F12	single profile from Matricardi w/ ±10% random scaling	GMI Daily profiles in 2006		Only available from GMI
CCl4	single profile from Matricardi w/ ±10% random scaling	GMI Monthly profiles in 2006		Only available from GMI
HNO3	single profile from Matricardi scaled to get 0.4 DU, then randomly varied by ln(q')=ln(q)±ln(5) (varied by factor of 5)	GMI Daily profiles in 2006	GEOS-Chem MOZART	GMI has wider altitudinal coverage
SO2	single US Standard Atmosphere profile scaled to get 0.1 DU, then randomly scaled (on a log scale) to get random range of 0.09 to 900 DU. The scale factor is a two-piece hyperbolas of log(p), with the maximum factor D (and zero vertical derivative) at 235 mb tapering to D/1000 at the top and D/100 at the bottom. The rate of tapering was arbitrary.	No Update	GEOS-Chem MOZART	No enough altitudinal coverage and variability from GEOS-Chem and MOZART
OCS	constant with height at 500 pptv from surface up to 20 km an then linearly decrease to 0 at 50 km; the suggested dynamic range (randomized) is ±10% (per S. Tjemkes)	AER 2D model (19 latitudes x 12 months)		Only available from GMI 2D
CF4	dynamic range (randomized) of 50 to 70 pptv, constant profile (per S. Tjemkes),	No Update	[1]Harnisch et al. 1996 (Single value, no profile information. CF4 trend: 62 ppt to 75 ppt from 1982 to 1995)	
NH3	Derived from profiles over Australian fires and sugar cane fields provided by Guergana Guerova, University of Wollongong	GEOS-Chem Monthly profiles in 2006	MOZART	
HCOOH	ATMOS profile	GEOS-Chem Monthly profiles		Dylan Millet via Karen Cady-Pereira
CH3OH	GEOS-CHEM profile provided by Dylan Millet. Harvard University	GEOS-Chem Monthly profiles	MOZART	Dylan Millet via Karen Cady-Pereira
C2H2	Remedios [MIPAS team]: mean profile and 1-STD variability. For the training we should use N-STD	GEOS-Chem Monthly profiles		Yaping Xiao
C2H4	ATMOS profile	GEOS-Chem Monthly profiles	MOZART	May Fu
HCN	Remedios [MIPAS team]: mean profile and 1-STD variability. For the training we should use N-STD	GEOS-Chem Monthly profiles		Qinbin Li via Yaping Xiao
CHClF2 (F22)	Remedios [MIPAS team]: mean profile and 1-STD variability. For the training we should use N-STD	No Update		

[1]Harnisch J., et al., Tropospheric trends for CF4 and C2F6 since 1982 derived from SF6 dated stratospheric air, GRL, 23(10), 1099-1102, 1996
[2] Hoffmann, et al., Envisat MIPAS measurements of CFC-11: retrieval, validation, and climatology, Atmos. Chem. Phys., 8, 3671-3688, 2008

Some technical details for the selecting process:

- Since the GMI daily profiles are in the order of ~1E6 (144x91x365=4782960), we need to choose a smaller dataset to shorten the selecting process. We construct a new dataset by picking one month in each season (Jan, Apr, July, Oct), and picking 3 days in a month (one day every 10-day interval), i.e., the total profiles used for the selecting algorithm is 144x91x4x3 = ~160,000.

- Based on the ~160,000 profiles sampled from above, we use the same sampling approach as we did for Mars profiles, and we selected 182-200 profiles after trying different cutoff criteria for the distance between profiles.

- The write_oss_var.pro program read the above 182-200 profiles,
  remove the profiles with spikes and avoid NaN resulting from interpolation
  convert units (from volume mixing ratio to mass ratio g/g),
  do vertical interpolation/extrapolation,
  and write the first 182 profiles into the final NEWOSS TXT file.

Total daily profiles → total used for selecting process → total selected →total for OSS
4782960 161179 182~200 182

Mapping the profiles from original model grids to 101 levels:

- The original global CTMs fields come in different vertical grids. The profiles as input to OSS are on 101 levels as shown in the OLD_OSS.txt file. So vertical interpolation and extrapolation are needed.
- All interpolation/extrapolation are based on logarithmic scale.
- For those levels below the lowest model pressure level (~1000 hPa), use constant values the same as the lowest model grid.
- For those levels above the top model level, use Gaussian extrapolation for NH3 (above 80mb), HCN (above 80 mb), CH3OH (above 80 mb), C2H4 (above 100 mb); and linear extrapolation for C2H2 (above 1 mb) and HCOOH (above 0.1 mb).

Available species from global chemical transport models (CTMs) and aircraft campaign (HIPPO):

1. **GEOS-Chem:**
   O3, CO2, CH4, CO, HNO3, SO2, OCS, NH3, CH2O2(HCOOH), CH3OH, C2H2, C2H4, HCN (species not from GEOS-Chem: N2O, CCl4)
   Note: these could be from different years

```
/data_static/carbon/TESCH4/GC_2006_2x25/ctm.2006MM01.bpch
CH4: /project/carbon_tmp/mars/yxiao/CH4.GEOS3_2x25_2001.bpch
C2H2: /data_static/carbon/TESCH4/yaping_tmp/ctm.bpch.newAS
C2H4: /project/carbon_tmp/mars/yxiao/C2H4.bpch
HCN: /project/carbon_tmp/mars/yxiao/HCN.ctm.bpch.ori_1tr
*CO2: /project/carbon_tmp/mars/yxiao/CO2.2006-2007-v80302.bpch
HCOOH: /project/carbon_tmp/mars/yxiao/oss_hcooh/
NH3: /data_static/carbon/TESCH4/GC_2006_2x25/ctm.2006MM01.bpch
```

* CO2 monthly mean profiles from Ray Nassar (ray.nassar@ec.gc.ca)