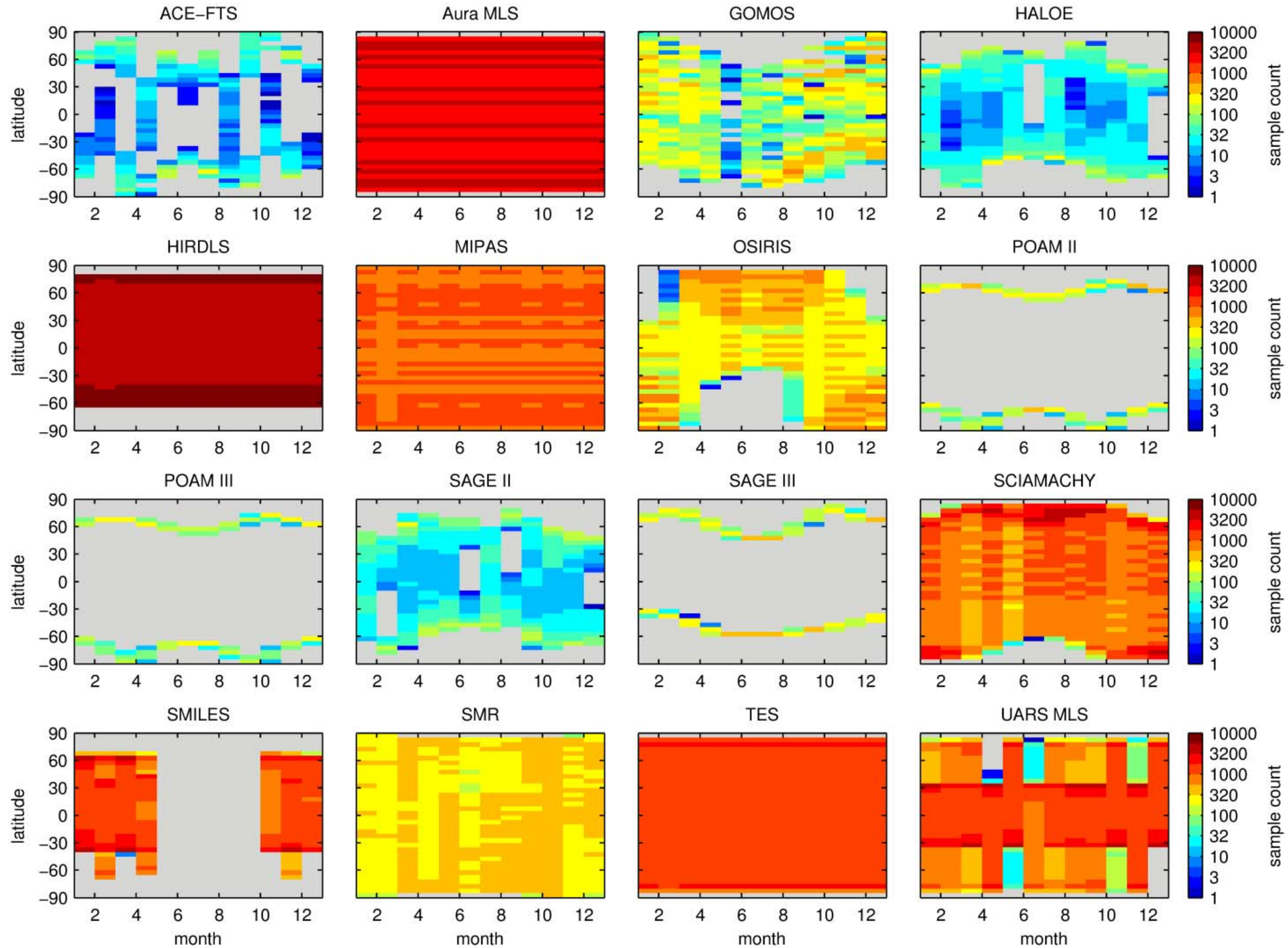
GEOMAR

Method

- Use chemical fields from a coupled chemistry model, (e.g., WACCM, CMAM)
- Sample model fields based on space-time sampling patterns of specific instruments (“*satellite simulator*”)
  1. Difference of sample mean and population mean (with full resolution model fields) gives estimate of potential sample bias in climatologies.
  2. Repeated resampling of model fields used to estimate of random error in climatologies.

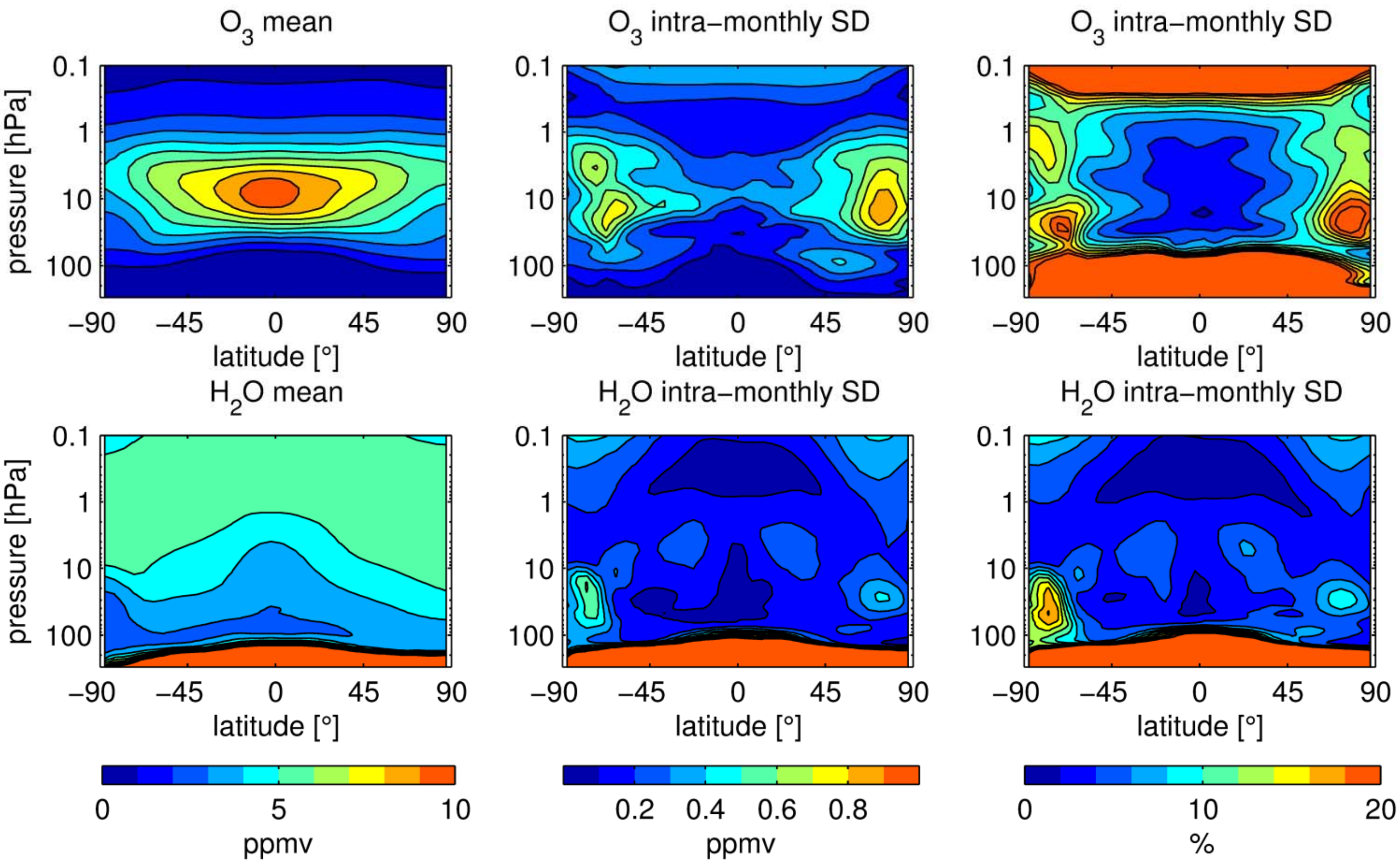
Instrument	Full Name	Sample reference period
ACE-FTS	Atmospheric Chemistry Experiment (ACE) - Fourier Transform Spectrometer	2005
Aura MLS	Aura - Microwave Limb Sounder	Jan 2005
GOMOS	Global Ozone Monitoring by Occultation of Stars	2003
HALOE	The Halogen Occultation Experiment	1994
HIRDLS	High Resolution Dynamics Limb Sounder	Sep 2006
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding	Jan 2009
OSIRIS	Optical Spectrograph and InfraRed Imager System	2009
POAM II	Polar Ozone and Aerosol Measurement II	1995
POAM III	Polar Ozone and Aerosol Measurement III	2001
SAGE II	Stratospheric Aerosol and Gas Experiment II	1990
SAGE III	Stratospheric Aerosol and Gas Experiment III	2003
SCIAMACHY	Scanning Imaging Absorption spectroMeter for Atmospheric CHartographY	2010
SMILES	Superconducting Submillimeter-Wave Limb Emission Sounder	Oct 2009 – Apr 2010
SMR	Sub-Millimetre Radiometer	2010
TES	Tropospheric Emission Spectrometer	Jul 2007
UARS MLS	UARS - Microwave Limb Sounder	1992

Sampling patterns: sample counts



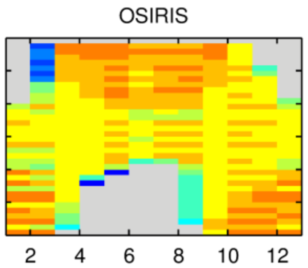
GEOMAR

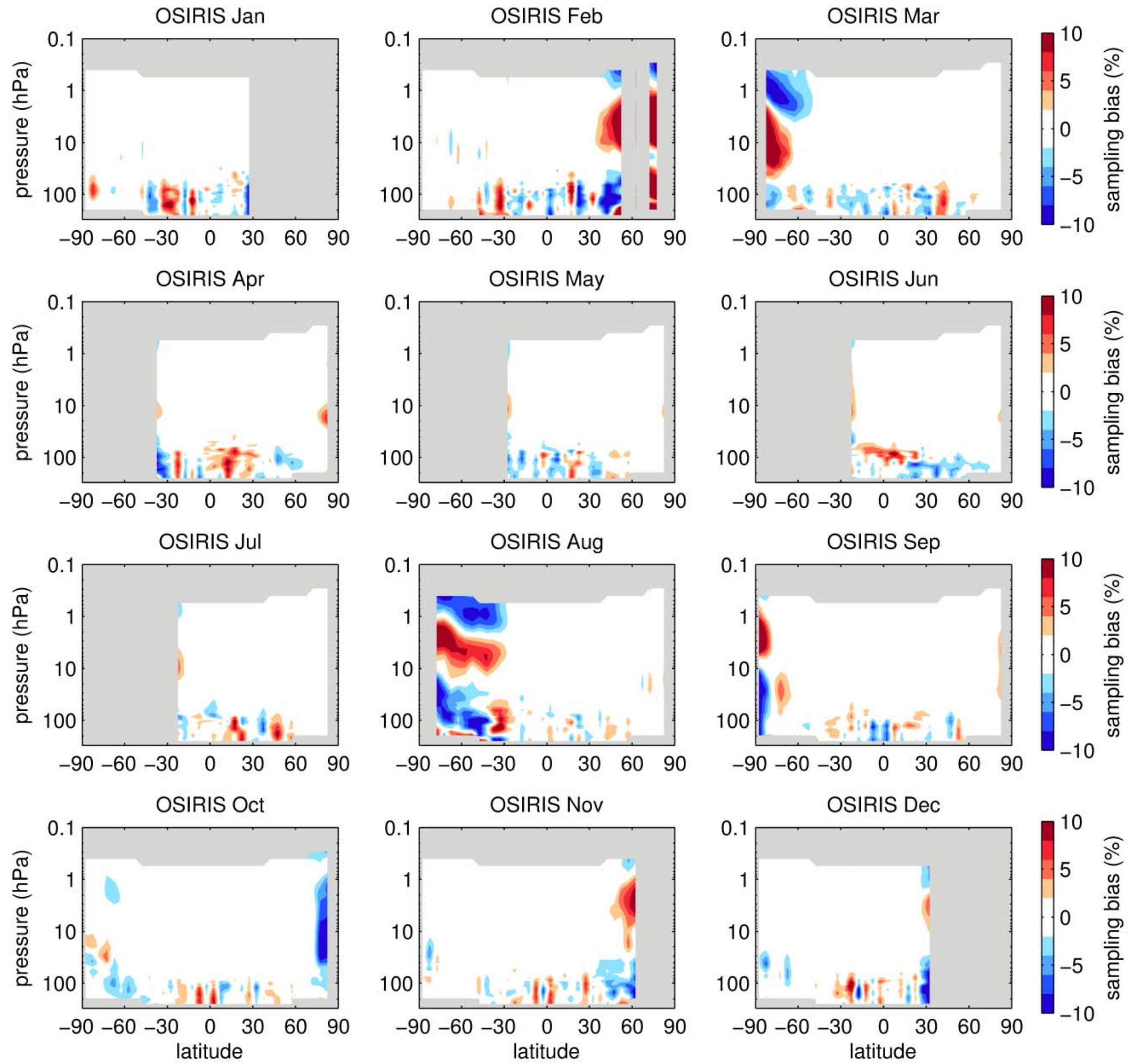
WACCM O_3 and H_2O



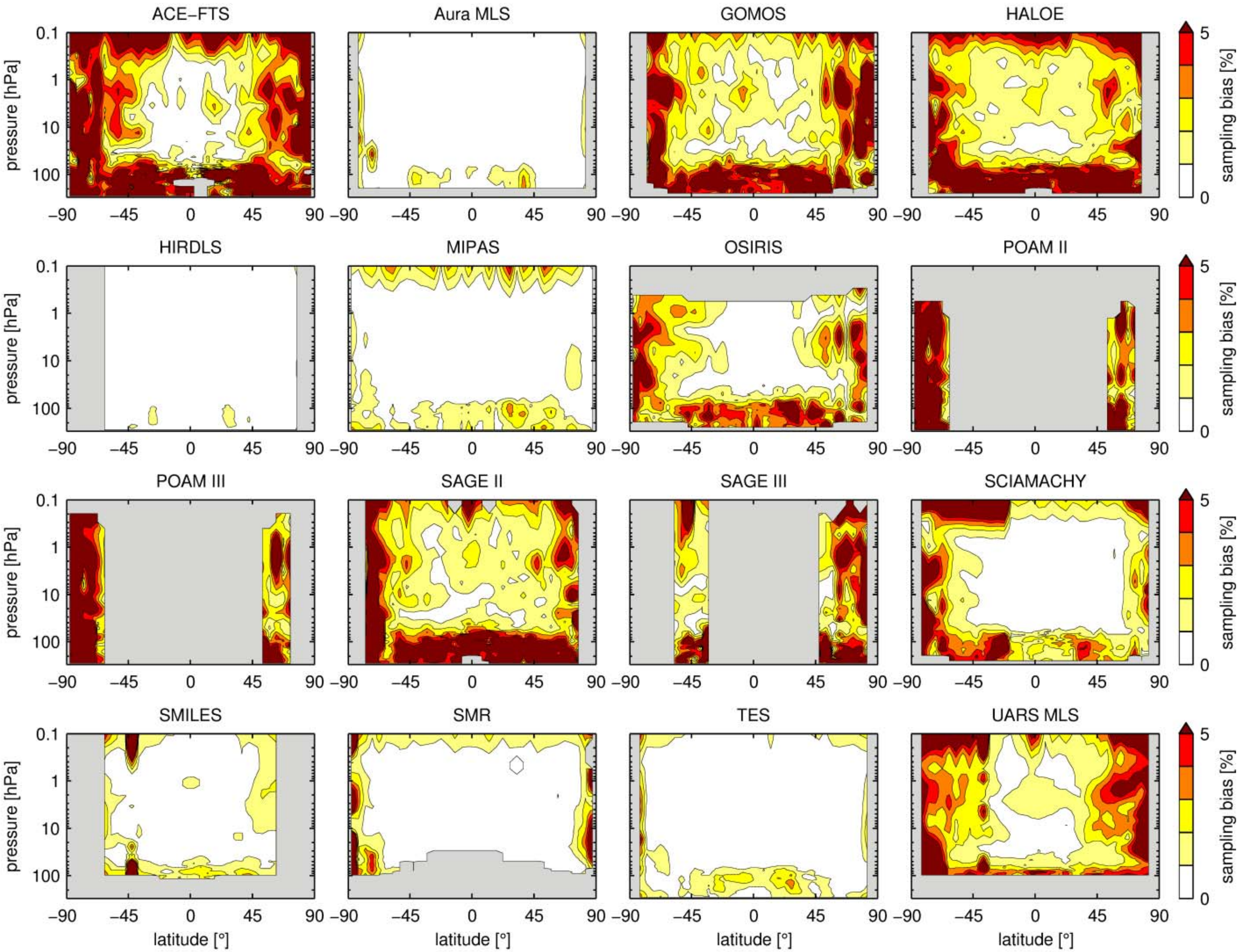
Single Instrument, 12 months of O_3 sampling bias

Sampling bias = sample mean – population mean



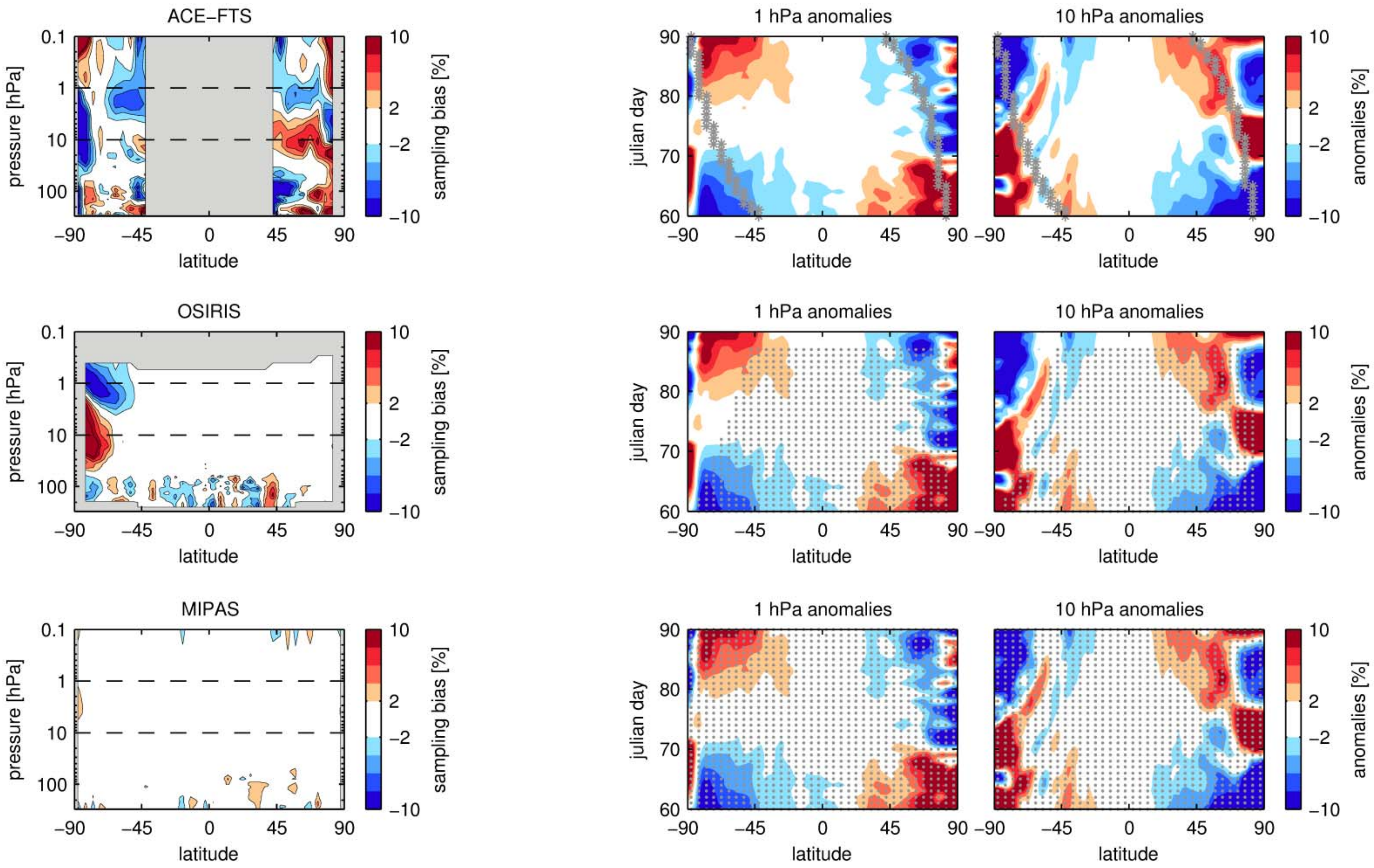


RMS monthly O_3 sampling bias



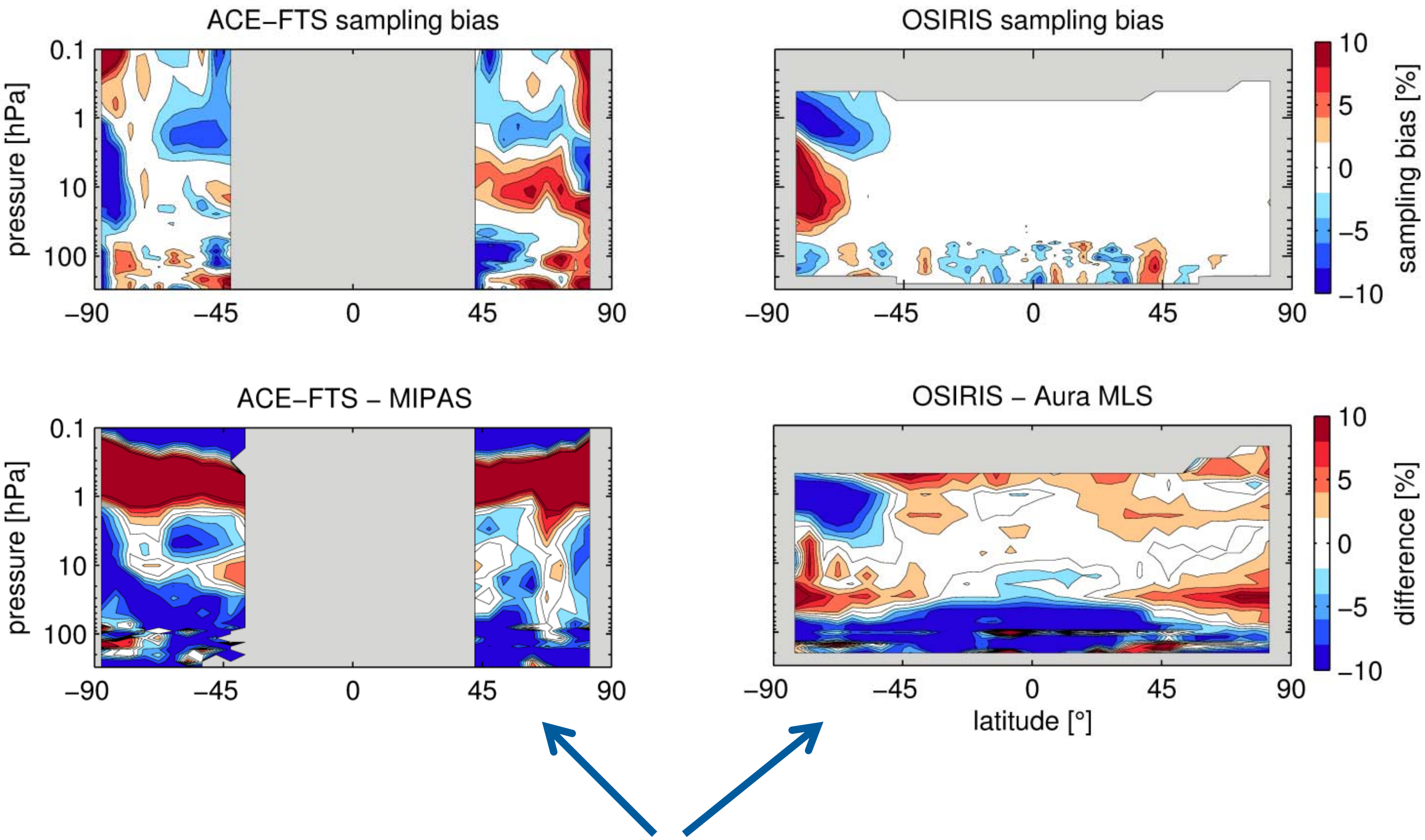
GEOMAR

March O_3 case study: impact of temporal non-uniformity



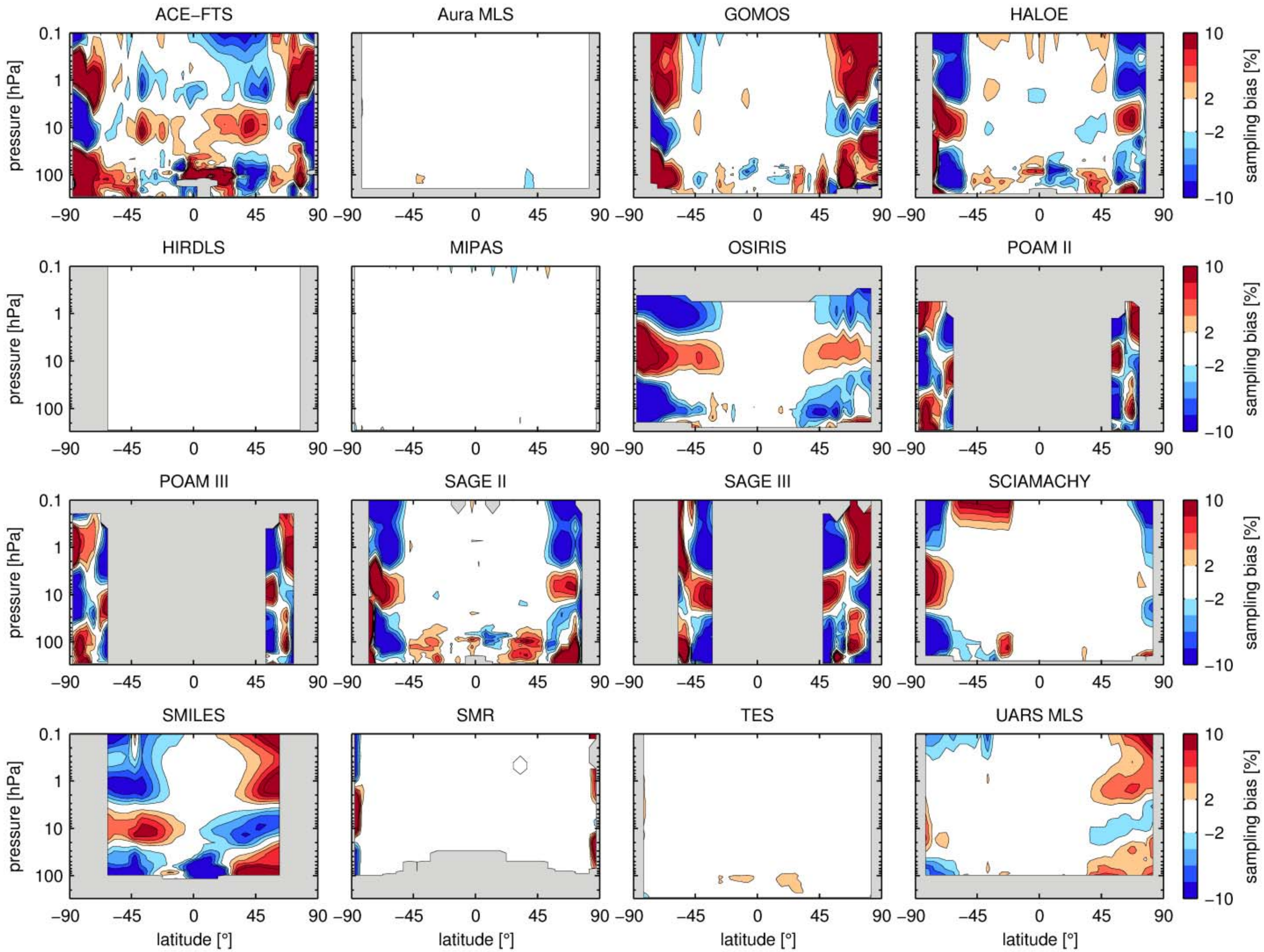
GEOMAR

March O_3 Case study: reality check



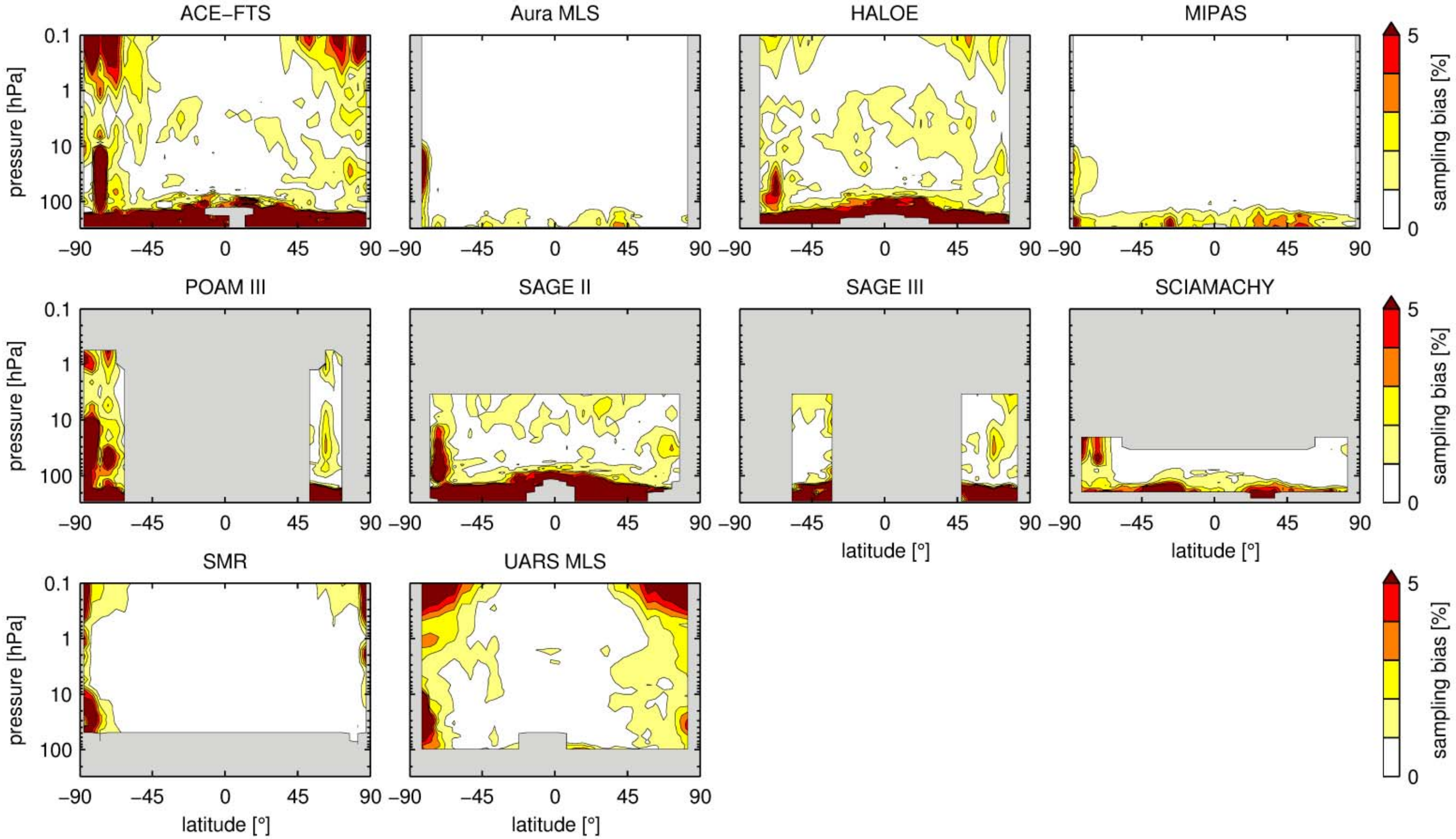
Real SPARC DI data!

Annual mean O_3 sampling bias



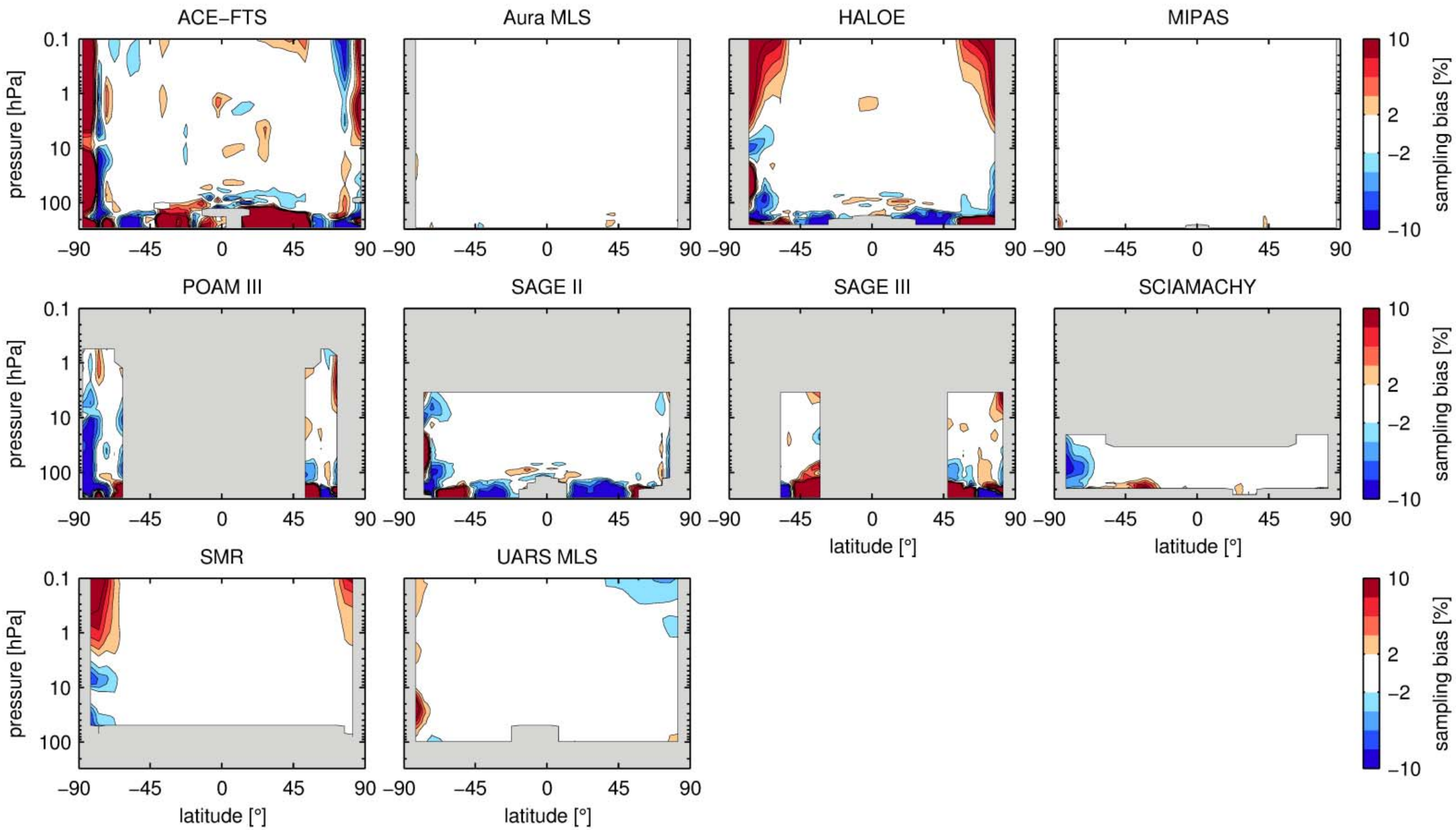
GEOMAR

RMS monthly H_2O sampling bias



GEOMAR

Annual mean H_2O sampling bias



GEOMAR

Part 2: random error of climatologies

- Is the standard SEM estimator,

$$\mathrm{SEM} = \frac{\hat{\sigma}_x}{\sqrt{N}}$$

appropriate for zonal mean climatologies from satellite measurements?

1. It assumes uncorrelated measurements!
2. It assumes random sampling!

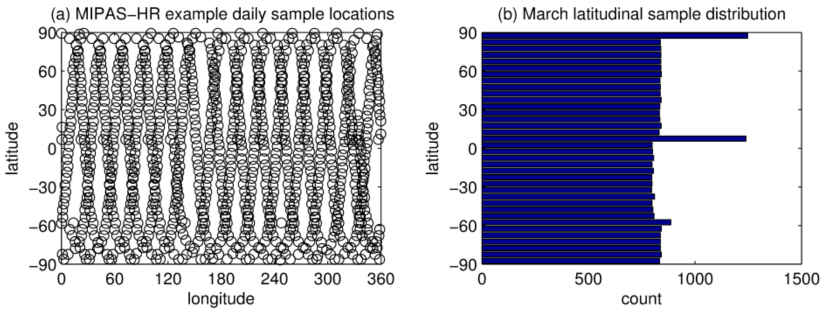
GEOMAR

Part 2: random error of climatologies

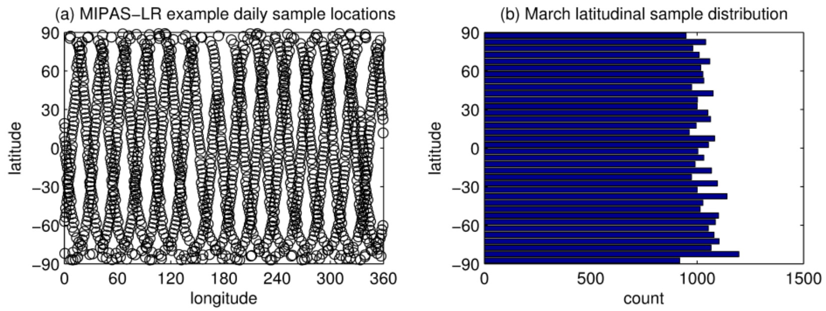
- What is the standard error of the mean?
  - Different samples (of size n) drawn from a population give different values of the sample mean. The **standard error of the mean** is the *standard deviation of those sample means* over all possible samples (of size n) drawn from the population.
- Idea: use model O_3 fields (CMAM this time), draw multiple* sample sets, calculate the mean for each set, and then calculate the SD of those sample means.
  - Compare with “classic” estimate σ/sqrt(n)

* duplicate sample sets created by shifting longitude and time such that LST of measurements remains constant

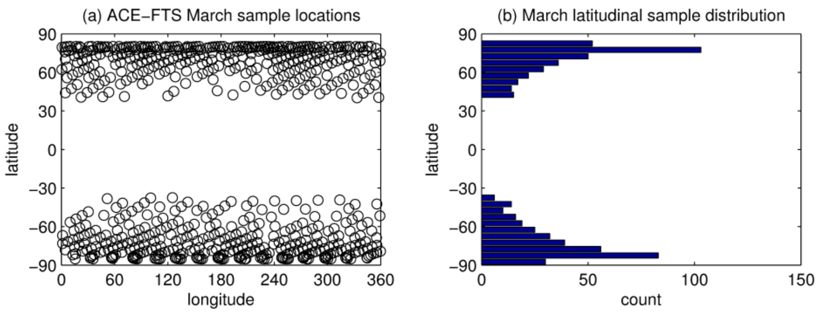
Sampling patterns: MIPAS and ACE-FTS



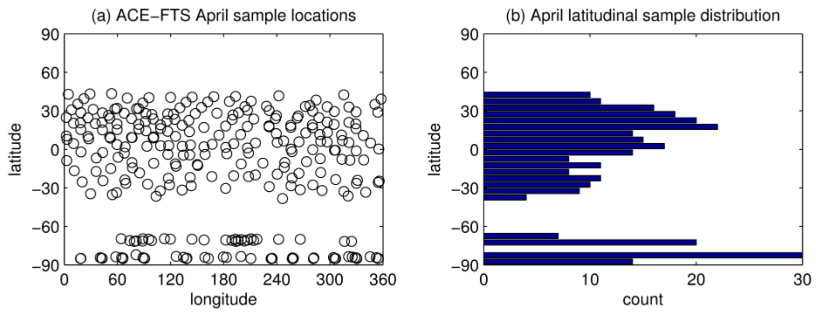
MIPAS-HR



MIPAS-LR



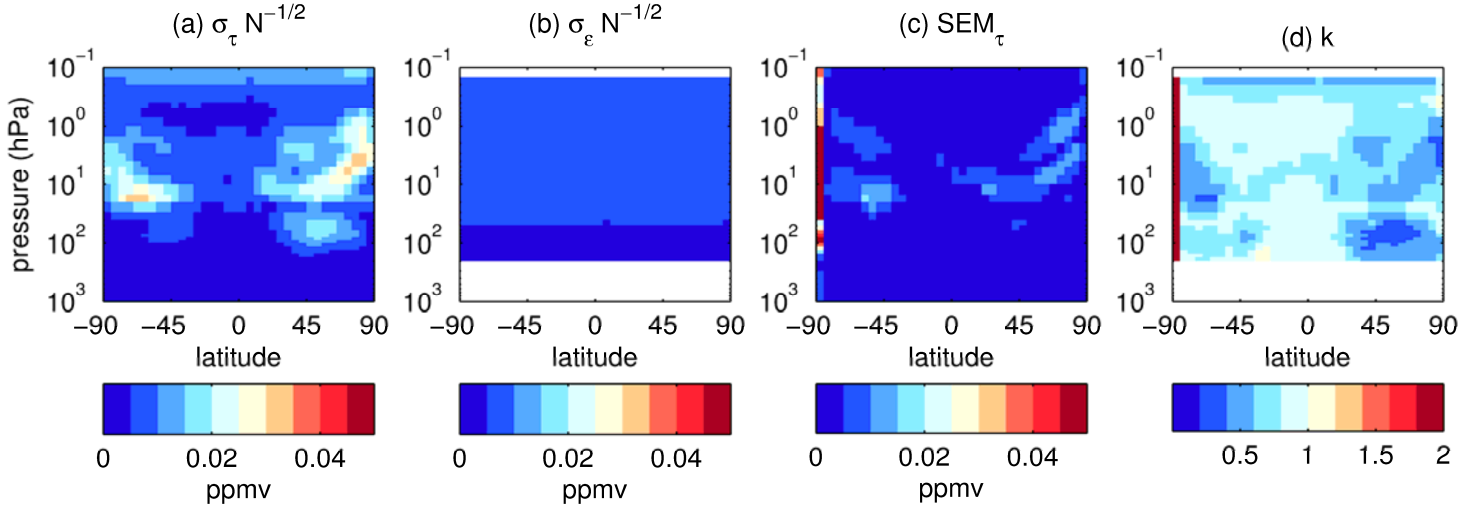
ACE-FTS March



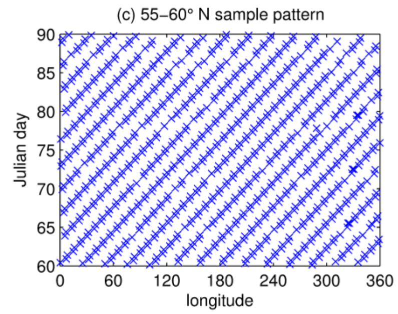
ACE-FTS April

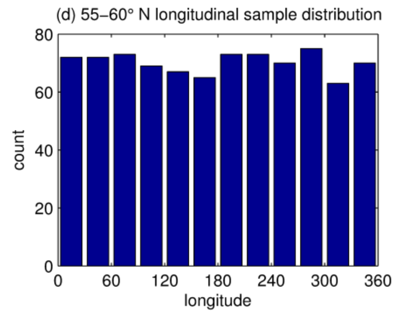
Results: MIPAS-HR

"classic" SEM estimate | Instrumental random error | "true" SEM | "true"/"classic" ratio



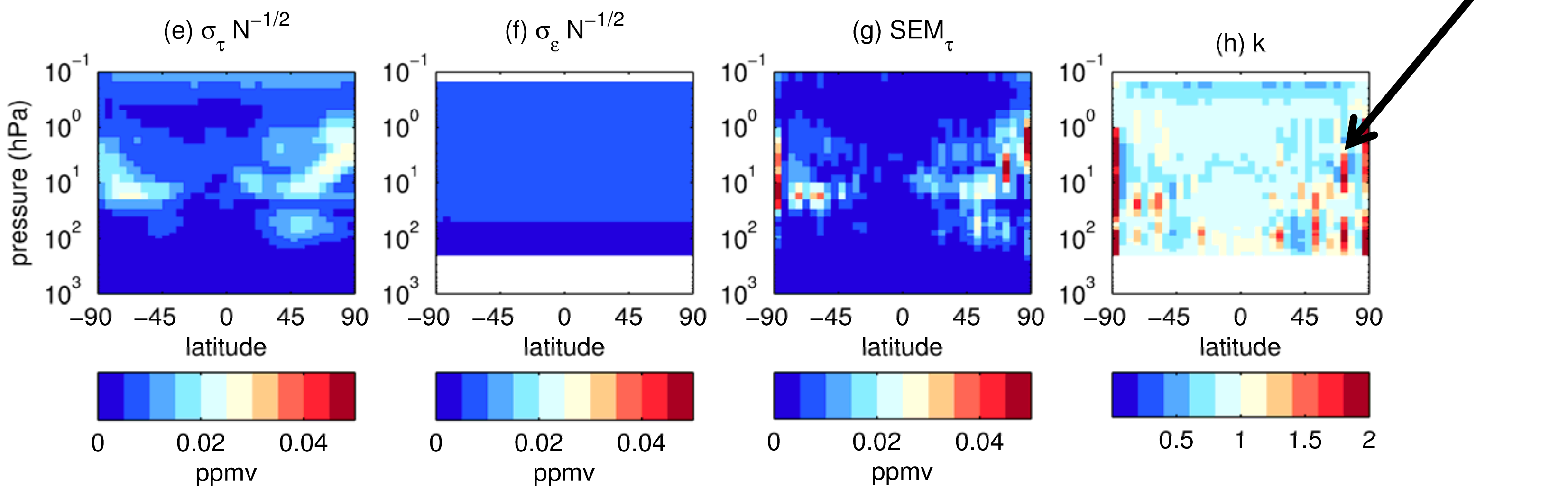
- k<1 → classic SEM estimate overestimates SEM ☺
- Explanation: ultra-uniform MIPAS-HR sampling produces better estimate of σ (and hence SEM) for given *n* than random sampling would.



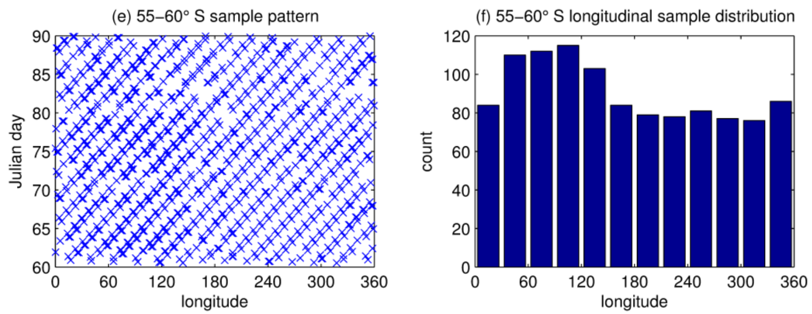


Results: MIPAS-LR

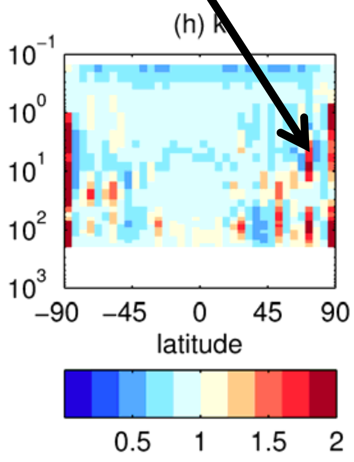
"classic" SEM estimate | Instrumental random error | "true" SEM | "true"/"classic" ratio



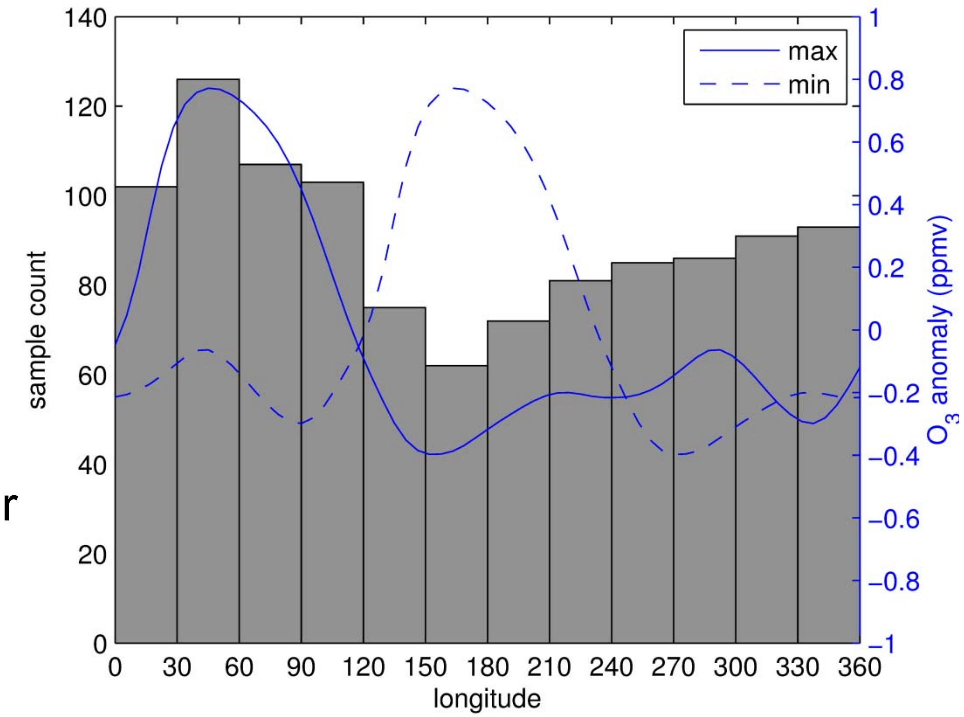
- $k>1$ → classic SEM estimate underestimates SEM ☠
- Explanation: systematic non-uniformity of MIPAS-LR sampling in certain latitude bins



When/where is σ/sqrt(n) a bad estimate?



Zonal sampling distribution for March in 50-55° N



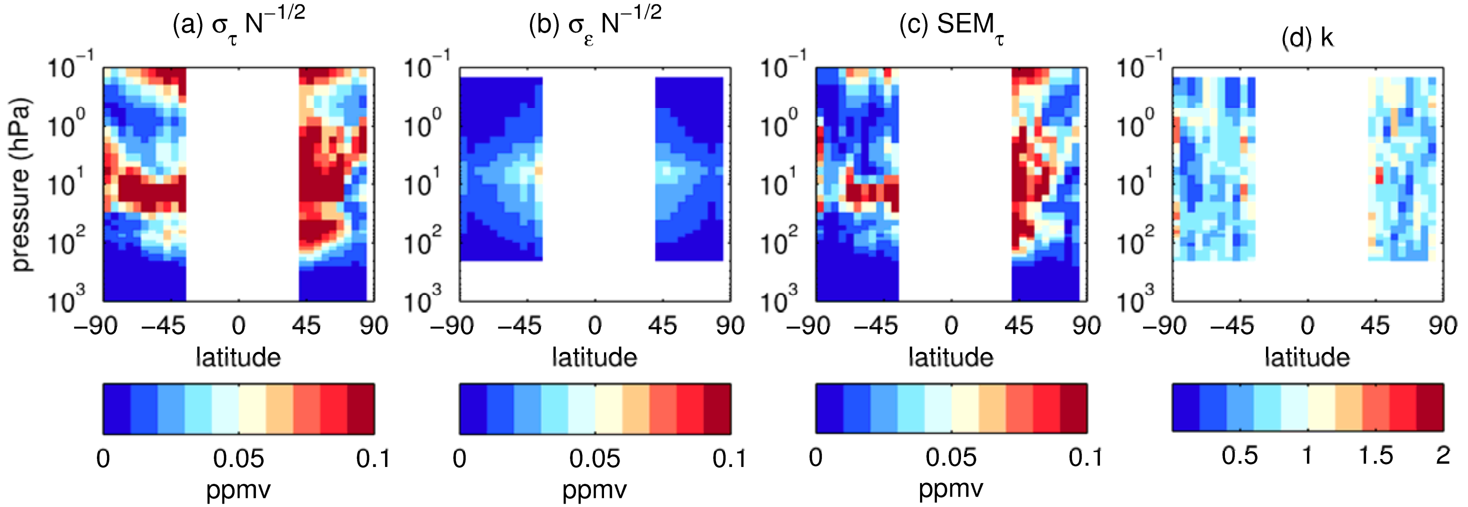
O_3 anomaly field at 52.4°N, 29 hPa

Non-uniformity of zonal sampling leads to random sample error when measured field shows zonal variability of same form as sampling

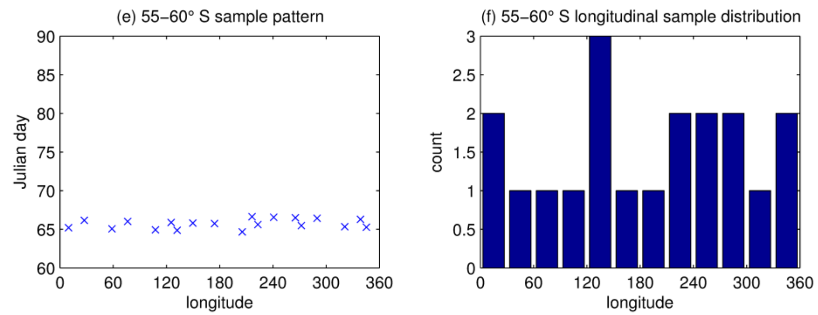
GEOMAR

Results: ACE-FTS March

"classic" SEM estimate | Instrumental random error | "true" SEM | "true"/"classic" ratio

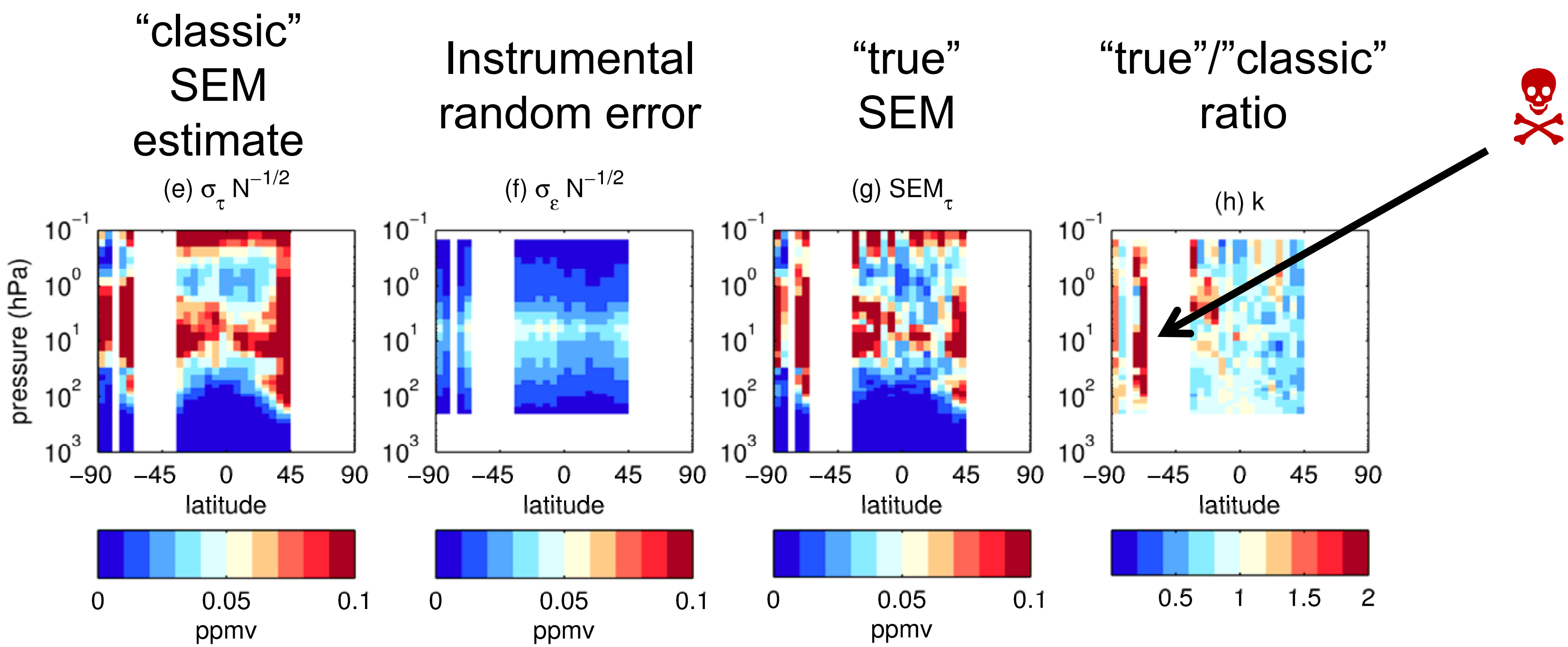


- *k*<1 → classic SEM estimate overestimates SEM ☺
- Explanation: even with relatively small *n*, ACE-FTS sampling is often more uniform than random sampling.

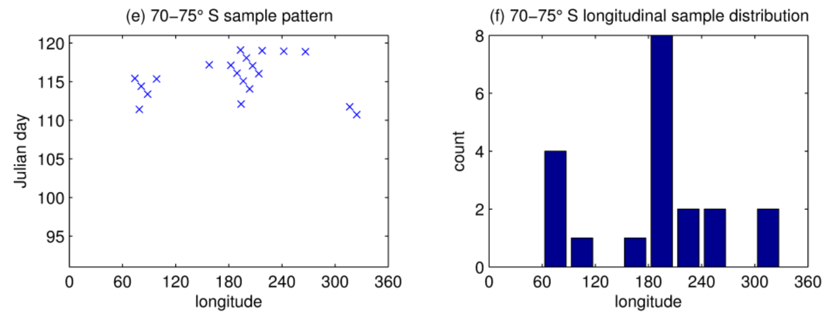


GEOMAR

Results: ACE-FTS April



- k>1 → classic SEM estimate underestimates SEM ☠
- Explanation: systematic non-uniformity of ACE-FTS sampling in certain latitude bins



For precise zonal means, zonal sampling uniformity is important!

GEOMAR

Caveats

- Results apply only to constructed monthly zonal means, produced by binning measurements!
- Concern horizontal and temporal sampling only, not vertical resolution.
- Results based on single years of model data
  - give a flavour of magnitude and important locations for sampling uncertainties
- Results based on single years of sampling
  - Variations in year-to-year sampling may also be important

Conclusions

- Sampling bias is an important consideration in the interpretation of zonal monthly mean climatologies
  - Monthly O_3 biases reach >10% in high latitudes for some instruments.
  - Annual mean O_3 sampling bias also often ~10%.
  - Non-uniform temporal sampling is most important mechanism, but non-uniform latitude and longitude sampling (not shown here) also play a role.

→*Toohey et al., submitted to JGR*

- The oft-used SEM estimator σ/sqrt(n) is *generally* a conservative estimate of the random error in zonal monthly means
  - True even for sparse samplers like ACE-FTS
  - Exceptions occur when zonal sampling distribution is non-uniform, and coincidident with similar zonal non-uniformity of the measured field (not limited to sparse samplers!)

→*Toohey and von Clarmann, AMT (2013).*

GEOMAR