united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

Course on "Inverse Methods in Atmospheric Science"
1 - 12 October 2001

301/1332-10

"Retrieval Algorithms for MIPAS Data Processing"

M. CARLOTTI
Università di Bologna
Italy

Please note: These are preliminary notes intended for internal distribution only.

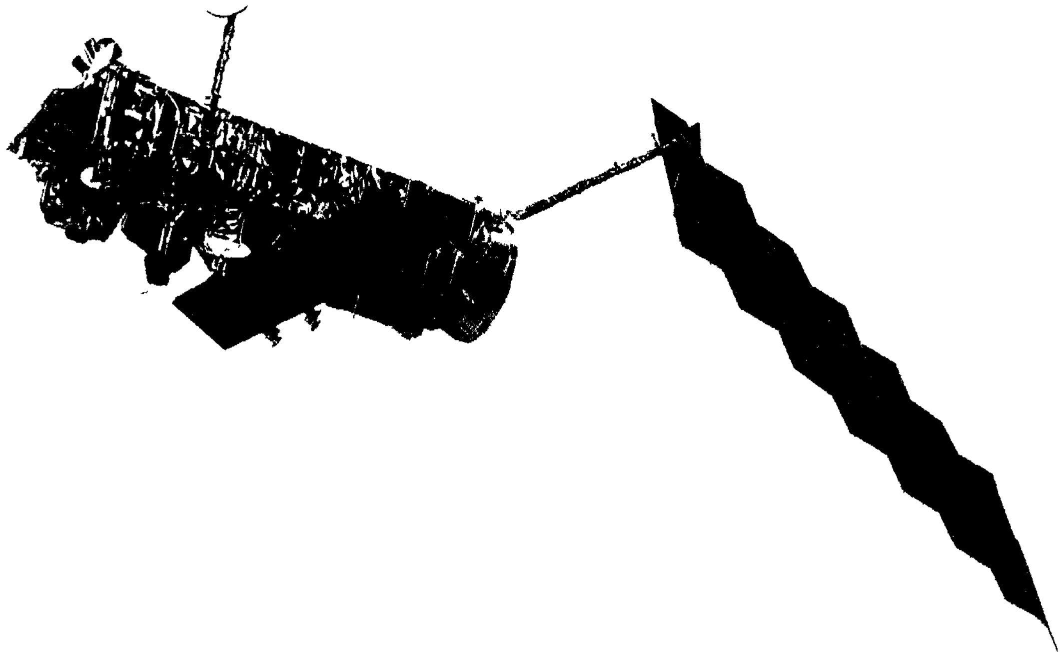
strada costiera, 11 - 34014 trieste italy - tel +39 0402240111 fax +39 040224163 - sci_info@ictp trieste it - www ictp trieste it

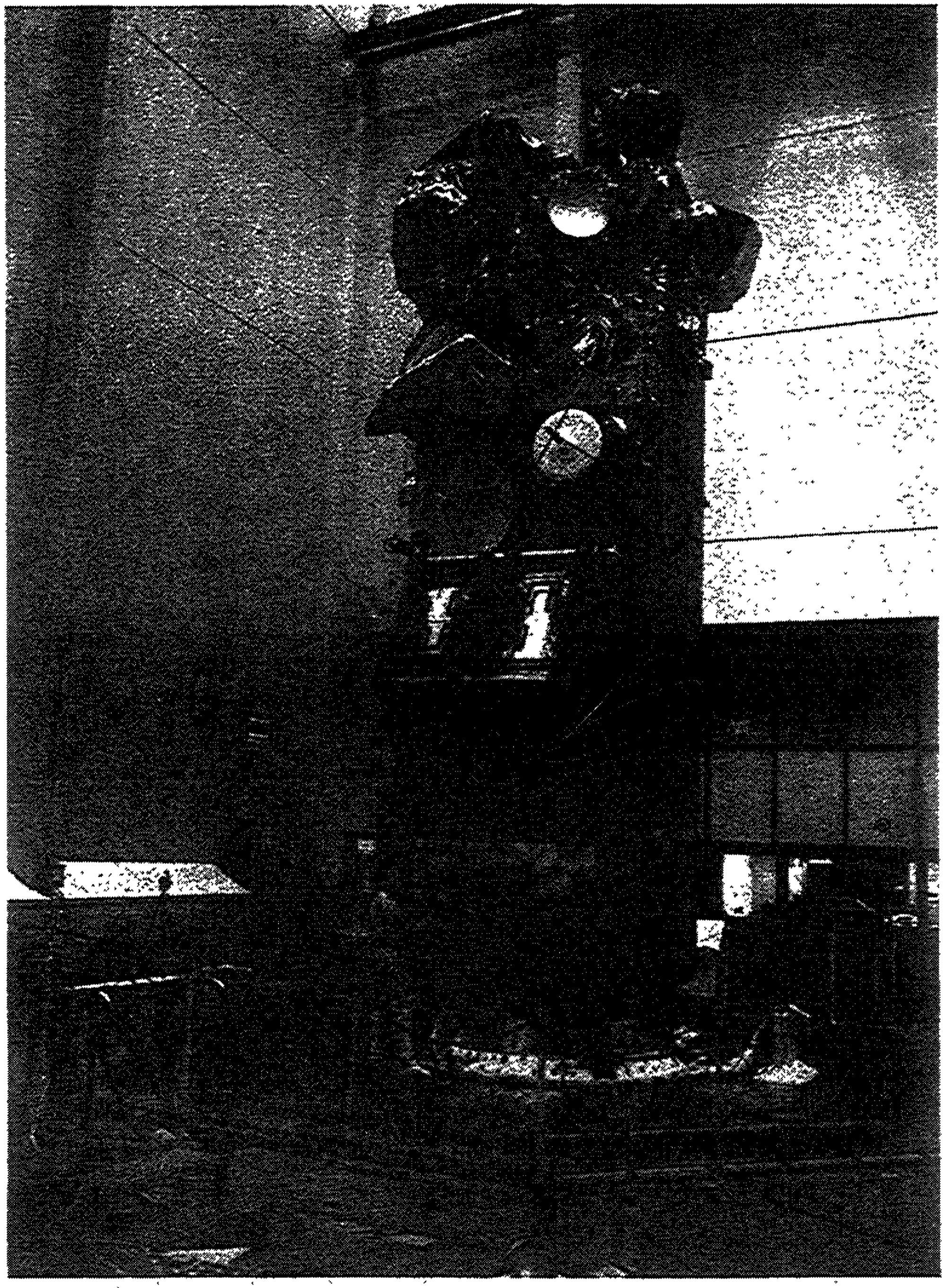
RETRIEVAL ALGORITHMS FOR MIPAS DATA PROCESSING

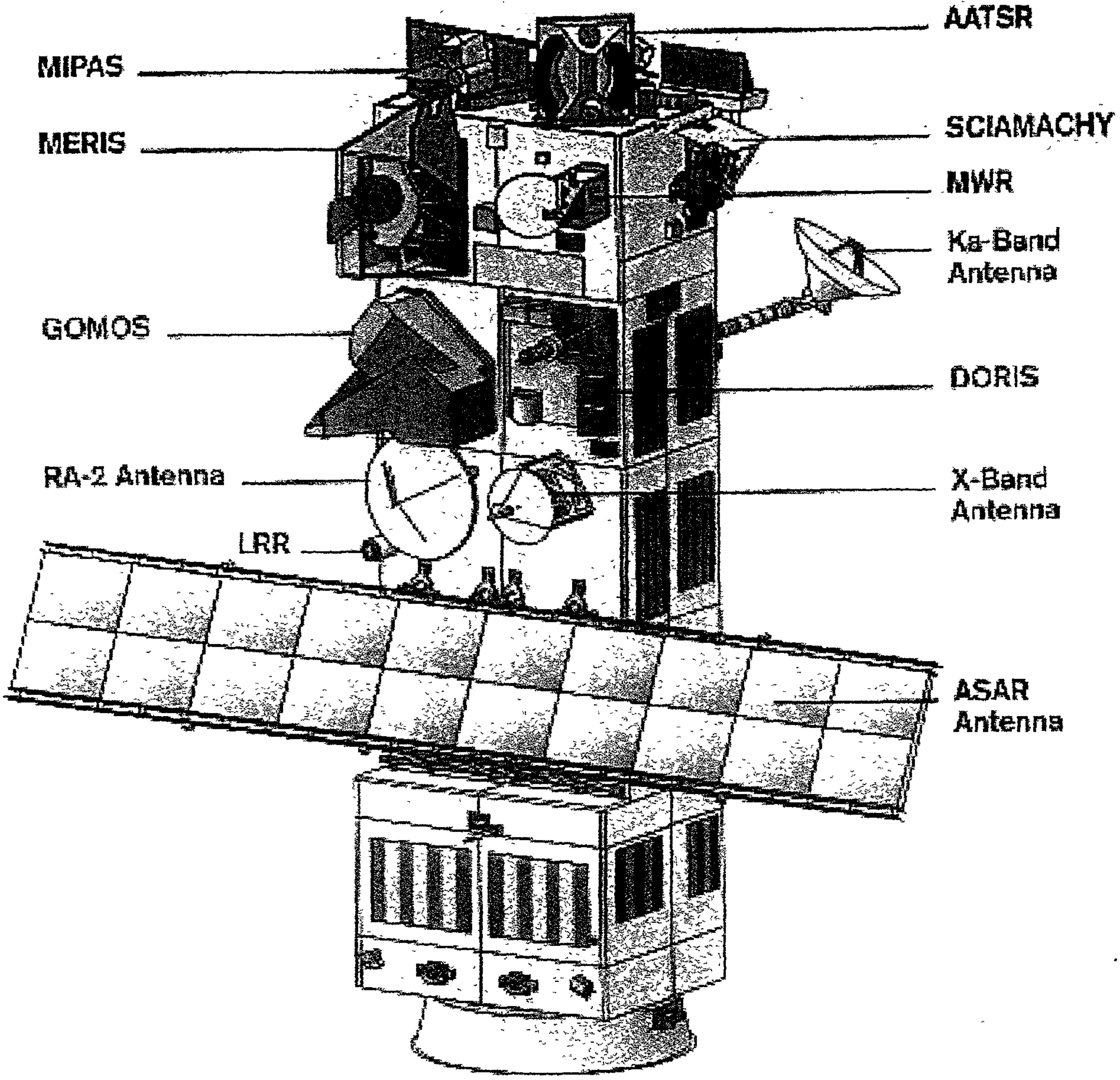
Massimo Carlotti
Universita' di Bologna
Italia

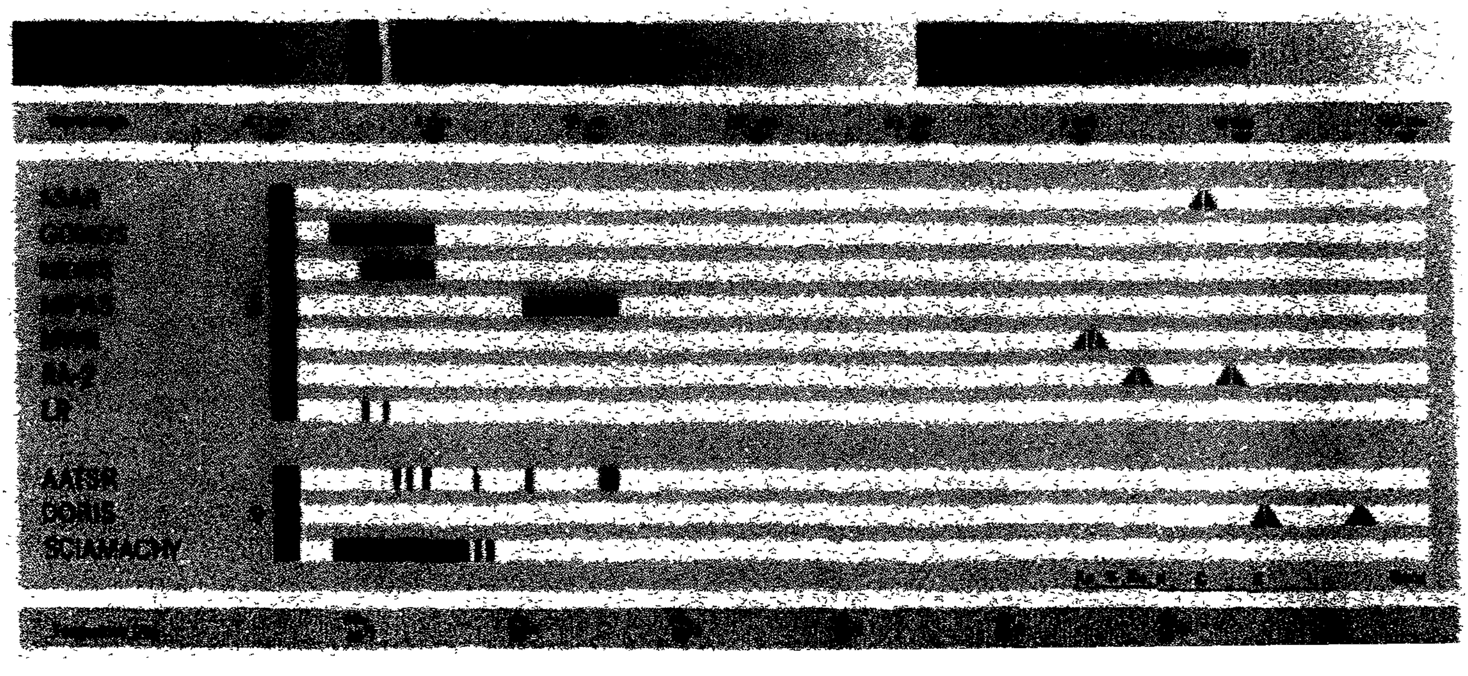
"***Inverse Methods in Atmospheric Science***" Trieste 1-12 October 2001 esa

ENVISAT
CARING FOR THE EARTH



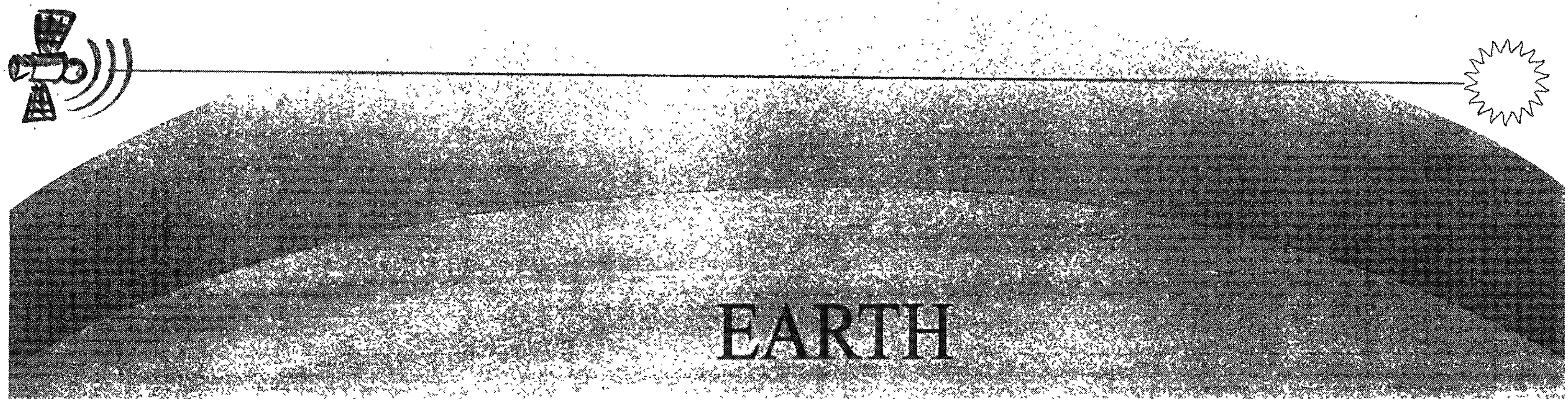



AATSR
MIPAS
MERIS
SCIAMACHY
MWR
Ka-Band
Antenna
GOMOS
DORIS
RA-2 Antenna
X-Band
Antenna
LRR
ASAR
Antenna


AATSR
DORIS
SCIAMACHY

THE **MIPAS** INSTRUMENT
(Michelson Interferometer for Passive Atmospheric Sounding)

- ESA core instrument will be operated on-board ENVISAT
- performs limb-scanning measurements of the atmospheric emission spectrum in the middle infrared (685 - 2410 cm^{-1})
- main target species of MIPAS: H_2O, O_3, HNO_3, CH_4, N_2O, NO_2
- spectral resolution: 0.025 cm^{-1} (FWHM, unapodized)
- limb-scanning mode (standard mode): 6 - 68 km in 3 km steps (5 and 8 km steps in the upper stratosphere)
- instrument FOV: 3 x 30 km
- time per elevation scan: 75 s


EARTH

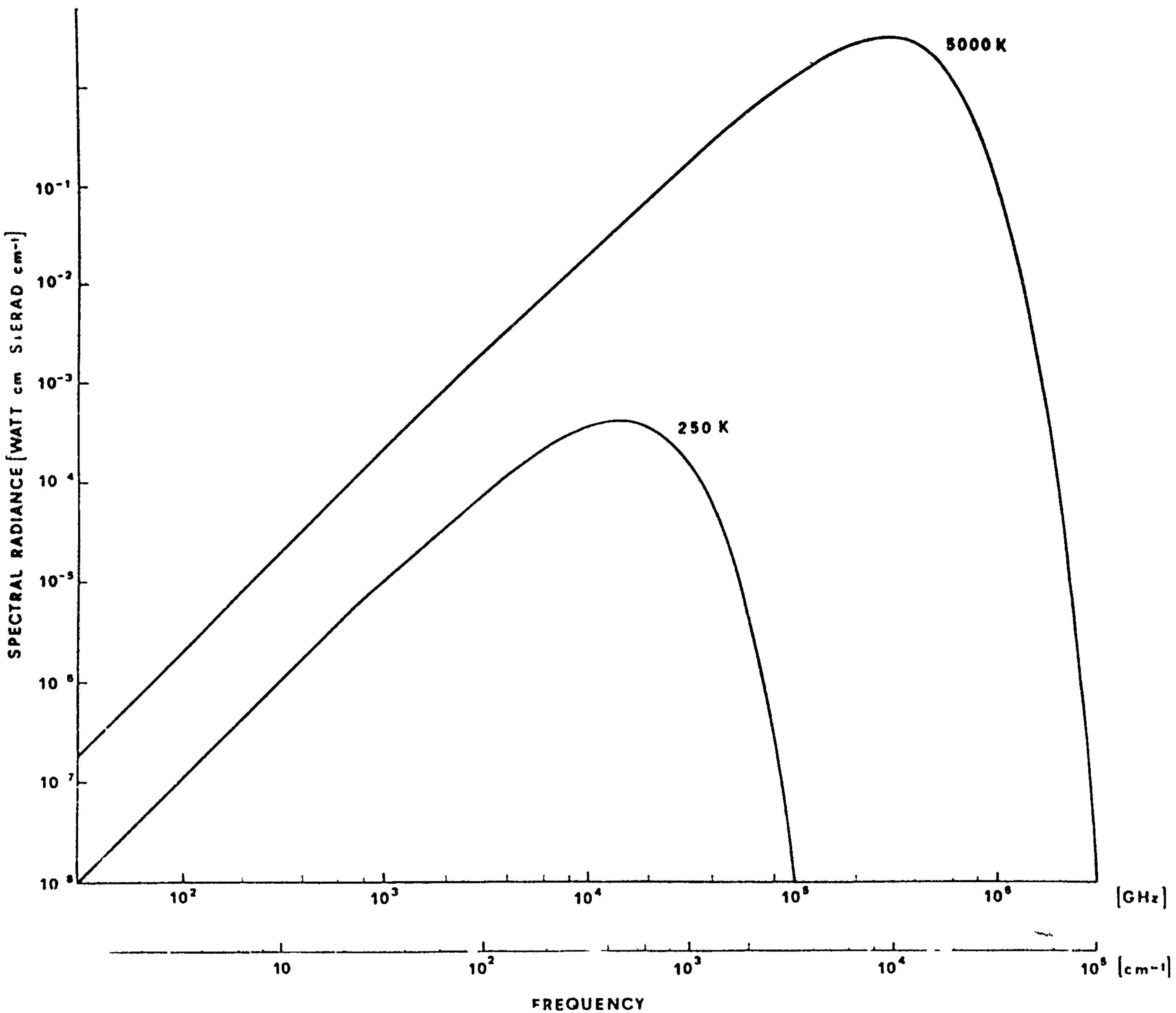
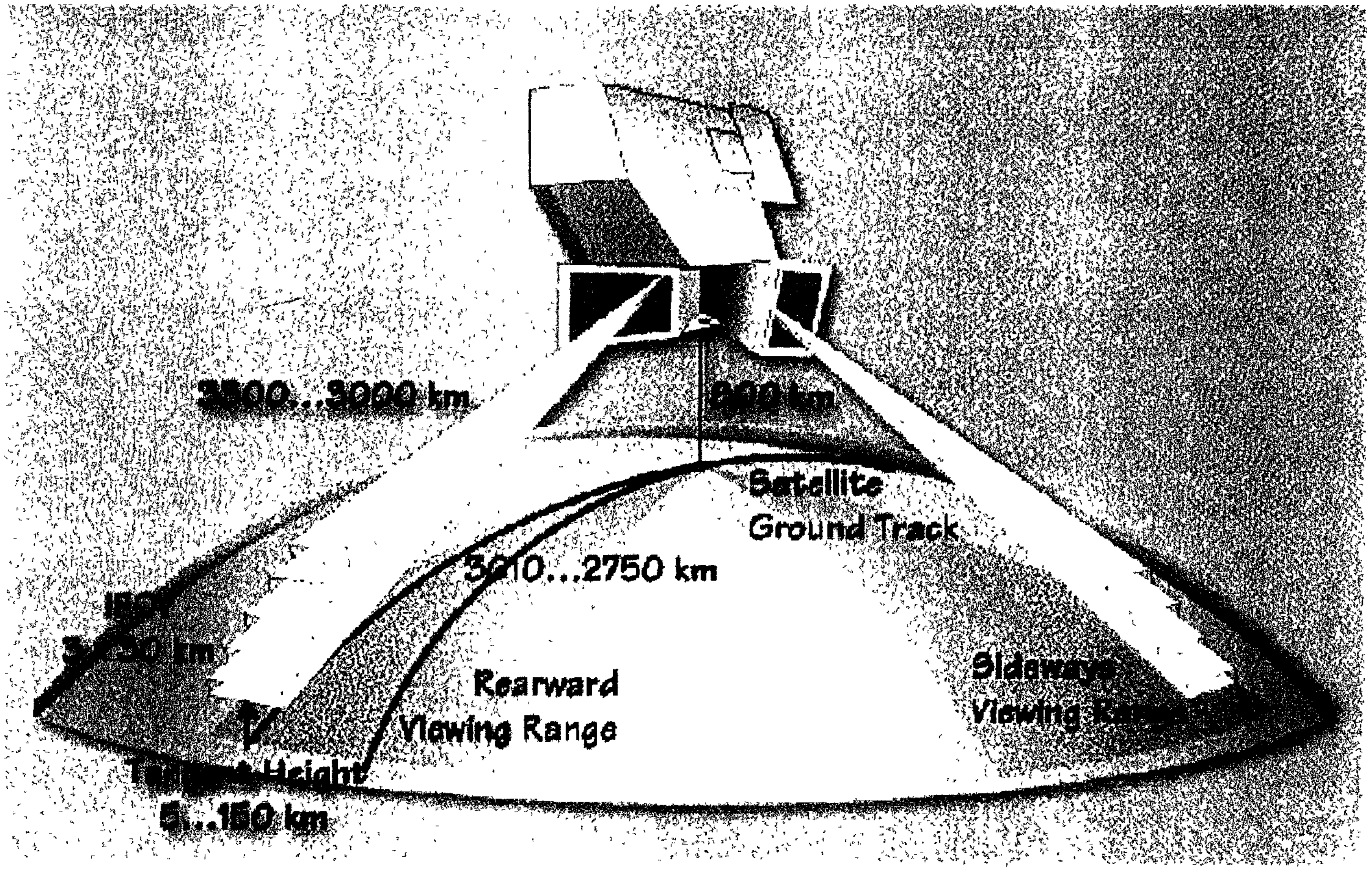
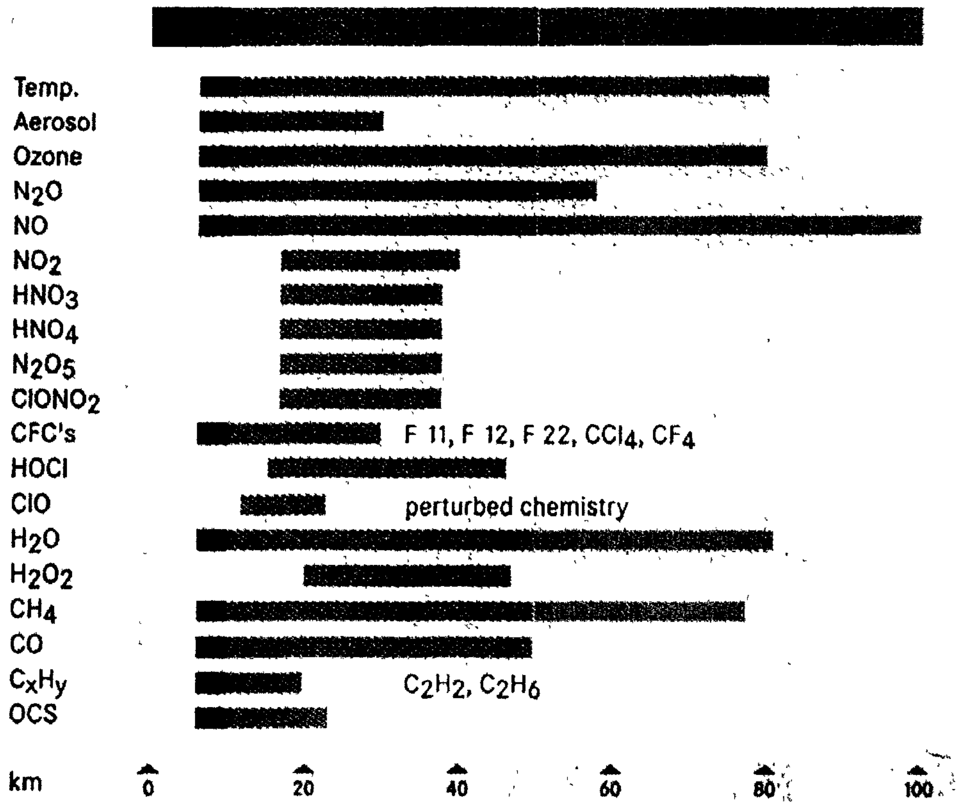

5000 K
250 K
SPECTRAL RADIANCE [WATT cm S.ERAD cm⁻¹]
10⁻¹
10⁻²
10⁻³
10 ⁴
10⁻⁵
10 ⁶
10 ⁷
10 ⁸
10²
10³
10⁴
10⁵
10⁶
[GHz]
10
10²
10³
10⁴
10⁵
[cm⁻¹]
FREQUENCY

FIG. 3

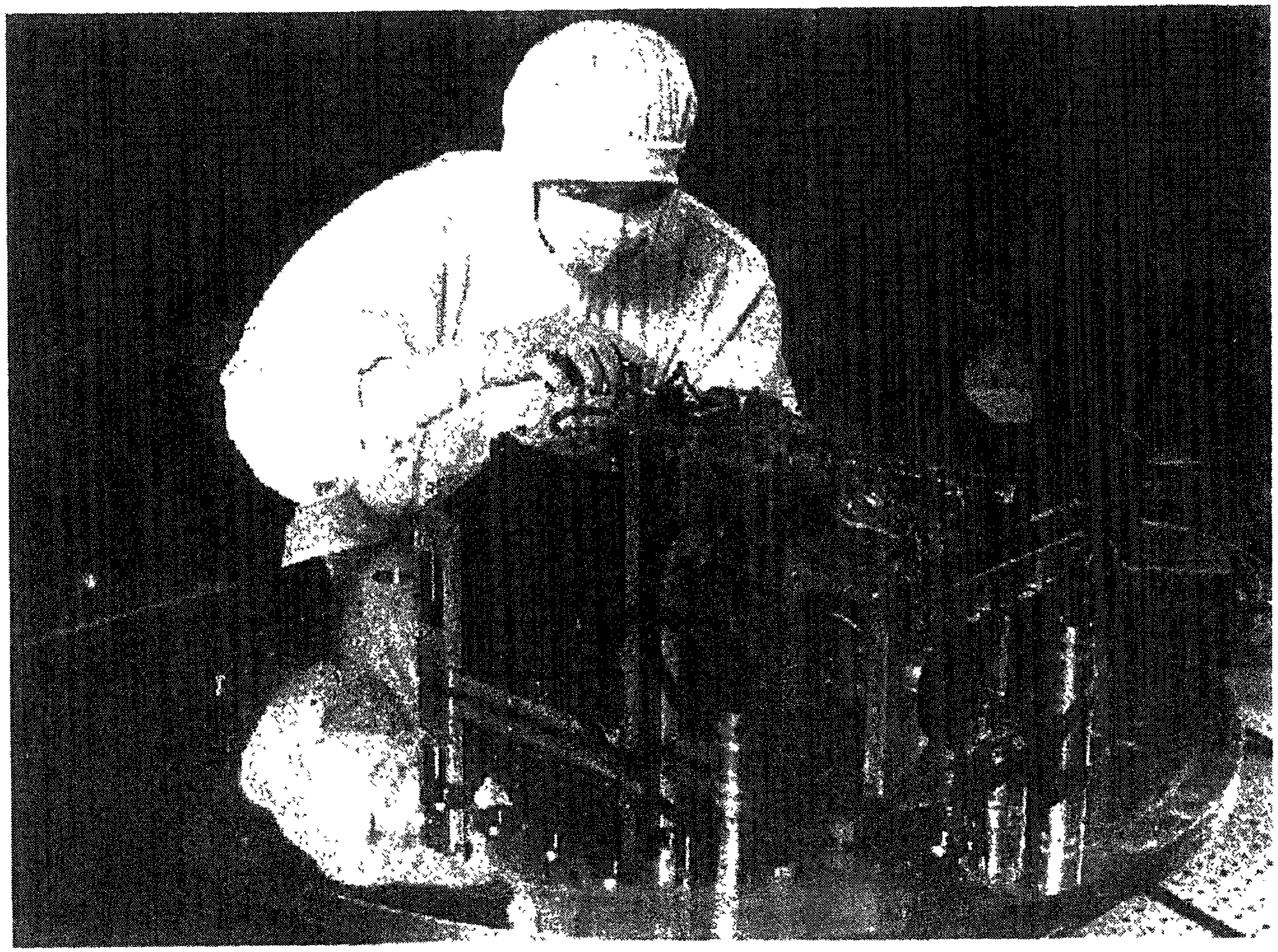
MIPAS observation geometry

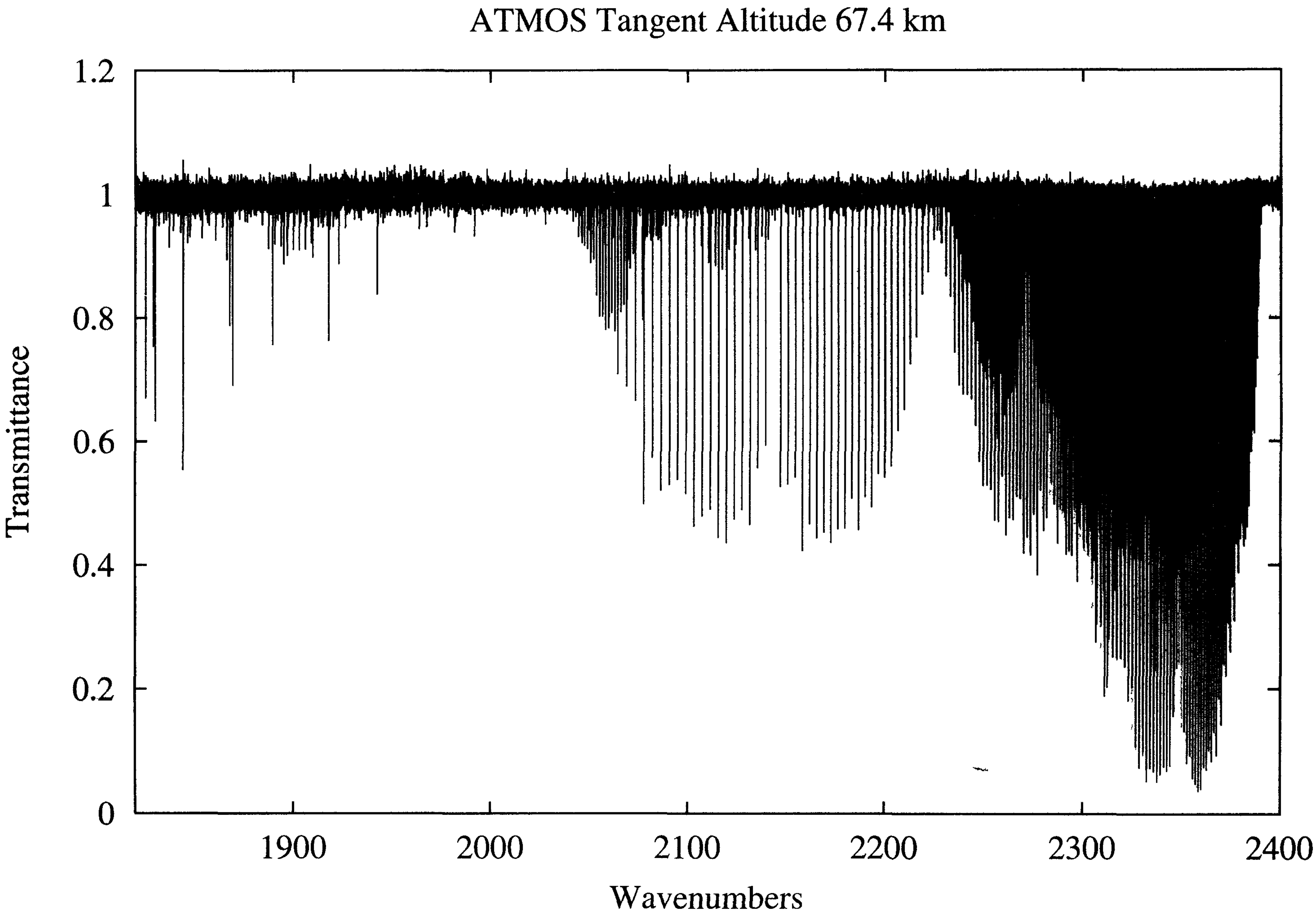


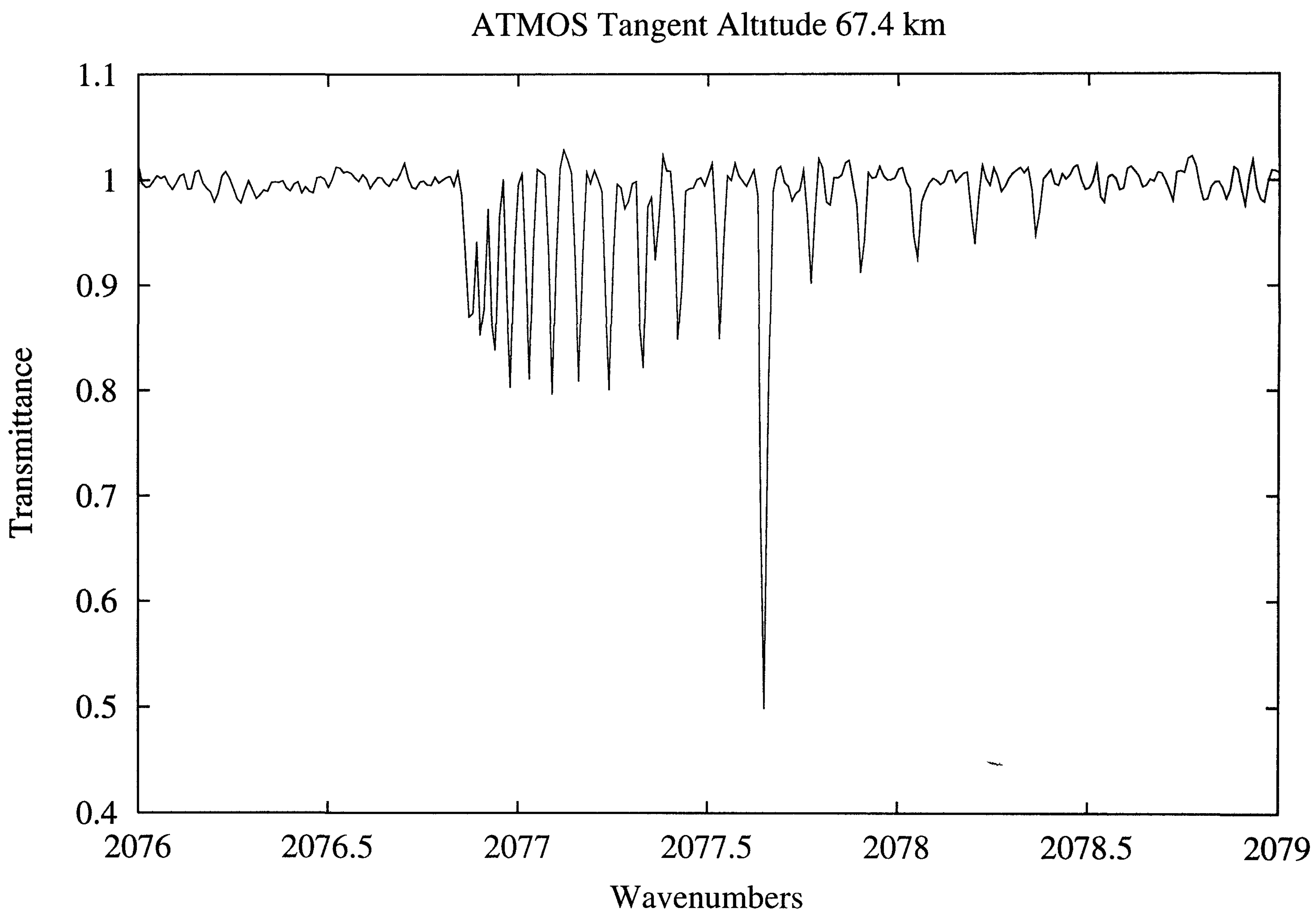
Chemical species detected by MIPAS

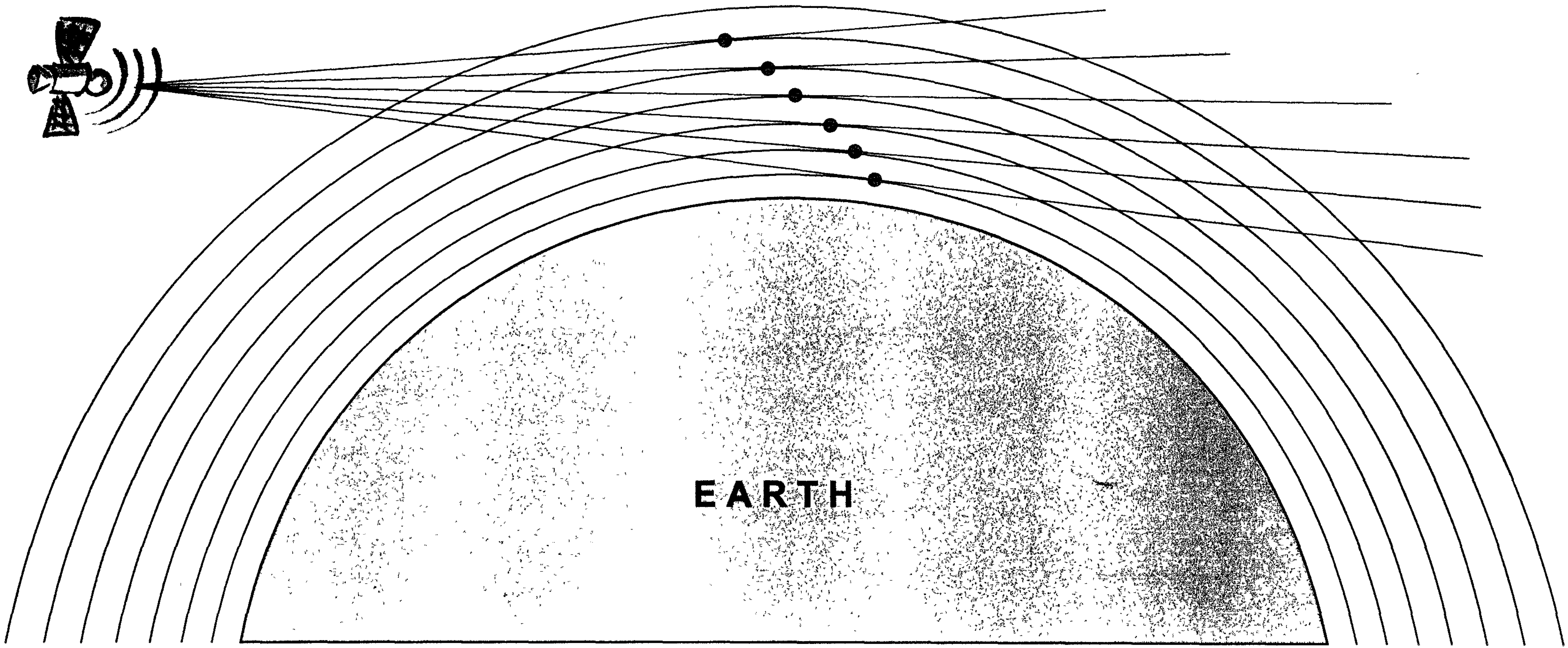


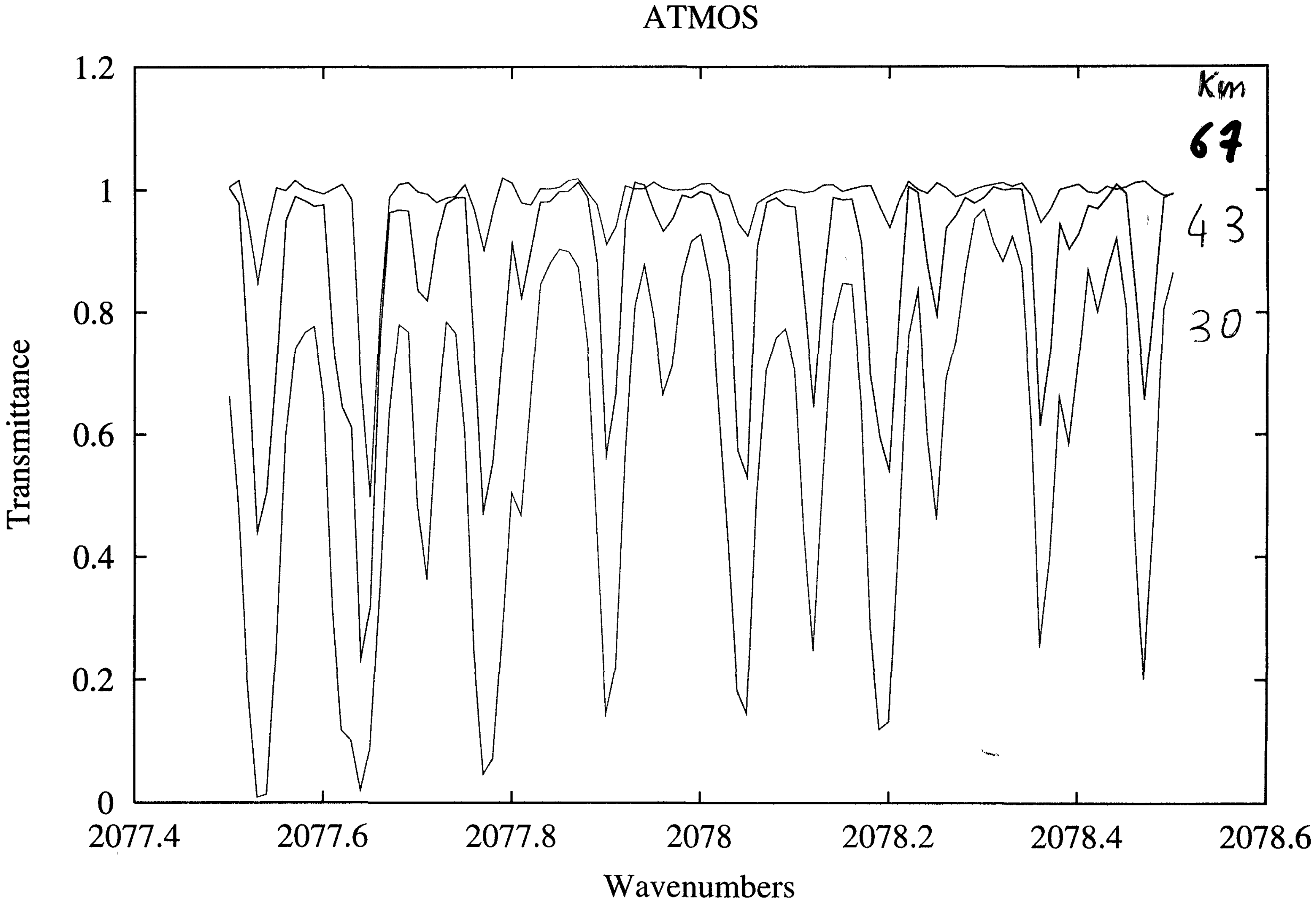
MIPAS




ATMOS Tangent Altitude 67.4 km
Transmittance
Wavenumbers
1.2
1
0.8
0.6
0.4
0.2
0
1900
2000
2100
2200
2300
2400


ATMOS Tangent Altıtude 67.4 km
Transmittance
Wavenumbers
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
2076
2076.5
2077
2077.5
2078
2078.5
2079


EARTH


ATMOS
1.2
1
0.8
0.6
0.4
0.2
0
Transmittance
2077.4
2077.6
2077.8
2078
2078.2
2078.4
2078.6
Wavenumbers
Km
67
43
30

Optimized Retrieval Model (ORM) Study

ESA - ESTEC Contract

Study Team:

• B. Carli, C. Piccolo, P. Raspollini	IROE - CNR, Italy
• M. Carlotti, L. Magnani, M. Ridolfi	University of Bologna, Italy
• B.M. Dinelli	ISM - CNR, Italy
• T. von Clarmann, M. Höpfner, H. Oelhaf, G. Stiller	IMK, Germany
• A. Dudhia, V. Jay, C. Rodgers, R.J. Wells	University of Oxford, UK
• J. Remedios	University of Leicester, UK
• J.M. Flaud	University of Paris, France

ESA - ESTEC Technical and Scientific Staff: J. Langen, G. Levrini, H. Nett

OBJECTIVES OF THE ORM

Development of scientific prototype of Level 2 code for near real time (NRT) retrievals from MIPAS spectra:

- temperature (error < 2 K)
- tangent pressure (error < 3%)
- volume mixing ratio of H_2O, O_3, HNO_3, CH_4 N_2O (NO_2) with 5% accuracy
- altitude range 6 - 68 km
- vertical resolution 3 km
- computing time of total retrieval < 1 hour on SUN SPARC 20

THE RETRIEVAL PROBLEM

$$\mathbf{n} = \mathbf{K}\,\mathbf{y}$$

- **n** is the difference between observation j and the corresponding simulation
- **K** is a matrix (usually denoted as Jacobian matrix). The entry k_{ij} of **K** is the derivative of observation i made with respect to parameter j
- $\mathbf{y} = \mathbf{q_Z} - \tilde{\mathbf{q}}_\mathbf{Z}$ is a vector. The entry y_i of **y** is the correction needed to the assumed value of parameter $\tilde{q}_Z$ in order to obtain its corrected value q_Z

The goal of the retrieval is the determination of vector **y**

Search for a "solution matrix" **D** that, multiplied by vector **n** provides **y**

$$\mathbf{y} = \mathbf{D}\,\mathbf{n}$$

If the vector $\mathbf{n}$ is characterised by the variance-covariance matrix $\mathbf{V}^n$

$$\mathbf{D} = [\mathbf{K}^{\mathrm{T}}(\mathbf{V}^n)^{-1}\mathbf{K}]^{-1}\mathbf{K}^{\mathrm{T}}(\mathbf{V}^n)^{-1}$$

That minimizes the χ^2 function defined as:

$$\chi^2 = \mathbf{n}^{\mathrm{T}}(\mathbf{V}^n)^{-1}\mathbf{n}$$

The errors associated with the solution to the inversion procedure can be characterised by the variance-covariance matrix $\mathbf{V}^q$ given by.

$$\mathbf{V}^q = \mathbf{D}(\mathbf{V}^n)\mathbf{D}^T = (\mathbf{K}^T(\mathbf{V}^n)^{-1}\mathbf{K})^{-1}$$

GENERAL FEATURES OF THE ADOPTED APPROACH

- use of narrow-band spectral intervals (microwindows)
- non-linear least-square fit
- global fit of limb scanning sequence
- sequential fit of species: pT, H_2O, O_3, HNO_3, CH_4, N_2O, NO_2

Unknowns of each retrieval

- profile of the parameter (one value for each tangent altitude)
- profile of the atmospheric continuum for each micro-window at altitudes < 30 km
- altitude independent instrumental continuum for each micro-window

DIAGRAM OF THE OPERATIONS FOR EACH RETRIEVAL

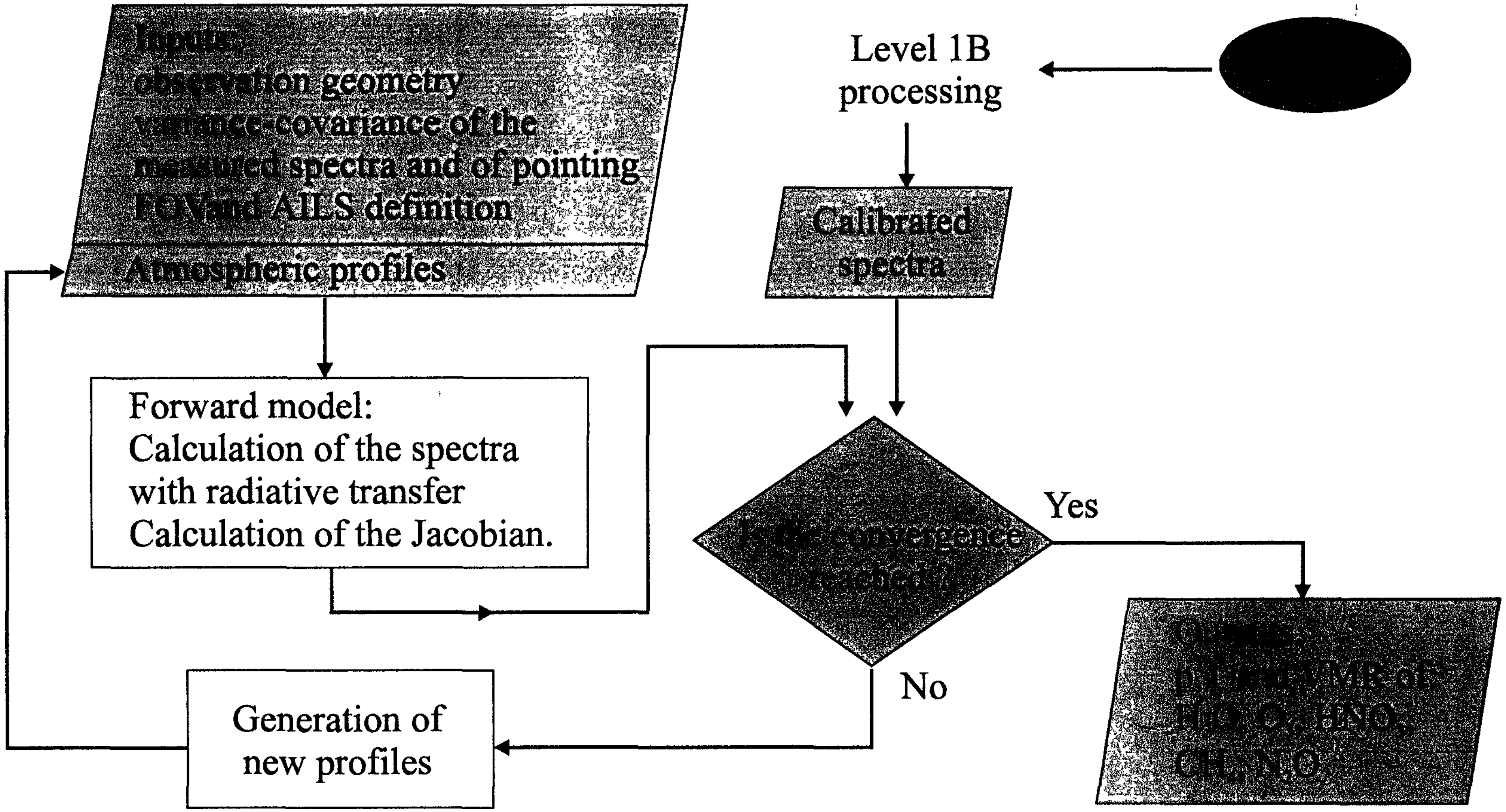
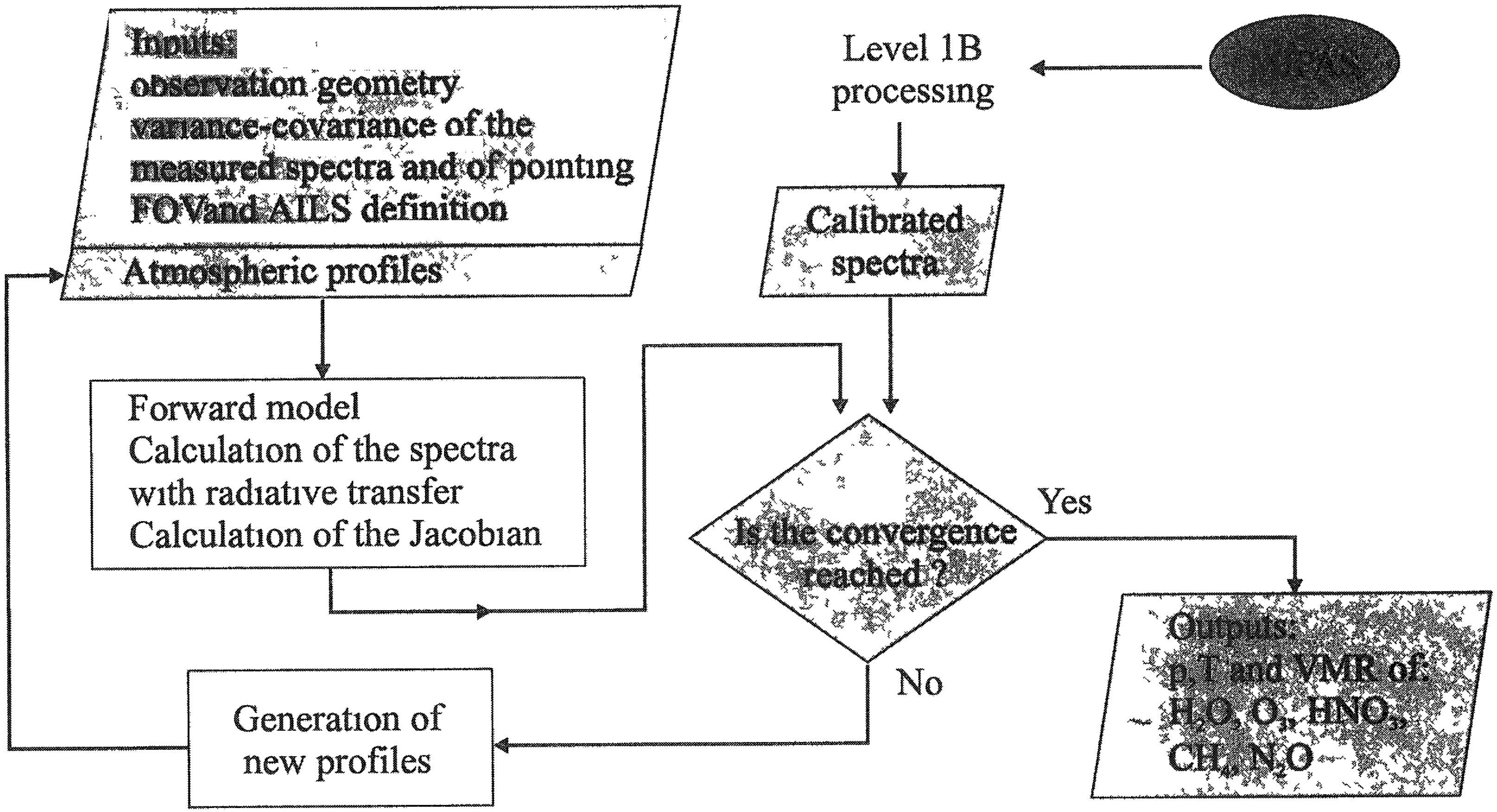


DIAGRAM OF THE OPERATIONS FOR EACH RETRIEVAL



Main components of the retrieval

a) Forward Model calculation

b) Jacobian calculation

c) Covergence criteria

a) Forward model

Equations

The forward model simulates the measurements. The measured signal is equal to the atmospheric radiance convoluted with instrument effects:

$$S_I(z_t,\sigma) = \iint \left(L(\sigma - \sigma', z_t - z_{t'}) \cdot AILS(\sigma')d\sigma' \right) \cdot FOV(z_{t'})dz_{t'}$$

and the atmospheric radiance is equal to:

$$L(\sigma, z_t) = \int_{x_i}^{x_0} [B(\sigma,x)c(\sigma,x)\eta(x)] \exp\left(-\int_{x}^{x_0} c(\sigma,x')\eta(x')dx' \right) dx$$

a) Forward model

Multiplicity of parameters and variables

- Cross sections. Summation of spectroscopic transitions
- Radiative transfer. Integral along the optical path
- Dependence on spectral frequency
- Dependence on limb angle and on instrumental field of view (FOV)

a) Forward model

Spectroscopic cross sections

- In the code, the cross-sections can either be calculated line-by-line (LBL) using pre-selected spectroscopic data and fast Voigt profile computations or be read from look-up tables (LUTs).
- The relative speed of the two approaches depends on the access time of the memory used to store LUTs.
- The baseline is to use Singular Value Decomposition to compress the LUTs so that they can be read from RAM.

a) Forward model

Radiative transfer

- use of Edlen model for refraction and WGS84 geoid model for geometrical corrections
- use of an atmospheric stratification common to all spectra
- use of relatively thick layers with "equivalent" pressure and temperature (Curtis-Godson approximation)
- use of a subset of the possible values of "equivalent" pressure and temperature (secant law approximation)

a) Forward model

Dependence on spectral frequency

The data analysis is performed at a “coarse grid” of 0.025 cm^{-1}.

Radiative transfer calculations require a “fine grid” which may be as small as 0.0005 cm^{-1}, but only a subset of fine grid points is used (“irregular grid”).

The irregular grid is altitude and limb angle independent.

a) Forward model

Dependence on spectral frequency

Interpolation of irregular grid and convolution with the apodised instrument line shape (AILS) are performed simultaneously

a) Forward model

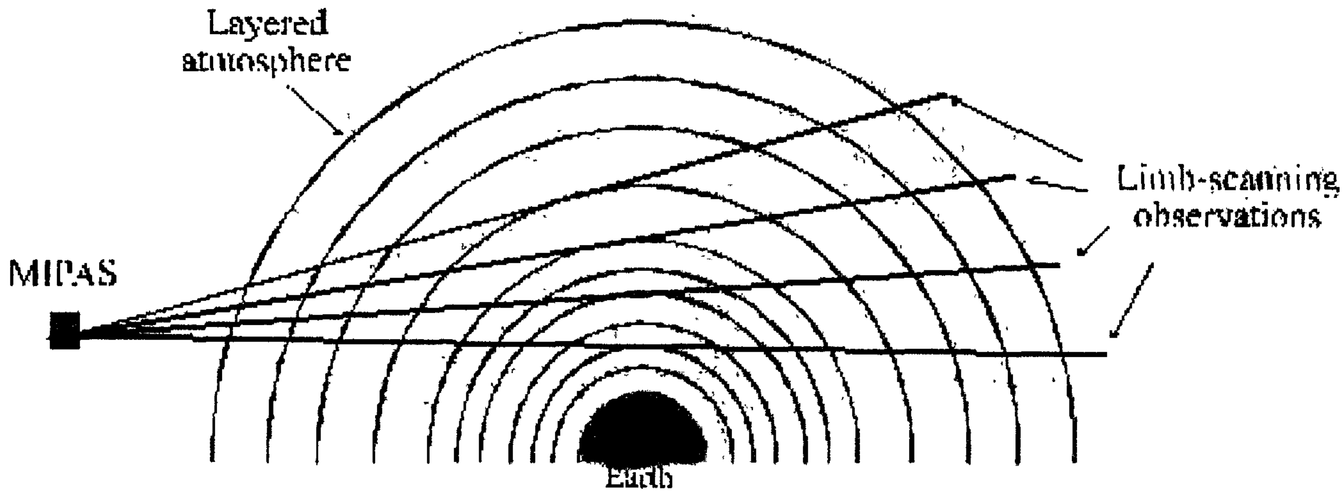
Convolution with instrument FOV

The spectra calculated at the limb scanning angles are interpolated to obtain the variation of the atmospheric radiance with the limb angle.

The interpolated function is used for the convolution with instrument FOV.

a) Forward model

- Use of secant law approximation: Curtis-Godson p and T (and therefore cross-sections) are calculated for only a sub-set of the actual paths in the atmosphere.



b) Jacobian

The calculation of the Jacobian is performed when intermediate results of forward model are available. For some unknowns an analytical optimisation is also possible.

Tangent pressure	Analytical
Temperature	Numerical
VMR	Analytical
Continuum	Analytical

c) Convergence criteria

Three convergence criteria are simultaneously used:

- **Linearity**: small difference between the calculated value of χ^2 and the value expected in linear approximation
- **Attained accuracy**: small variation in the estimate of the unknown
- **Computing time**: maximum number of iterations.

Implemented choices

A) Use of complementary information

B) Retrieval vertical grid

C) Vertical resolution

D) Atmospheric continuum

E) Pressure shift and line mixing

F) NLTE

A) Use of complementary information

- Hydrostatic equilibrium equation is used in p, T retrieval

- Engineering data on pointing are used in p, T retrieval

hydrostatic equilibrium

In a stratified atmosphere the pressure profile can be calculated from the temperature profile by means of the hydrostatic equation:

$$d\mathbf{P} = -\rho\ \mathbf{g}\ d\mathbf{z}$$

P pressure,
ρ density of the atmosphere,
g gravity acceleration,
z altitude.

Integrating:

$$\mathbf{P}_n = \mathbf{P}_o \exp\left(- \frac{M}{R}\ \mathbf{g} \sum_{i=0}^{n} \frac{1}{T_i}\ \Delta\mathbf{z}\right)$$

$\mathbf{P}_n$ pressure at a given altitude,
$\mathbf{P}_o$ pressure at the reference altitude,
M average molecular weight of the atmosphere,
R universal gas constant,
T_i temperature of the atmospheric layer i,
$\Delta\mathbf{z}$ altitude step between atmospheric layers.

B) Retrieval vertical grid

- The vertical profiles are retrieved at an altitude grid defined by the tangent altitude levels, since this provides the most accurate results
- The representation of the profile at an user defined vertical grid requires interpolation.

C) Vertical resolution

A 3 km step is a good resolution/accuracy compromise for both the measurement and the retrieval grid

D) Atmospheric continuum

- All sources that have a constant amplitude in a micro-window (3 cm^{-1}) are not modelled and are fitted as a continuum offset .
- The continuum is modelled as an altitude dependent absorption coefficient present at altitude < 30 km.
- Values of continuum in nearby microwindows could be made equal in order to reduce number of fitted parameters

E) Pressure shift and line mixing

- Modelling of pressure shift and line mixing is presently not included
- It can be added to cross-section LUTs without an increase of code computing time.

F) NLTE

- NLTE effects are not modelled in the code
- NLTE is, however, a source of error which is considered in the micro-window selection

Products: data

- Tangent pressures
- temperature profile
- VMR profiles of five gasses
- continuum profiles of fitted micro-windows
- VCM of the above retrieved quantities
- instrument offset for each micro-window

  χ^2 for each retrieval
- annotation data

ALGORITHM VALIDATION

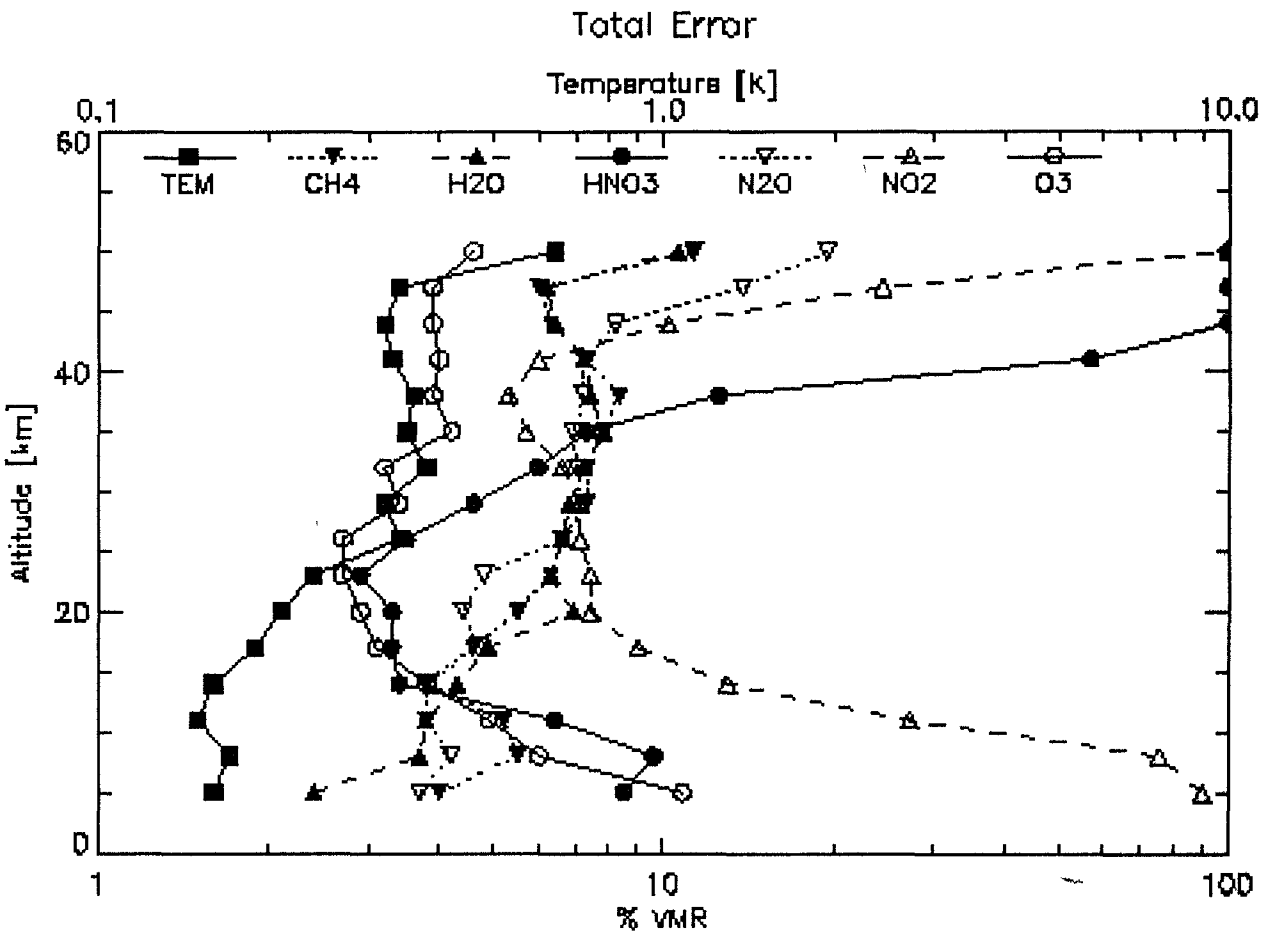
- The Forward Model was validated with respect to a Reference Forward Model (RFM)
- The retrieval code was tested with retrieval scenarios generated with both a local FM and with the RFM
- Tests with balloon (MIPAS-B2) measurements are in progress

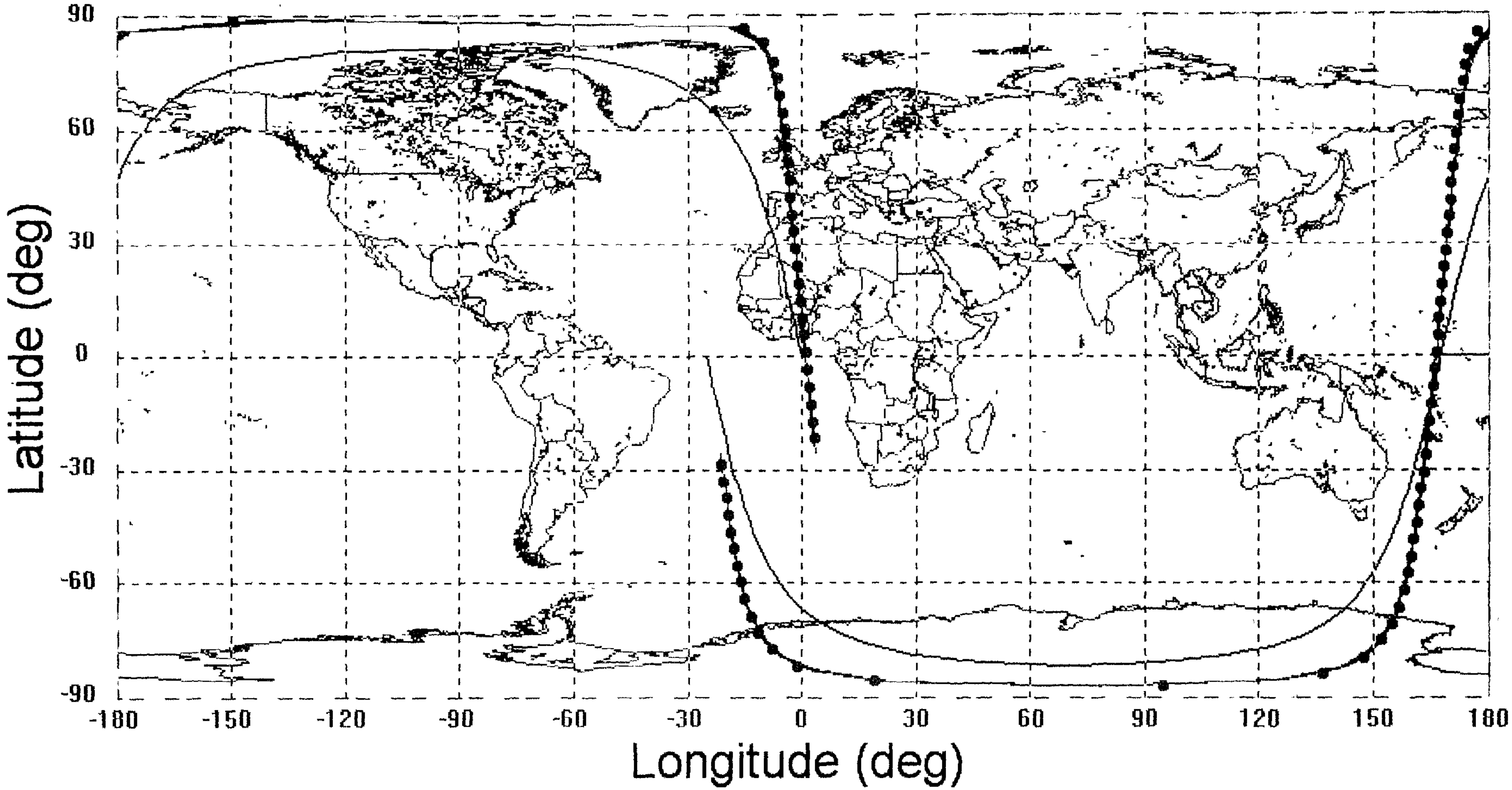
RUNTIME PERFORMANCE

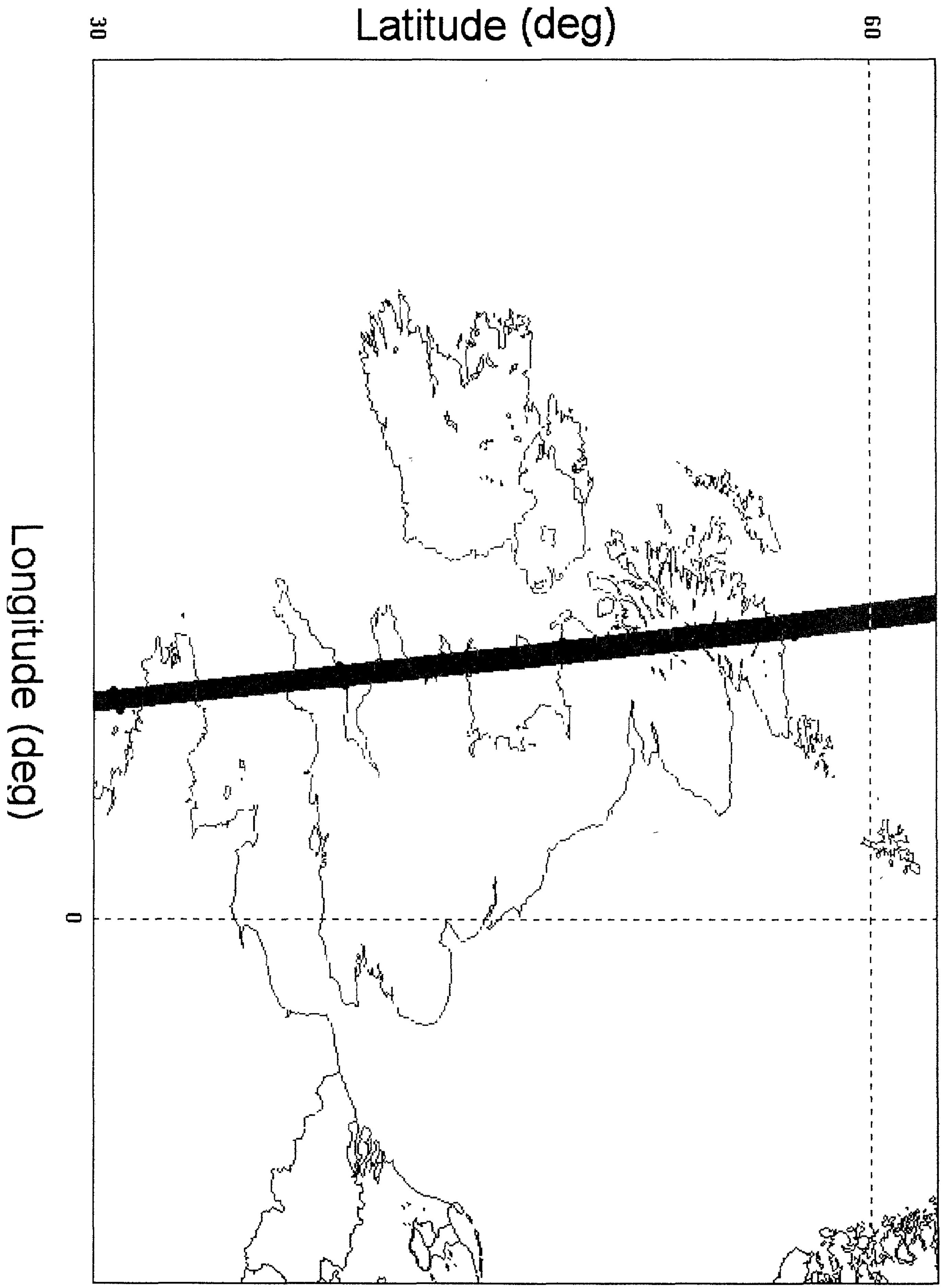
Computer description	Standard set of MWs	Optimized set of MWs
SUN SPARC station 20 120 MHz CPU, 128 Mb RAM	550 (*)	348 (*)
PENTIUM PC 200 MHz CPU, 256 Mb RAM	352	210
Ultra Sparc station 5	181	Not Available
IBM RS6000 Model 397	149	Not Available
Digital DEC-SERVER Mod 4100 600 MHz CPU, 1 Gb RAM	74	51

(*) Runtime strongly affected by the use of swap space

ORM ACCURACY FIGURES: TOTAL ERROR BUDGET SUMMARY




Latitude (deg)
90
60
30
0
-30
-60
-90
-180
-150
-120
-90
-60
-30
0
30
60
90
120
150
180
Longitude (deg)


Latitude (deg)
30
60
Longitude (deg)
0

SLIMCAT
Atmosphere

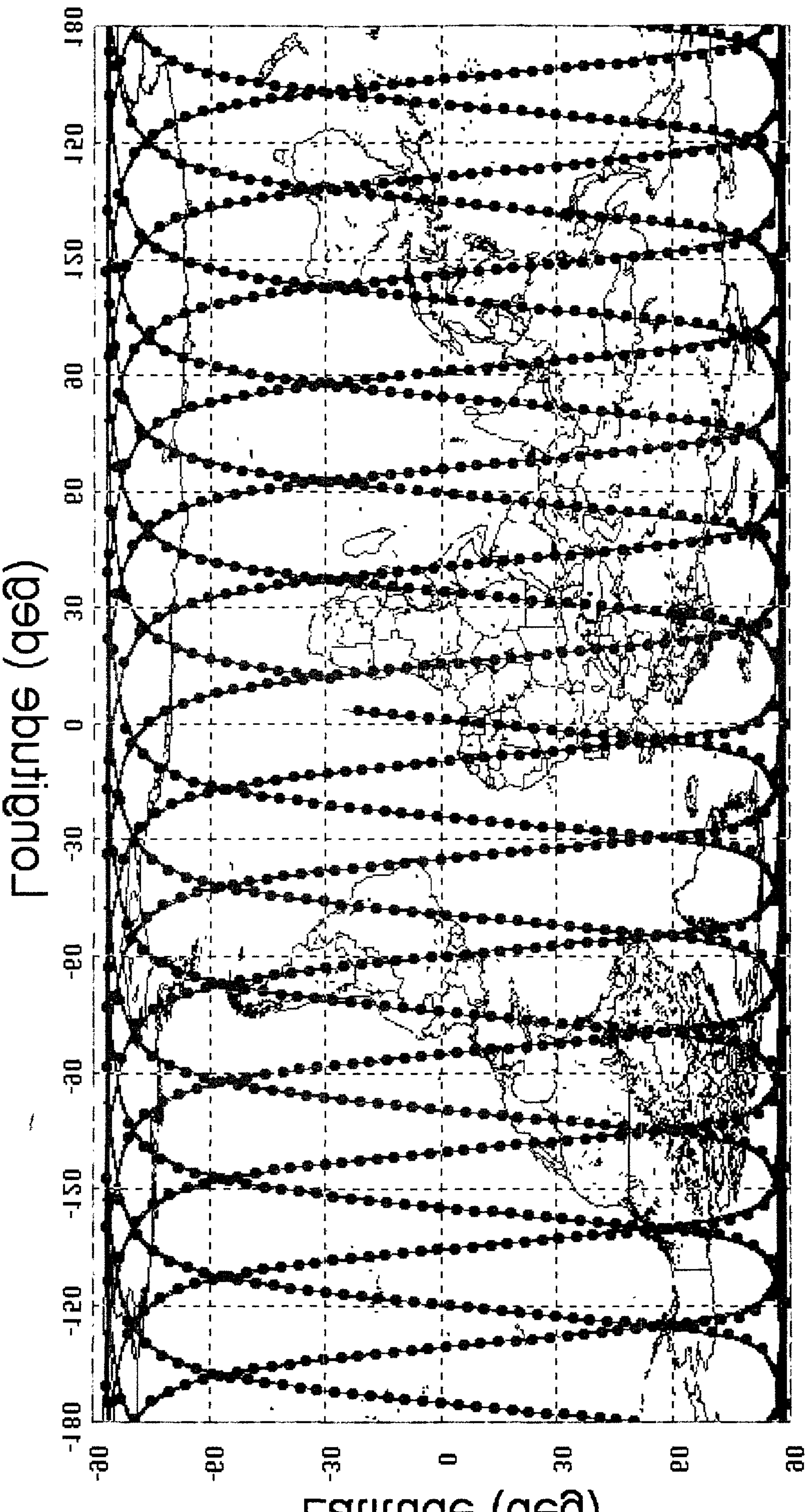
3-D chemical transport model

- latitude grid: 5°, 4.5° in polar regions
- longitude grid: 7.5°
- altitude grid: 8-57 km irregular steps of ≈ 2, 4, 5 km

Day: 27 September 1996 (ozone hole conditions)

Quantities: p, T, VMR of:
H_2O, O_3, N_2O, CO, CH_4, NO, NO_2, HNO_3, OH, HF, HCl, HBr, ClO, CH_2O, HOCl, H_2O_2, COF_2, HO_2, O, $ClONO_2$

M.P. Chipperfield, "Multiannual Simulations with a Three-Dimensional Chemical Transport Model", *J. Geophys. Res.*, **104**, 1781-1805 (1999).


Longitude (deg)
Latitude (deg)
-180 -120 -150 -90 -60 -30 0 30 60 90 120 150 180
-90 -60 -30 0 30 60 90

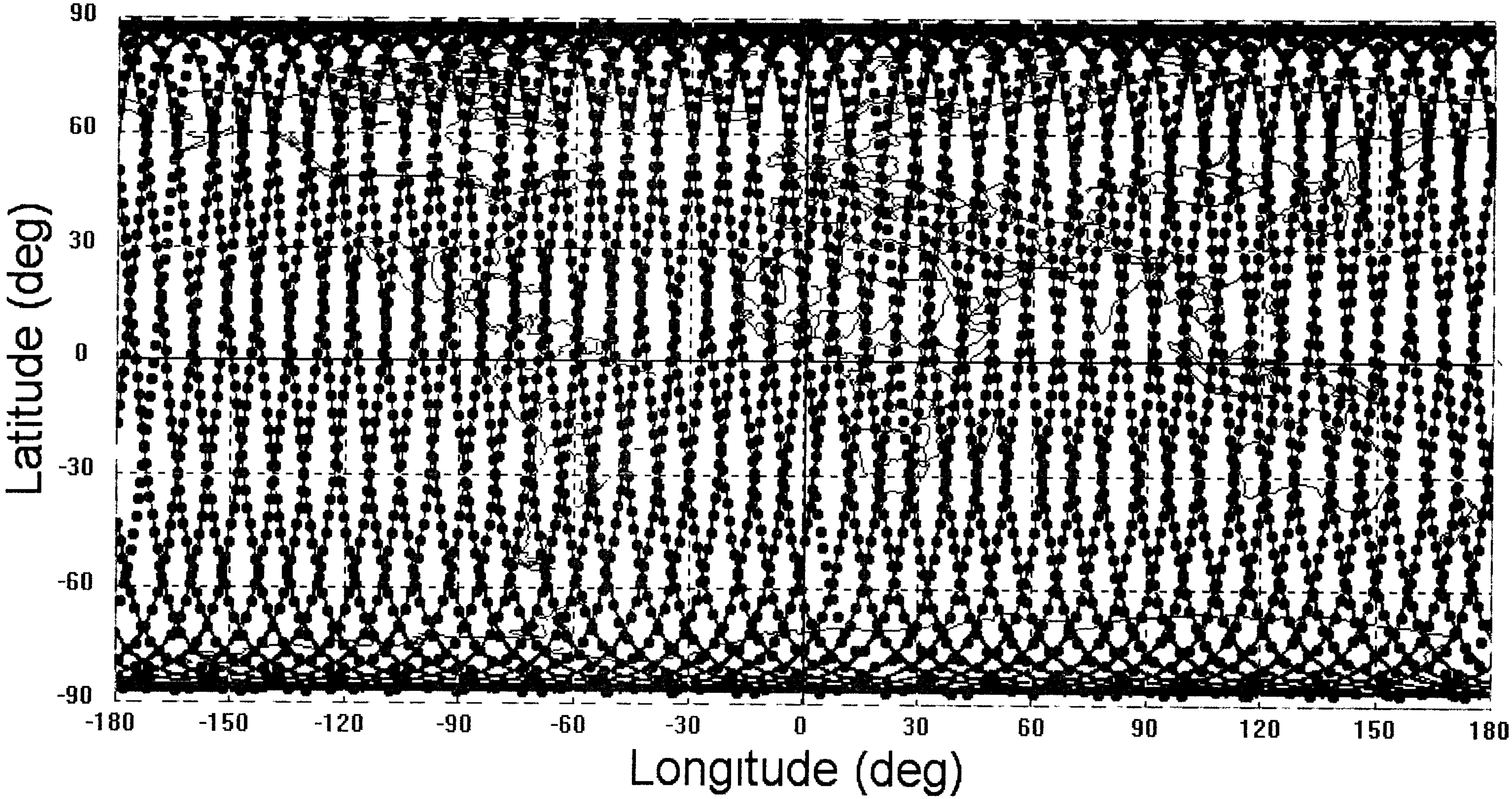

Latitude (deg)
90
60
30
0
-30
-60
-90
-180
-150
-120
-90
-60
-30
0
30
60
90
120
150
180
Longitude (deg)

FIGURE 52

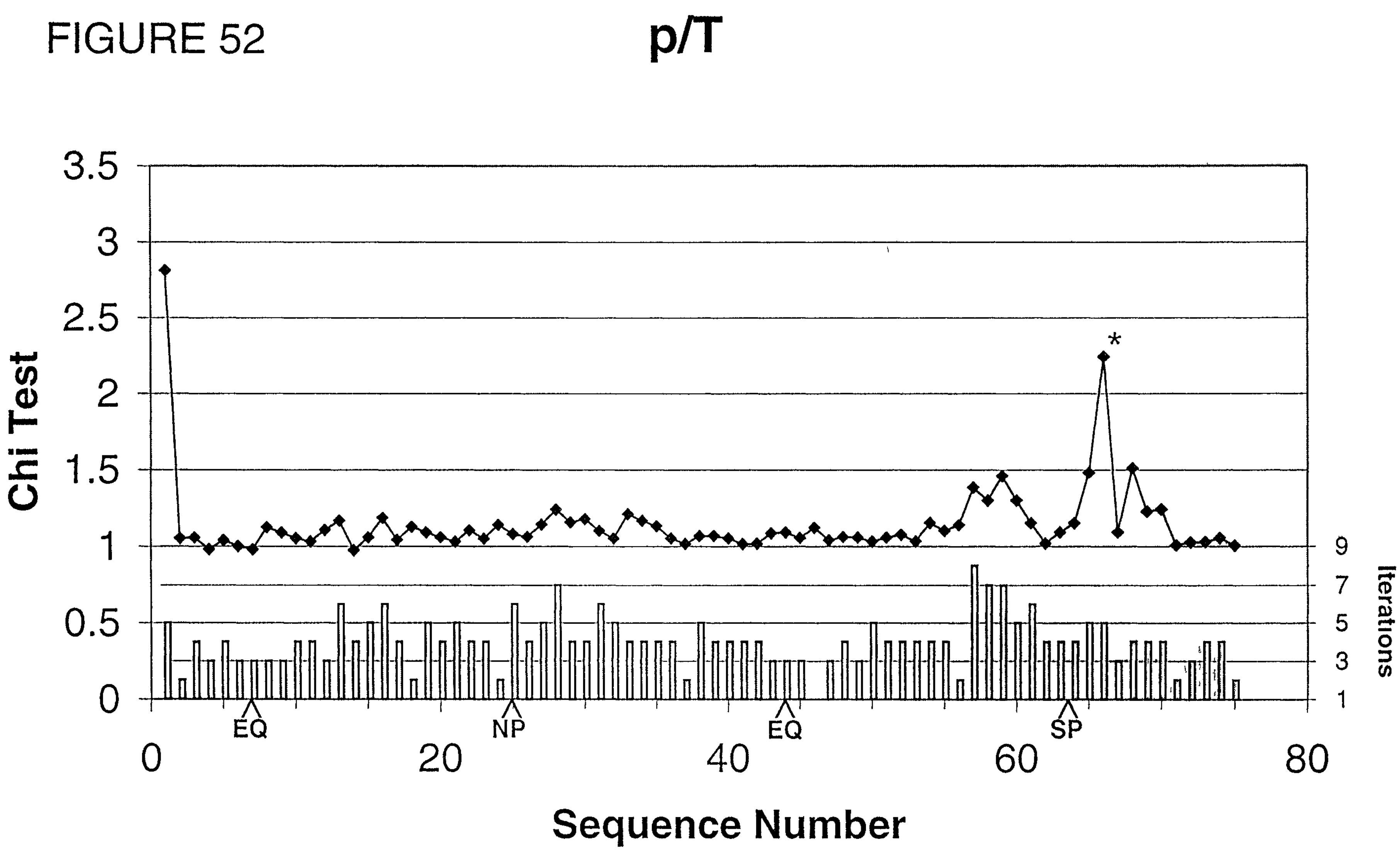

p/T
Chi Test
3.5
3
2.5
2
1.5
1
0.5
0
*
9
7
5
3
1
Iterations
EQ
NP
EQ
SP
0
20
40
60
80
Sequence Number

FIGURE 53

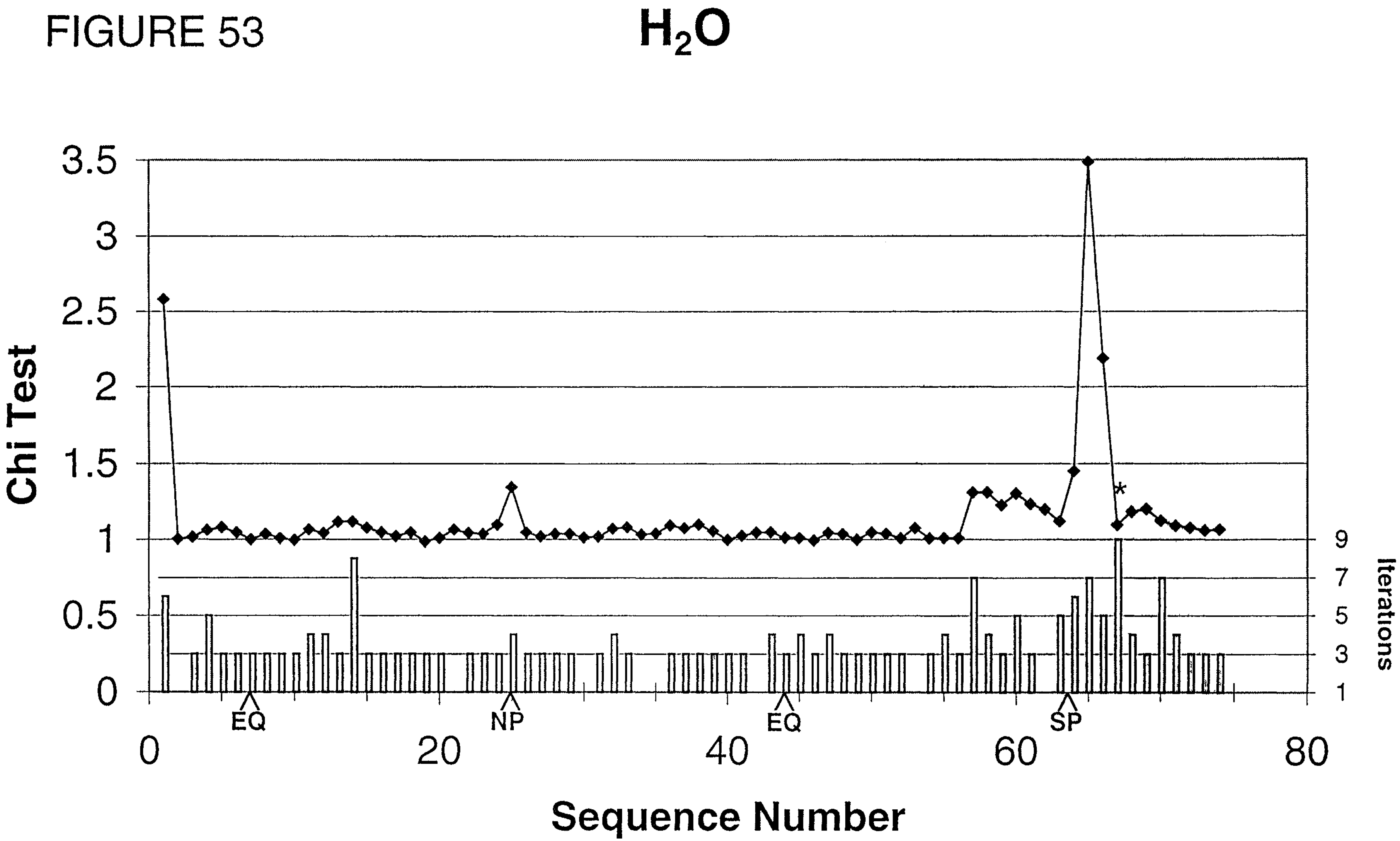
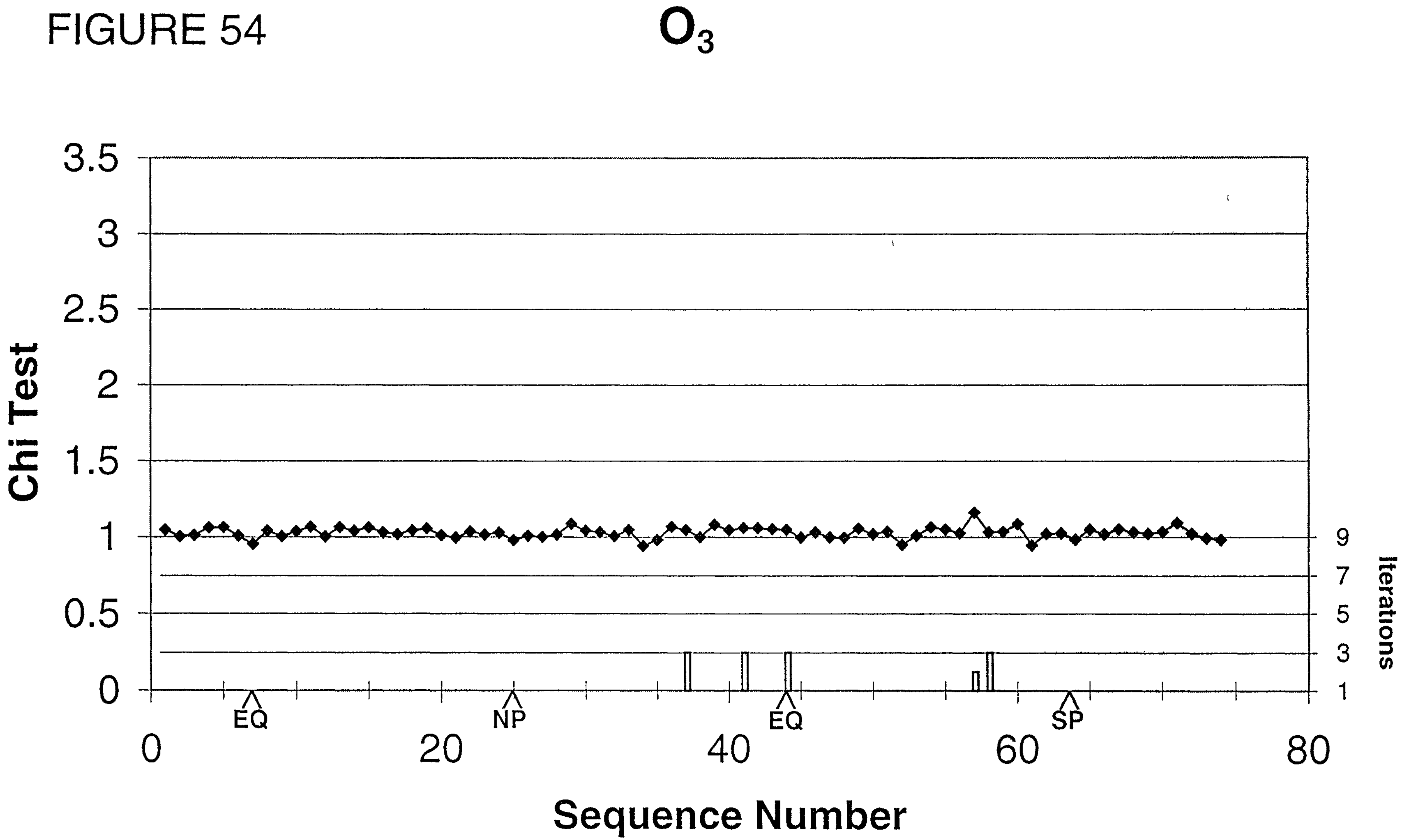
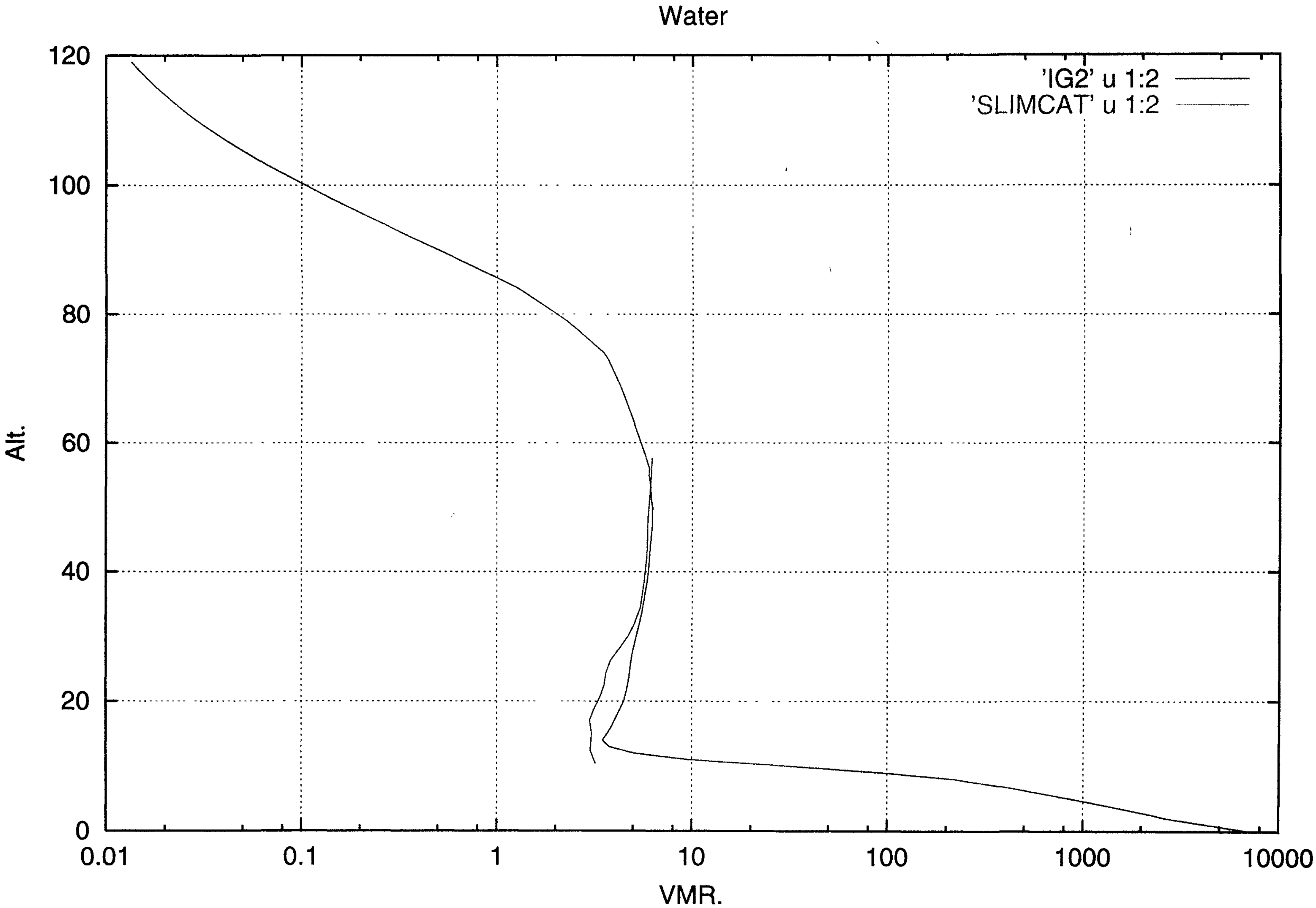

H2O
Chi Test
3.5
3
2.5
2
1.5
1
0.5
0
*
9
7
5
3
1
Iterations
EQ
NP
EQ
SP
0
20
40
60
80
Sequence Number

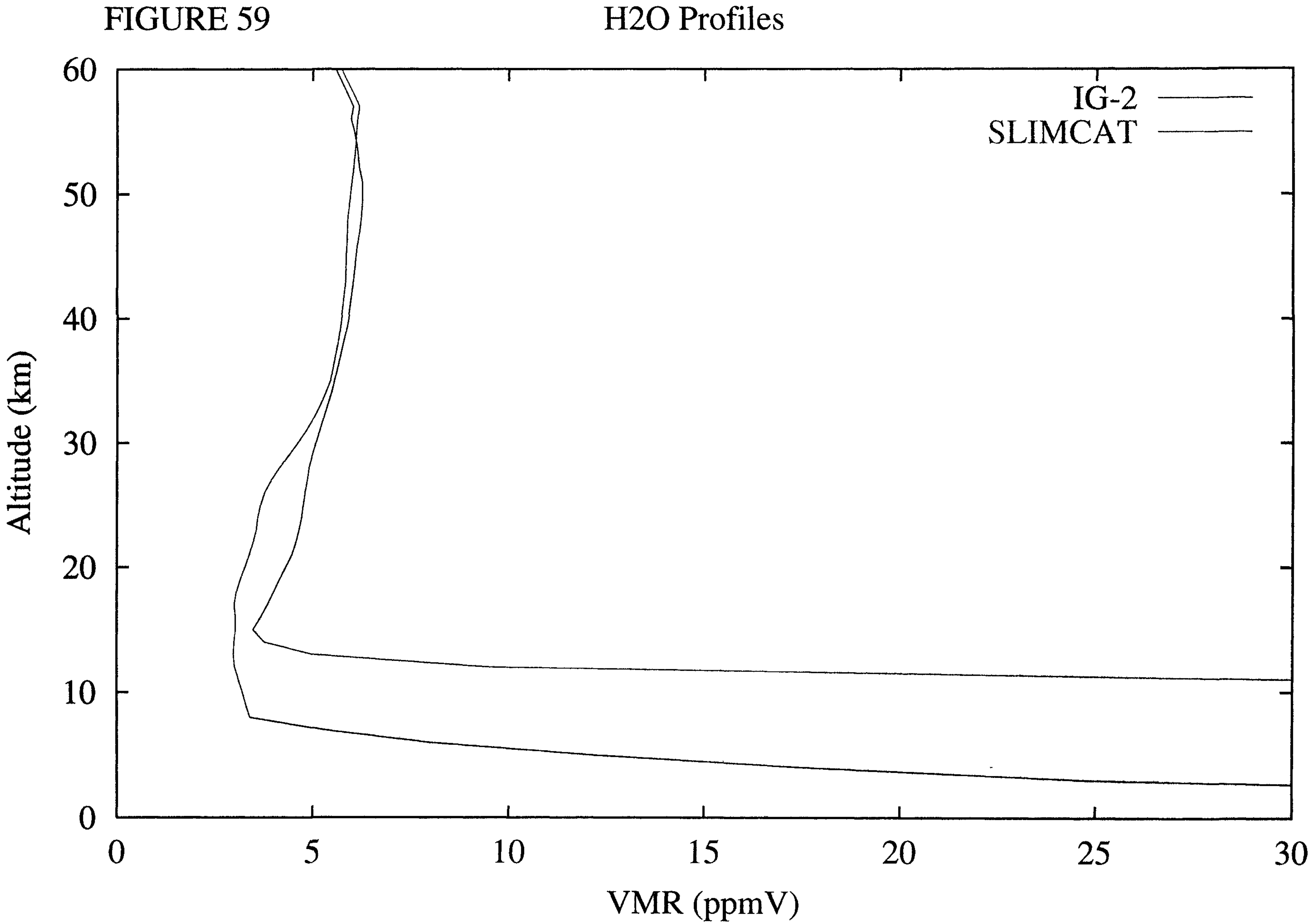
FIGURE 54


O$_3$
Chi Test
3.5
3
2.5
2
1.5
1
0.5
0
EQ
NP
EQ
SP
0
20
40
60
80
Sequence Number
9
7
5
3
1
Iterations



Water
'IG2' u 1:2
'SLIMCAT' u 1:2
Alt.
120
100
80
60
40
20
0
0.01
0.1
1
10
100
1000
10000
VMR.

FIGURE 59


H2O Profiles
IG-2
SLIMCAT
Altitude (km)
VMR (ppmV)
0
10
20
30
40
50
60
0
5
10
15
20
25
30

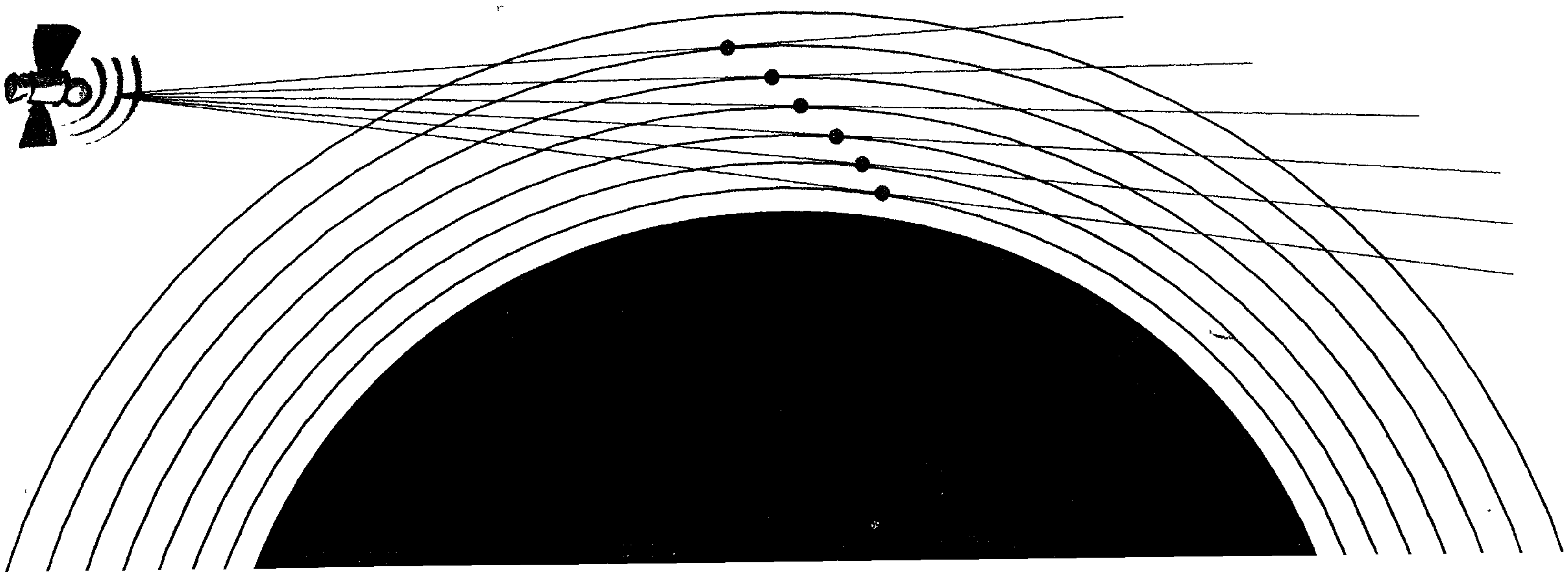
INVERSION METHODS

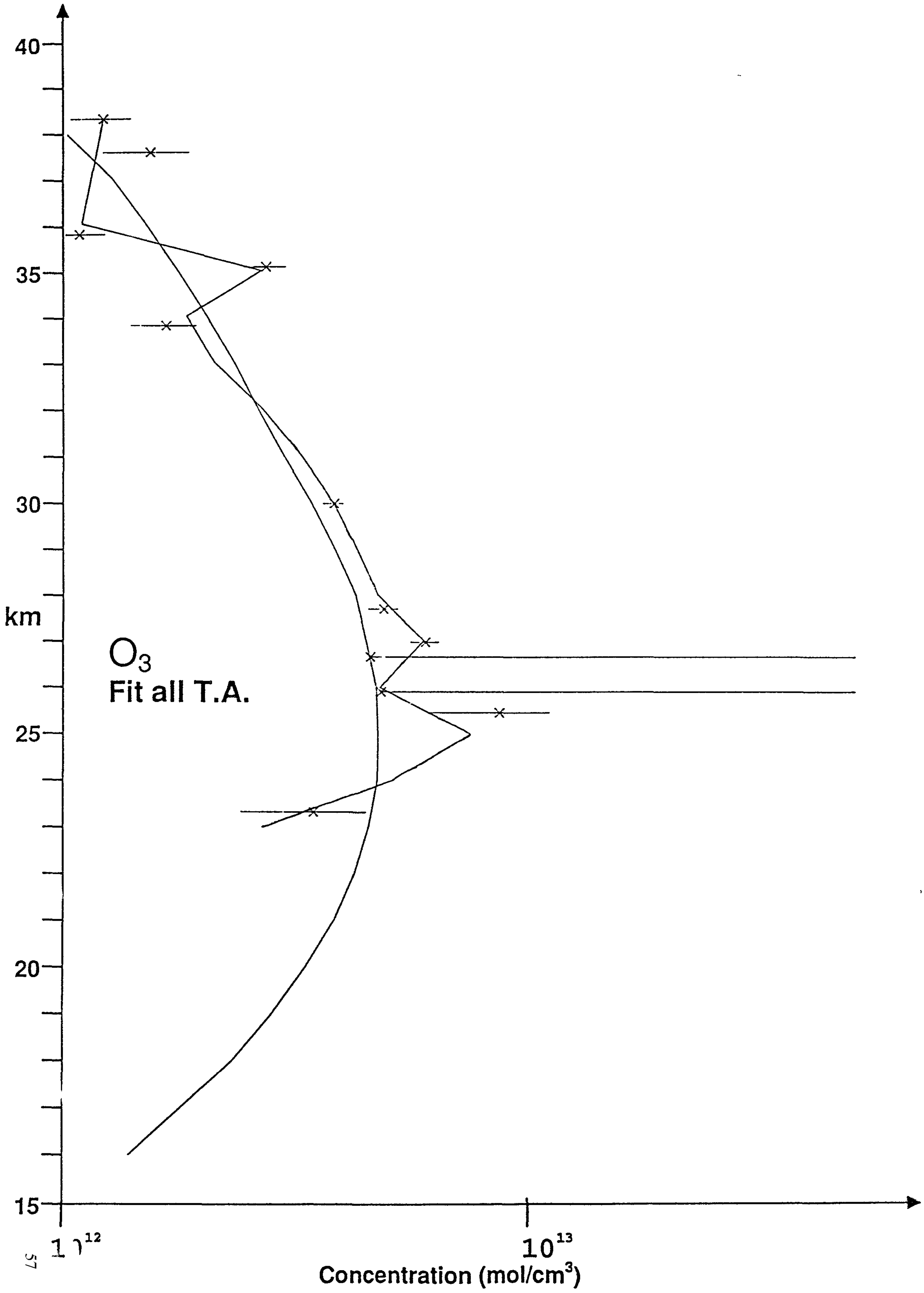
Onion peeling

- The retrieval process starts from the uppermost limb observation and proceeds downwards
- An error propagation applies which is difficult to assess
- One-to-one correspondence is established between the limb-views and the retrieval altitudes

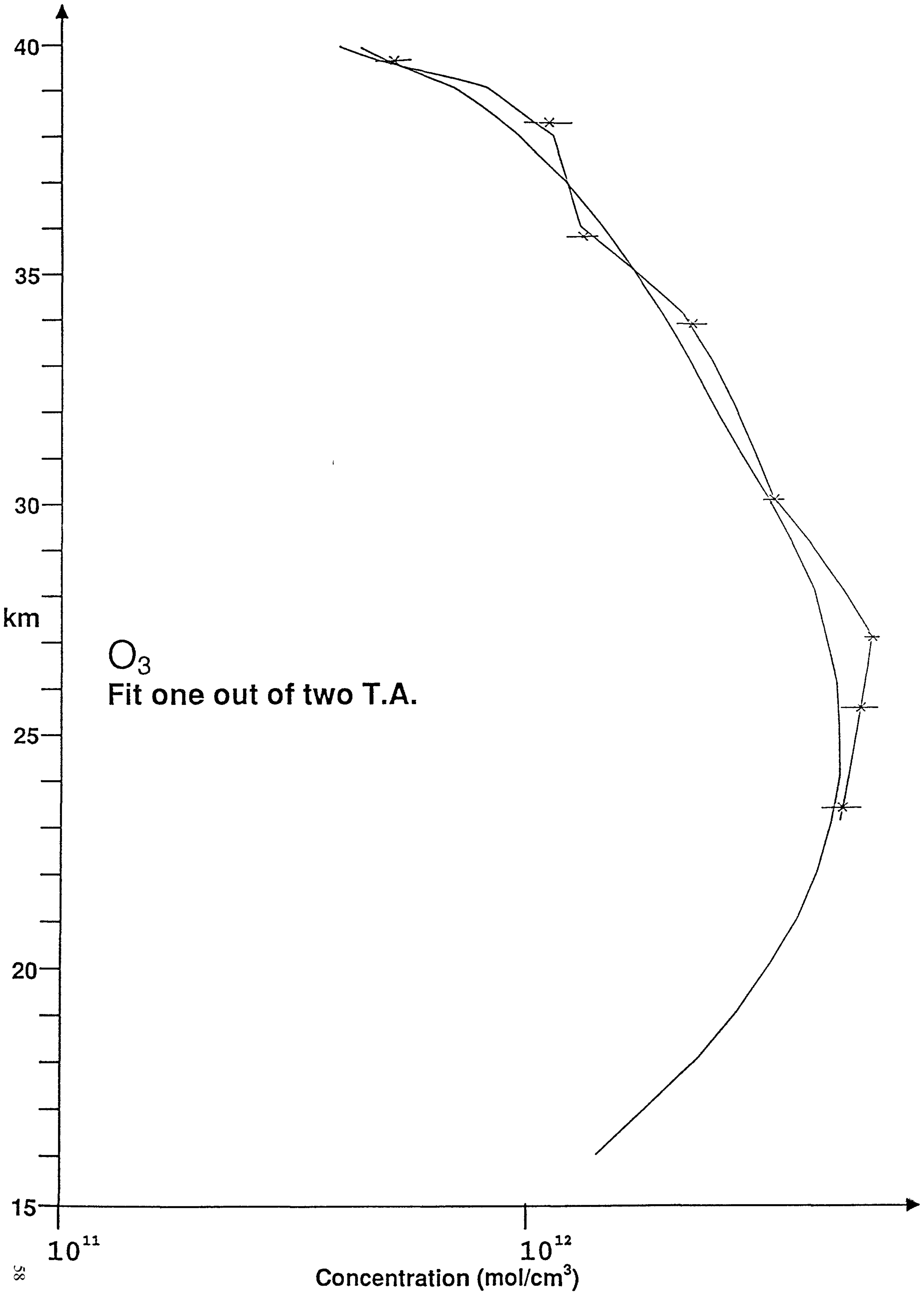
Global-Fit

- The whole profile is simultaneously retrieved to fit all the observations of a limb-scanning sequence
- Retrieval altitudes can be different from observation altitudes
- The ESDs derived from the VCM of the retrieval permit to assess the trade-off between uncertainty and vertical resolution





40
35
30
km
25
20
15
O_3
Fit all T.A.
10^{12}
10^{13}
Concentration (mol/cm^3)

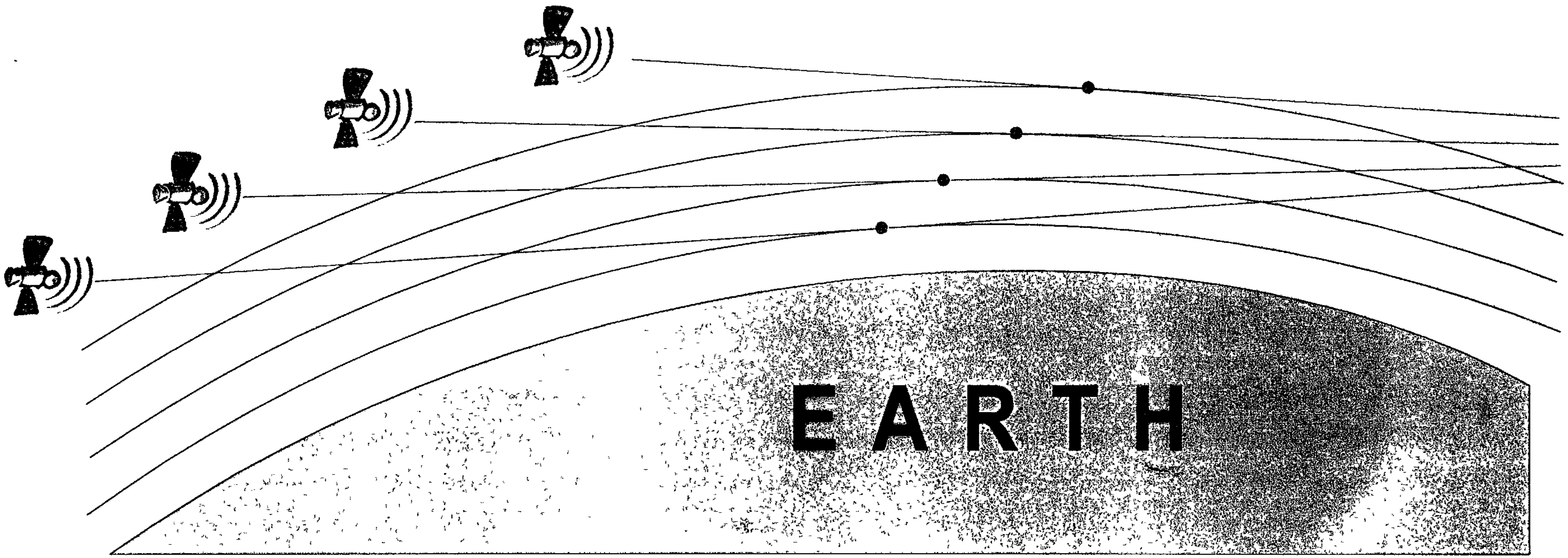

40
35
30
km
25
20
15
O_3
Fit one out of two T.A.
10^{11}
10^{12}
Concentration (mol/cm^3)

In the case of a steady platform

The assumption of horizontal homogeneity applies to a length of several hundreds of kilometers

In the case of an orbiting platform

The assumption of horizontal homogeneity extends to a larger scale also because of the movement of the platform during the recording time. For "along track" observations the assumption may extend to about 2000 km.



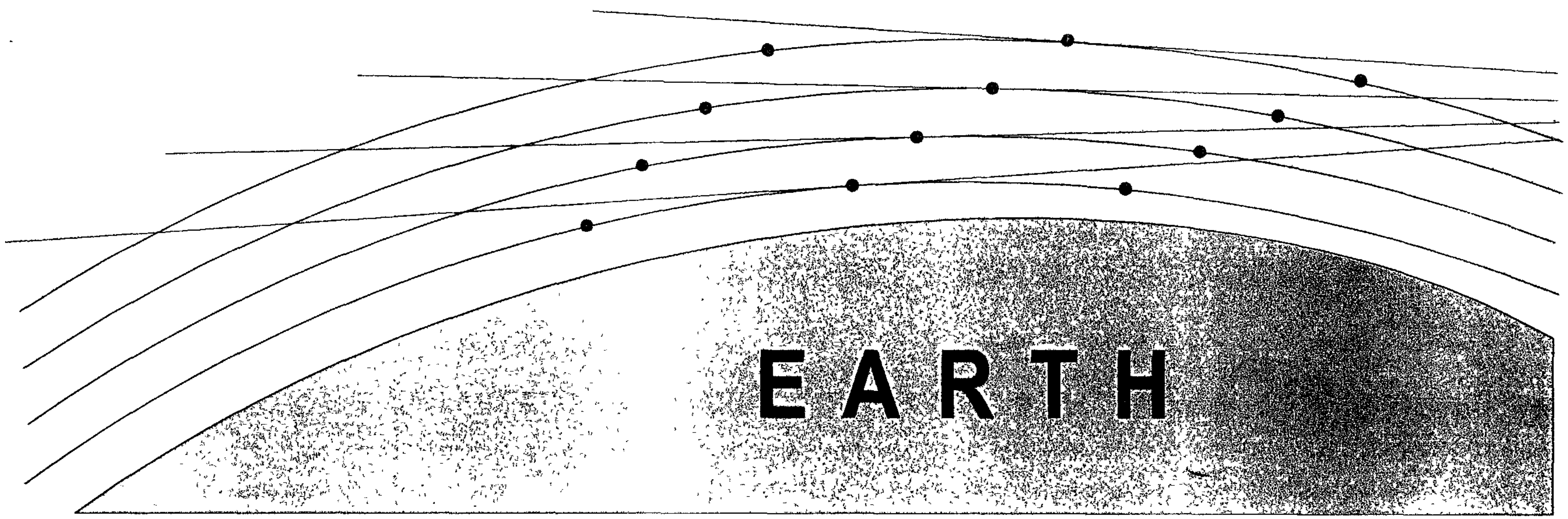
THE PROBLEM

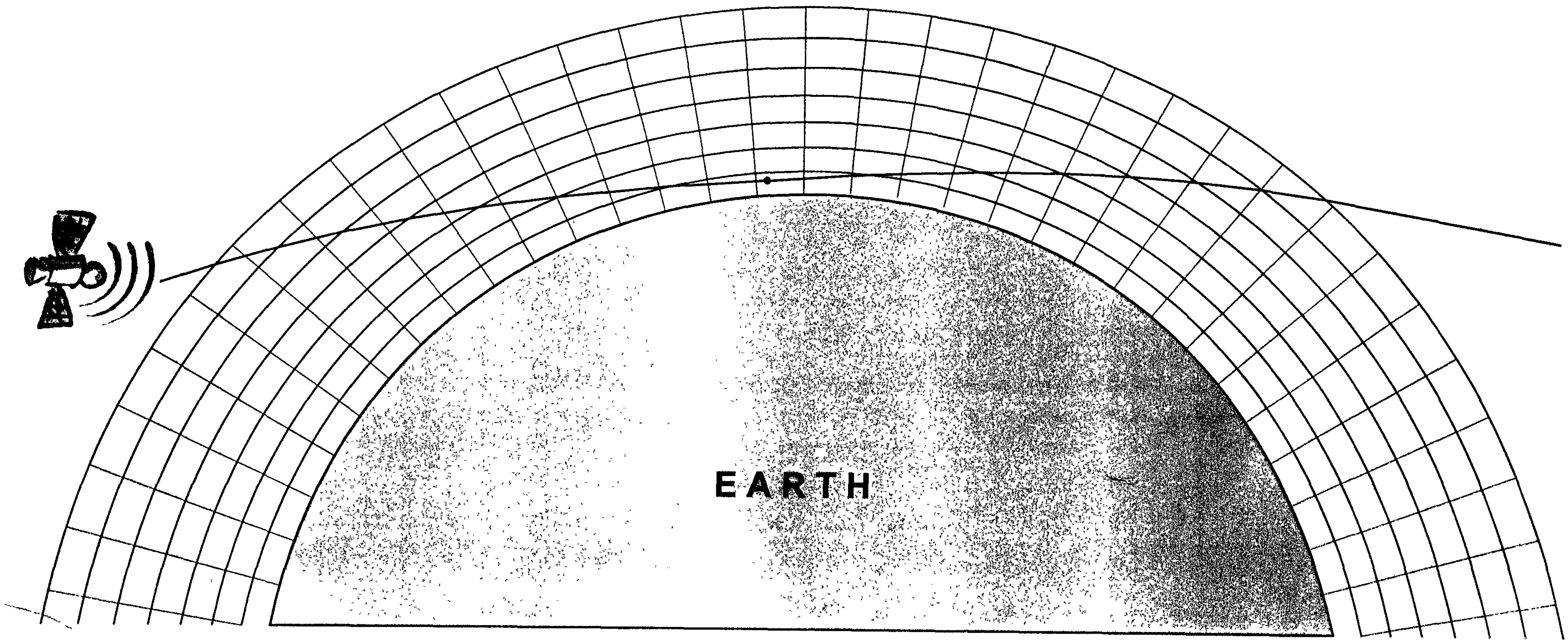
To model the horizontal variability of the atmosphere when retrieving physical and chemical parameters from limb-scanning observations

Rationale for *GEO-FIT*

- The information content about a given location of the atmosphere can be gathered from all the lines of sight that cross that location.

- Since the loop of cross-talk between nearby sequences closes when the starting sequence is reached again at the end of the orbit, in a retrieval analysis the full gathering of information can be obtained by merging in a simultaneous fit the observations of a complete orbit




EARTH

DEFINITION OF ATMOSPHERIC PROPERTIES

Atmospheric properties are associated to each clove starting from the quantities at the nodes that identify the clove itself. They are:

- ***average value of the refraction index***
- ***average value of the atmospheric continuum*** (c_{cont})

Atmospheric properties are associated to each individual path p (within a clove) that results from the ray tracing of a line of sight. For each gas (g) are calculated:

- ***slant column*** ($col_{g,p}$),
- ***equivalent pressure and temperature***
- ***absorption cross section*** ($c_{g,p}$)

RETRIEVAL PROCEDURE

$$\mathbf{y} = (\mathbf{K}^T \mathbf{S}_n^{-1} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}_n^{-1} \mathbf{n}$$

$\mathbf{y}$ = vector of the corrections to the assumed value of the parameters
$\mathbf{n}$ = vector of the differences between observations and simulations
$\mathbf{S_n}$ = VCM associated to vector $\mathbf{n}$
$\mathbf{K}$ = Jacobian matrix

Retrieval Grid

- *nominal* location of tangent points
- defined through an input file

PERFORMANCE OF GEO-FIT

ATMOSPHERIC MODEL

SLIMCAT: three-dimensional chemical transport model developed by Chipperfield for the day September 27, 1996.

pressure, temperature, and VMR profiles on a grid of 7.5° in longitude and 5° in latitude except for the polar regions where the latitudinal grid is 4 5°.

atmospheric-continuum profiles calculated using the Clough model

RETRIEVAL

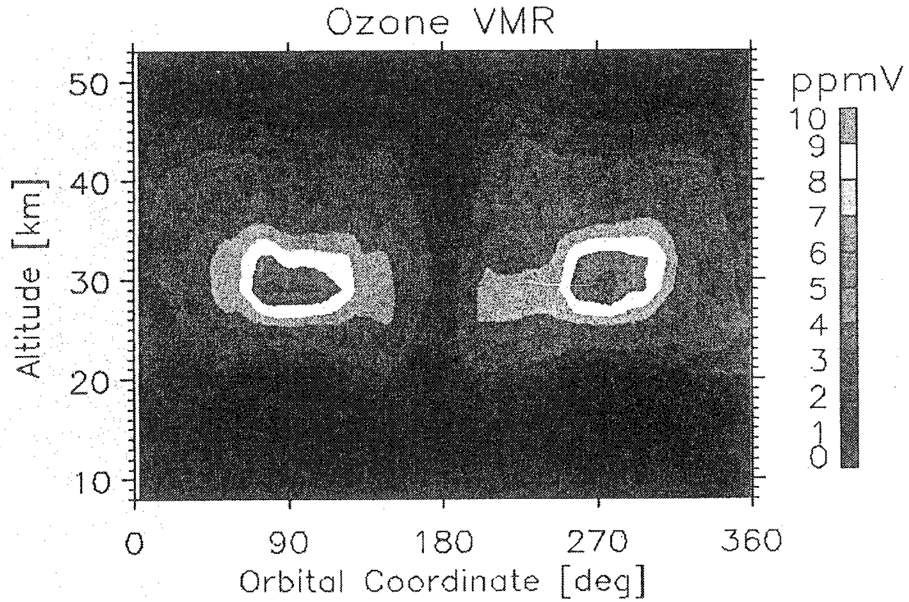
MIPAS ozone microwindows simulated for the **polar orbit** that overpasses Greenwich

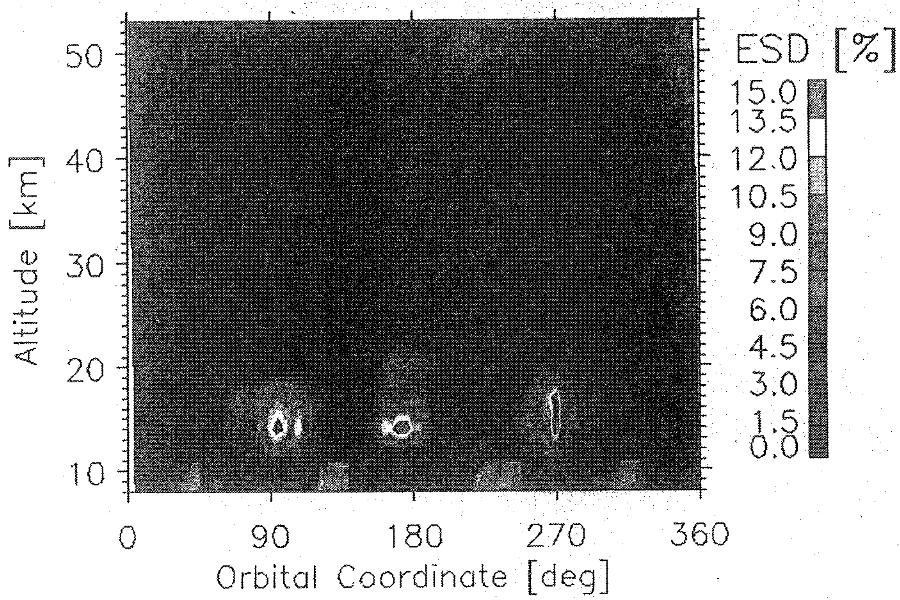
Initial-guess profiles obtained applying a perturbation factor to the reference profiles. This factor is different for each profile and is determined by a random-numbers generator within given thresholds. Threshold of ±30% for ozone VMR and ±50% for atmospheric-continuum.

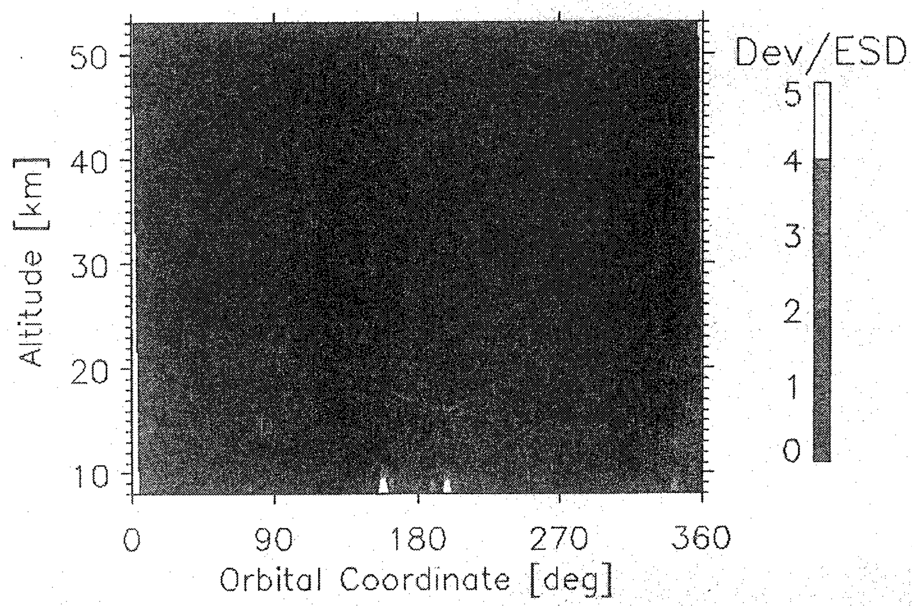
retrieval grid defined by the *nominal* location of the tangent altitudes

One Gauss-Newton iteration to reach convergence

GEO-FIT analysis
Retrieval at tangent altitudes







Performance of geo-fit on ozone VMR retrieval

Altitude (km)	Average ESD (%)	Maximum ESD (%)	Orbital coordinate	Average dev (%)	Maximum dev (%)	Orbital coordinate
53	4.0	6.1	352.0	2.4	8.0	104.7
50	5.7	8.7	352.0	3.8	12.6	37.2
47	4.4	8.0	347.9	2.8	10.8	347.9
44	3.1	5.6	357.2	1.9	8.0	222.2
41	2.4	4.9	352.0	1.6	6.3	294.4
38	2.1	4.3	352.0	1.4	4.6	347.9
35	2.0	4.2	357.2	1.4	5.0	321.8
32	2.1	3.8	347.9	1.6	6.5	357.2
29	2.5	3.9	11.8	2.0	6.1	340.9
26	3.5	4.9	241.5	2.6	9.0	241.5
23	3.2	5.1	174.2	1.9	7.2	169.7
20	3.1	9.0	165.5	1.8	6.0	174.2
17	4.8	16.5	269.2	2.7	16.8	269.2
14	6.0	17.9	269.2	4.6	22.6	179.4
11	2.9	6.4	89.6	2.1	7.3	103.1

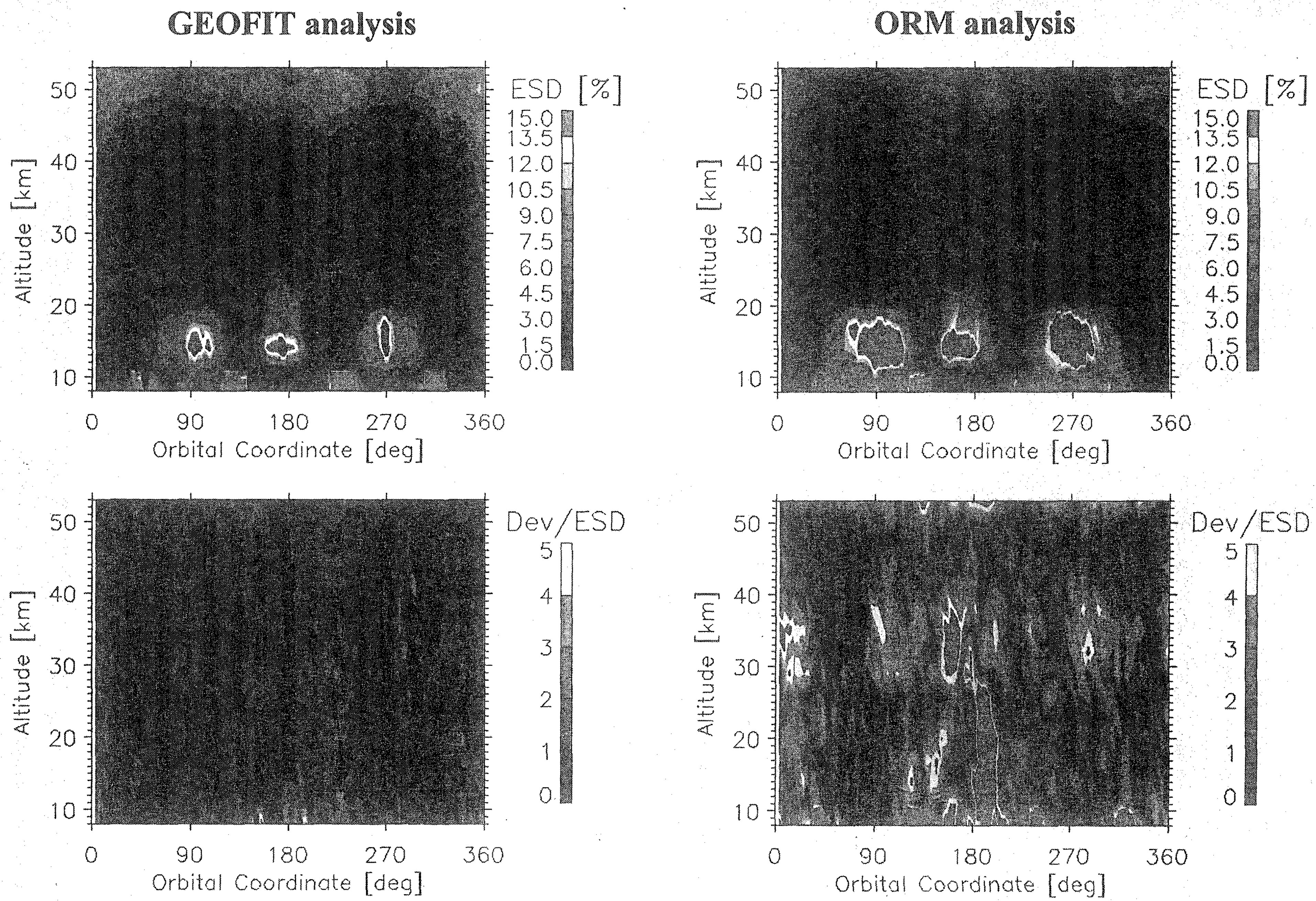
ASSESSMENT OF THE ERRORS DUE TO HORIZONTAL HOMOGENEITY ASSUMPTION

(Geo-fit vs. ORM)

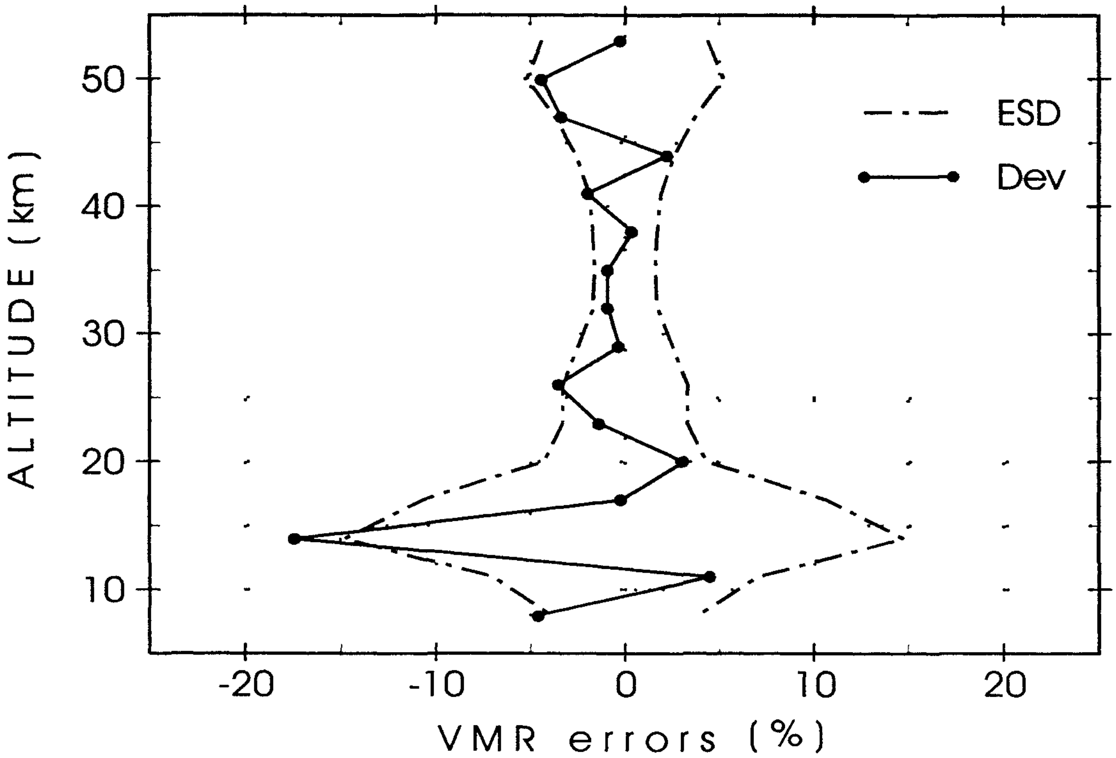
On observations generated by a homogeneous atmosphere:

In order to provide comparable ESDs, ORM requires a number of observations about 3 times larger than geo-fit.

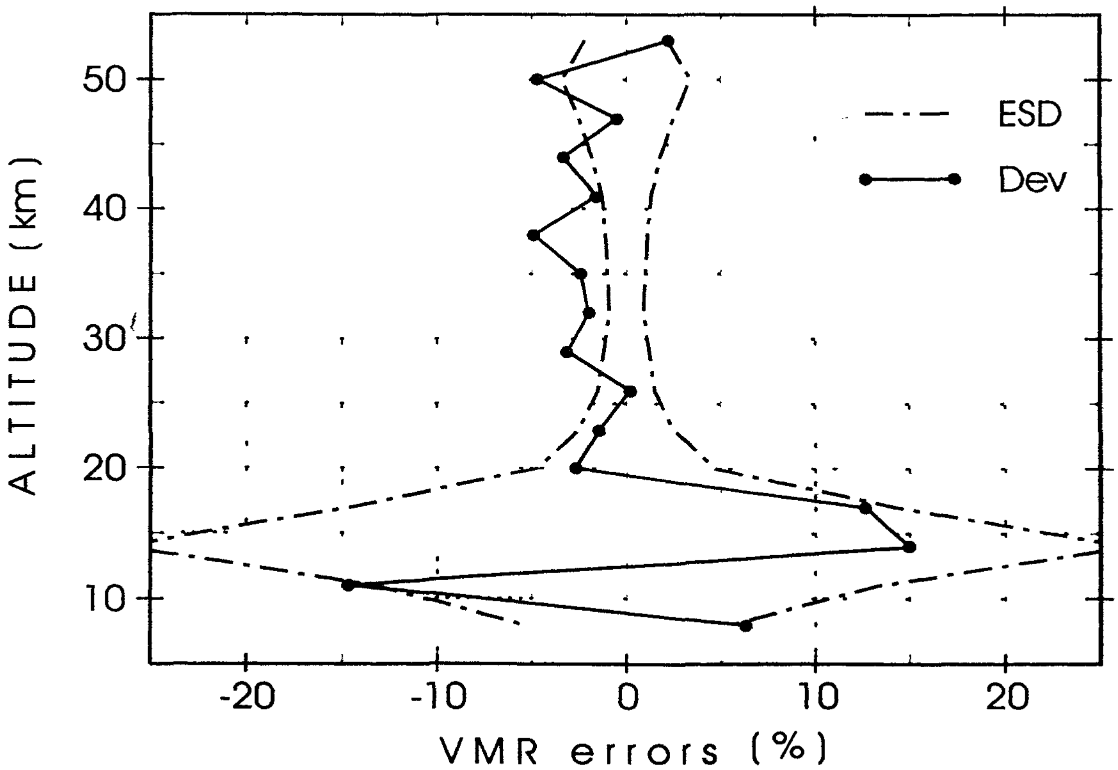
Geo-fit retrieval grid = ORM retrieval grid


GEOFIT analysis
ORM analysis
Altitude [km]
Orbital Coordinate [deg]
ESD [%]
15.0
13.5
12.0
10.5
9.0
7.5
6.0
4.5
3.0
1.5
0.0
Dev/ESD
5
4
3
2
1
0
0
90
180
270
360
10
20
30
40
50

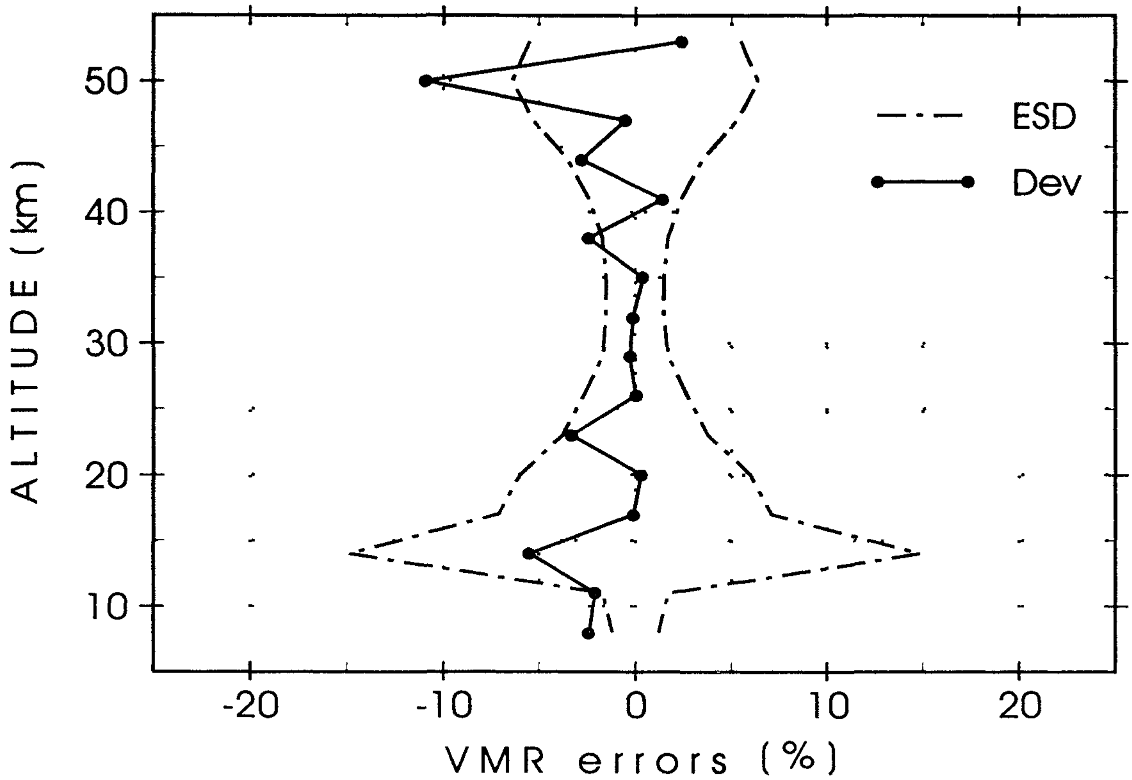
Sequence at equator: GEOFIT



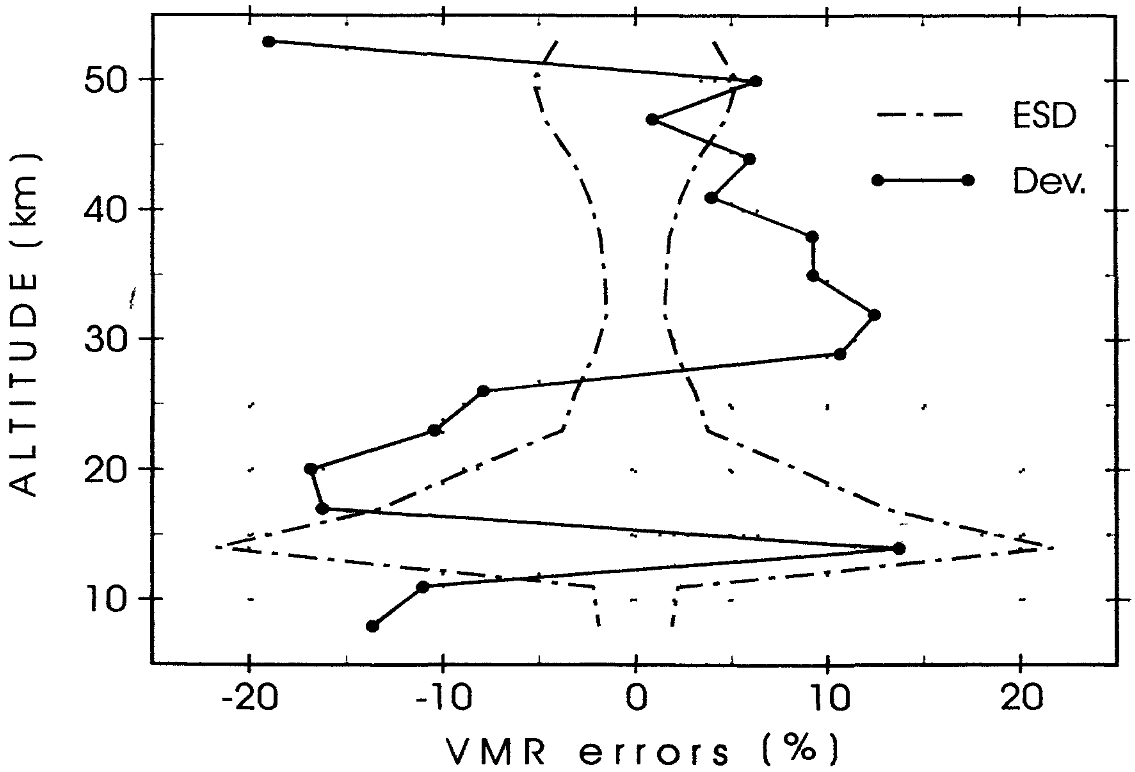
Sequence at equator: ORM



Sequence at polar vortex: GEOFIT

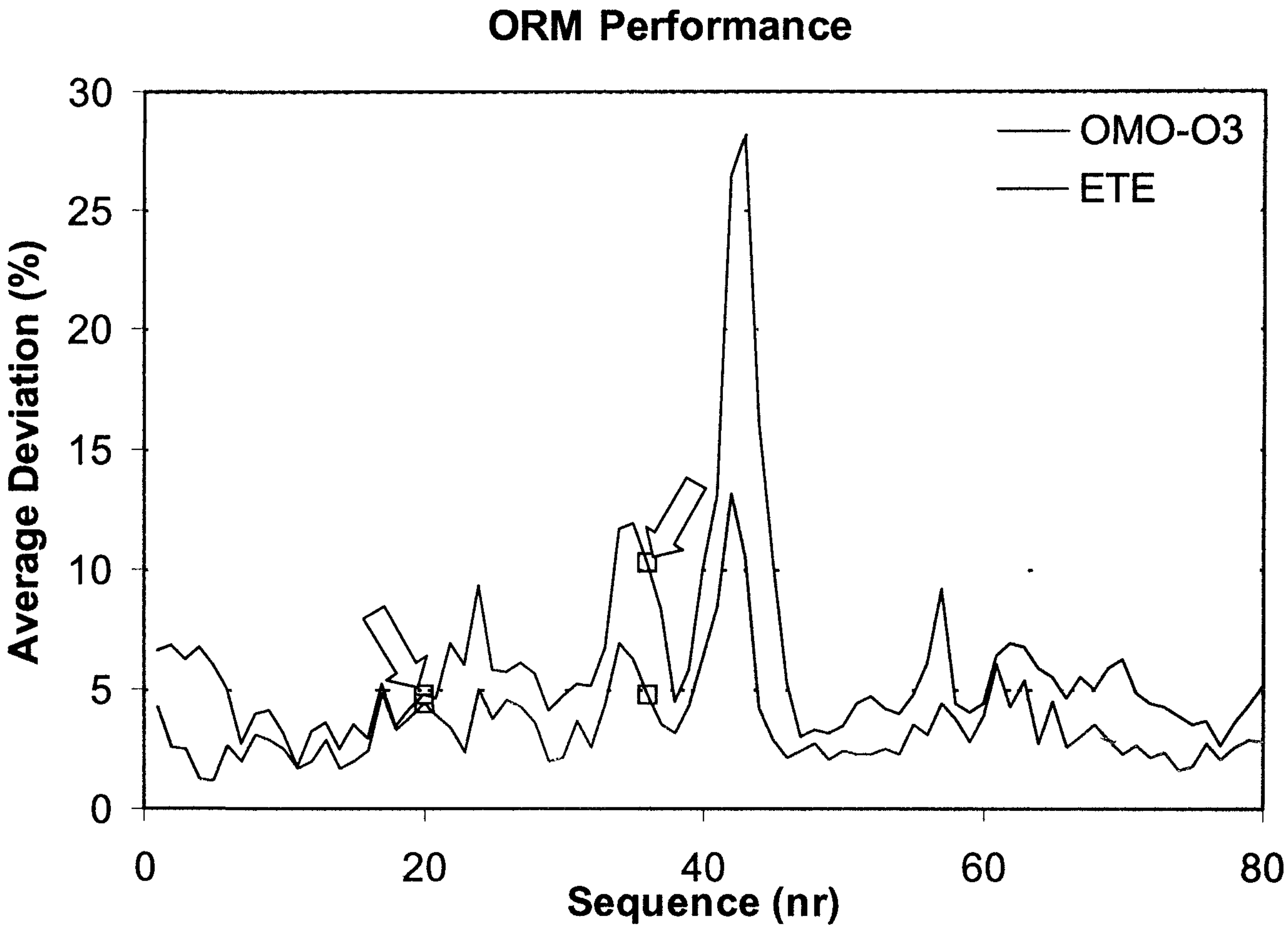


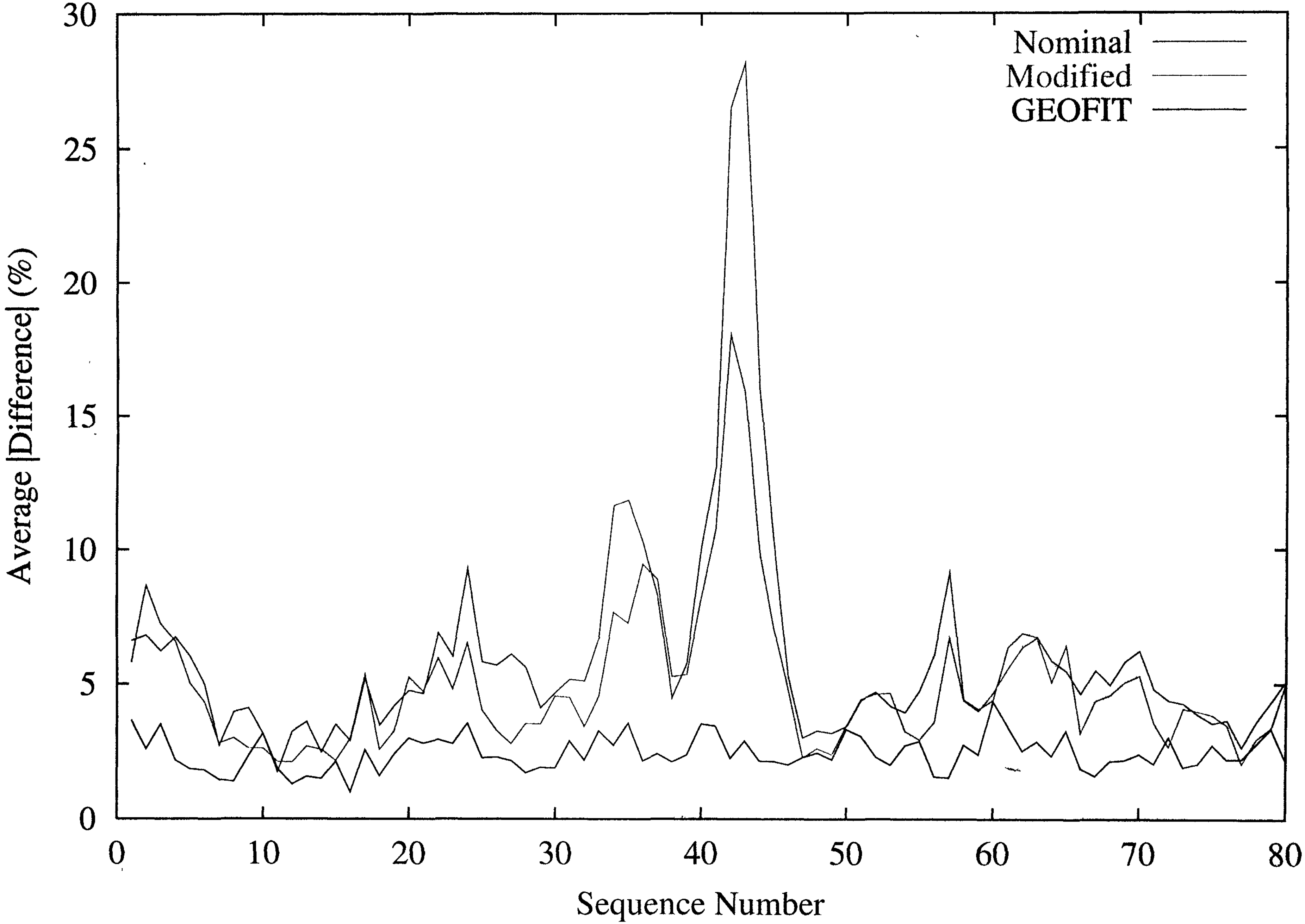
Sequence at polar vortex: ORM



Contribution of *p* and T

ORM retrievals have been repeated on observations generated using an atmosphere that is homogeneous with respect to pressure and temperature.




Nominal
Modified
GEOFIT
Average |Difference| (%)
0
5
10
15
20
25
30
0
10
20
30
40
50
60
70
80
Sequence Number

COMPUTATIONAL STRATEGY

80 limb-scanning sequences
949 spectral points for each sequence
75920 analyzed observations (m)

Atmospheric boundary at 80 km,
Atmosphere subdivided into 38 levels and 360 radii for a total of **13320** cloves

1280 VMR parameters
2240 atmospheric-continuum parameters (ımposed =0 above 30 km)
3520 retrieval parameters (n)

$$\mathbf{y}(n) \quad \mathbf{K}(m,n) \quad \mathbf{S_n}(m,m) \quad \mathbf{n}(m)$$

RAM requirement ≈ **700** Mb
disk space ≈ **3** Gb.
CPU (EV6 21264 of the "DEC-Alphaserver DS20) time ≈ **58** min
20% cross-sections, 30% radiative transfer and derivatives, 45% matrix algebra

REFERENCES

- The algorithms implemented in the ORM (identical to the ones implemented in the MIPAS NRT Level 2 processor of ESA) are described in:
  - Ridolfi M , B Carli, M Carlotti, T v Clarmann, B M Dinelli, A Dudhia, J -M Flaud, M Hopfner, P E Morris, P Raspollini, G Stiller, R J Wells, 'Optimized forward model and retrieval scheme for MIPAS near-real-time data processing' *Appl Optics*, Vol **39**, No 8, p 1323 – 1340 (10 March 2000)
  - HyperText Product Specifications (HTPS) will be made available by ESA, please refer to the ESA's web site *http //envisat estec esa nl* for questions on the HTPS
- New retrieval approach that overcomes horizontal homogeneity assumption:
  - M Carlotti, B M Dinelli, P,Raspollini, M Ridolfi, "Geo-fit approach to the analysis of satellite limb-scanning measurements', *Appl Optics*, Vol **40**, No 12, p 1872 - 1875 (April 2001)