TECHNICAL REPORT: THE MARS CLIMATE DATABASE IMPROVED LARGE SCALE VARIABILITY MODEL

ESTEC Contract 11369/95/NL/JG

M. Collins, S. R. Lewis and P. L. Read

January 20, 1998

Contents

1 Introduction

This technical note describes an improved version of the large scale variability model used in version 1.0 of the Mars Climate Database (MCD) and described in [1] and in the MCDv1.0 Detailed Design Document [2].

In the MCD, data is stored in 12 seasonal bins and at 12 local times of day within each season. Although this captures the main seasonal and diurnal components of variability, any intra-seasonal or day-to-day (synoptic) variations are averaged out. Thus there is a need to simulate this variability, especially if the user wishes to produce an ensemble of realizations of a variable at a particular seasonal date and local time of day which covers a realistic range of variability.

In version 1.0 of the MCD large scale variability is modelled in a novel manner by adding a series of functions to a mean vertical profile, $\overline{D}(z)$, i.e.,

$$D(z) \ = \overline{D}(z) \ + \ \sum_{i=1}^{N} \ p_i \mathbf{e}_i(z) \tag{1}$$

where the functions $\mathbf{e}_i(z)$ are eigenvectors of the covariance matrix of all the pre-averaged profiles generated by the GCM and p_i are the amplitudes of the functions. The eigenvectors, $\mathbf{e}_i$, are often called Empirical Orthogonal Functions (EOFs) and the p_i are referred to as the Principal Components (PCs) (see e.g. [3]). The set $\{\mathbf{e}_i\}$ form an optimal linear basis such that the variance capture is high even when the truncation limit is low.

In version 1.0 only correlations in altitude between variables were considered when calculating the covariance matrix. However, in order to retain cross-correlations between different variables (zonal wind, meridional wind, temperature, surface pressure and density) all were normalized and combined together to form a set of multivariate functions. Different sets of EOFs were computed for each of the 12 seasons on a low resolution grid (20° longitude × 20° latitude) and the series (1) was truncated at $N = 6$ to reduce the demands on data storage. Even so, typically $80 - 90\%$ of the variance was retained in the version 1.0

variability model at this level of truncation.

The amplitudes of the functions (PCs) were modelled using a superposition of a quadratic fit to the intra-seasonal (trend) component and a normally distributed random component for the remainder of the variance. This algorithm simulates realistic looking profiles (see figure 1). For further details see [1] and the MCDv1.0 Detailed Design Document [2].

2 Improving the Variability Model

2.1 Horizontal Correlations

Version 1.0 of the variability model only takes into account correlations in altitude between variables. In order to improve the model it is desirable to extend the spatial dimension to include correlations between variables in both the horizontal and the vertical.

Ultimately it would be desirable to include all the longitude, latitude and vertical grid-points in the analysis. A technical point, however, must be noted here. In computing the EOFs, the eigenvalues and eigenvectors of an $N \times N$ real symmetric matrix must be found. The order, N, of the matrix depends on the number of variables and on the number of spatial points. As the number of calculations needed to perform the eigenvector problem increases as N^3 there is a limit on the value of N that can be practically handled. The estimated CPU time and storage requirements for calculating the eigenvectors of the full problem (even on the low resolution grid with four three-dimensional and one two-dimensional variables, for which $N = 18 \times 9 \times (4 \times 25 + 1) = 16362$) is prohibitive.

We make the choice, therefore, to calculate EOFs in the two-dimensional, height-longitude plane which gives a manageable set of eigenvector problems. There is some physical basis for this choice, in that much of the variability the model must account for is in the form of baroclinic waves which, in general,

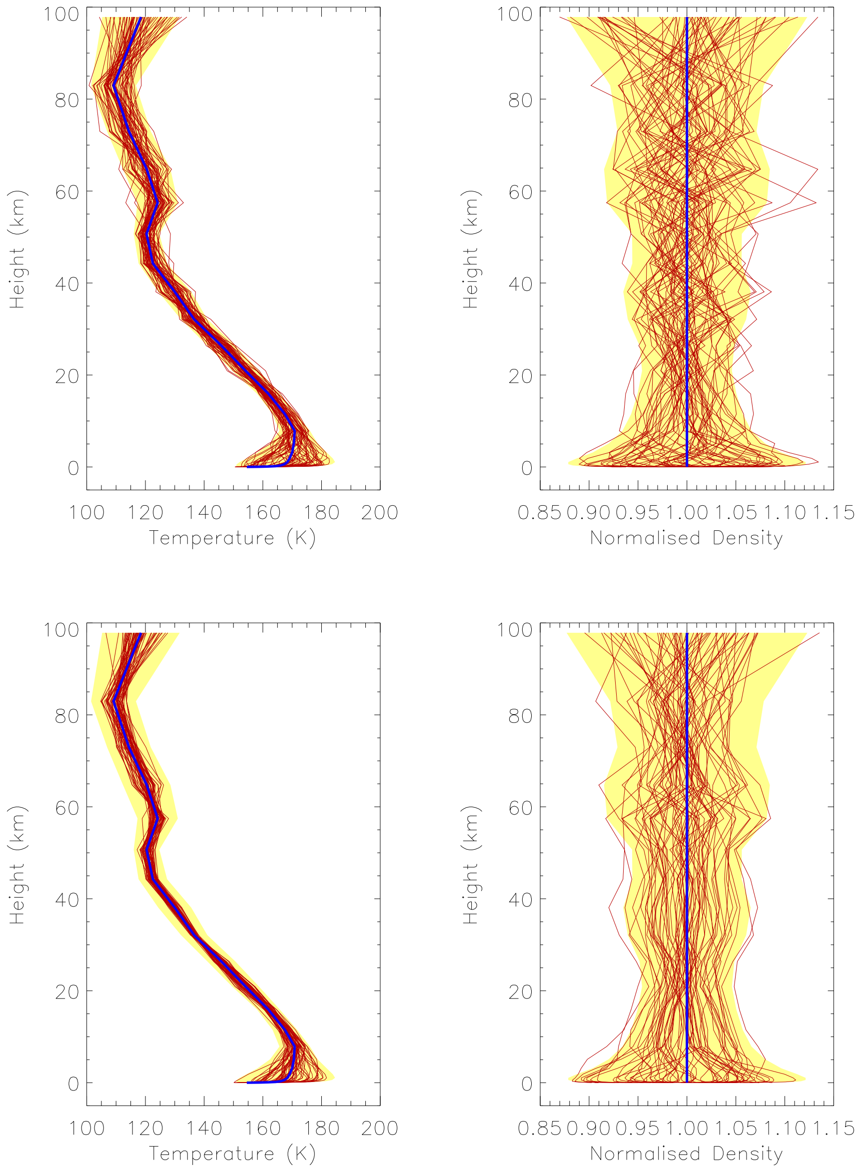

Figure 1: Top: Profiles of temperature (left panel) and normalized density (right panel) from a single horizontal point in the GCM. The mean profile is in black and the grey shading indicates 2 standard deviations from this mean. Bottom: Simulated profiles, corresponding to those shown above, computed using the variability model from version 1.0 of the MCD.

propagate west to east along lines of latitude.

2.2 Statistical Stability and PC Modelling

In version 1.0 of the MCD we calculated separate sets of the variability EOFs for each of the 12 seasons. However, due to the relatively small number of days in each season (50-70), this can lead to poor estimation of the EOFs. Greater statistical stability can be achieved by forming the covariance matrix over the entire annual cycle, although this means that more EOFs must be retained in the series (1) in order to still capture a relatively high fraction of the variance. Fortunately this is possible as there is then no need to store the EOFs at each season. In fact the series can be extended to include 72 ($= 12 \times 6$) EOFs rather than just 6 per season with no increase in the amount of space need to store the MCD.

This also overcomes a problem which occurred with the version 1.0 implementation of the PC modelling. Often if a time series of a variable was computed across a seasonal boundary there would be a discontinuity at the boundary where the quadratic fit to the intra-seasonal or trend components did not match. In the improved version 2.0 of the variability model we simply store a smooth version of each PC to represent the seasonal cycle together with a variance to represent synoptic activity. This both simplifies and improves the algorithm greatly.

2.3 Mathematical Details

Consider a time series of longitude-pressure vectors of zonal wind, $\mathbf{u}(\phi, p, t)$, meridional wind, $\mathbf{v}(\phi, p, t)$, temperature, $\mathbf{T}(\phi, p, t)$, density, $\rho(\phi, p, t)$ and surface pressure $p^*(\phi, t)$ at M discrete time points and on L spatial points. We form a time series of vectors $\mathbf{D}(t)$, where

$$\mathbf{D}(t) \;=\; (\hat{\mathbf{u}}(\phi, p, t), \hat{\mathbf{v}}(\phi, p, t), \hat{\mathbf{T}}(\phi, p, t), \hat{p^*}(\phi, t), \hat{\rho}(\phi, p, t)) \tag{2}$$

and the hat ˆ denotes a removal of the series mean and normalization of variance operator,

$$\overline{u} = \frac{1}{ML}\sum_{m=1}^{M}\sum_{l=1}^{L} u_{ml} \tag{3}$$

$$\hat{u}_{ml} = \frac{u_{ml} - \overline{u}}{\sqrt{\frac{1}{ML}\sum_{m=1}^{M}\sum_{l=1}^{L}(u_{ml} - \overline{u})^2}} \tag{4}$$

where the m denotes time and the l denotes spatial point. Hence, with this normalization, the variance of the entire time series of $\mathbf{D}(t)$ is unity.

We then form the $(N \times N)$ covariance matrix, $\mathbf{C}$, such that

$$\mathbf{C} = \frac{1}{N}\mathbf{D}\mathbf{D}^T \tag{5}$$

where $N =$ number of horizontal points $\times(4\times$ number of vertical points$+1)$, $\mathbf{D}$ is the $(N \times M)$ matrix whose rows are the vectors $\mathbf{D}(t)$ and the superscript T indicates the transpose.

The matrix $\mathbf{C}$ is real symmetric and we can find the eigenvectors and eigenvalues and order them in decreasing eigenvalue magnitude. We note that if $\mathbf{E}_i$ is the i'th eigenvector then

$$|\mathbf{E}_i| = 1 \tag{6}$$

where $|\cdot|$ is the Euclidean Norm, and

$$\sum_{i=1}^{N}\lambda_i = 1 \tag{7}$$

where λ_i is the i'th eigenvalue.

The i'th principal component (PC) at time m, p_{mi} is defined as

$$p_{mi} = \sum_{n=1}^{N} D_{mn}E_{in} = \mathbf{D}\cdot\mathbf{E}_i \tag{8}$$

and we note the result

$$\frac{1}{MN}\sum_{m=1}^{M}(p_{mi})^2 = \lambda_i \tag{9}$$

Each principal component has 669 values during one year (one per day). From the PCs we calculate a 30 day running mean, ps_{mi}, and a variance departure from that mean, pv_{mi}. The modelled PC, p'_{mi}, at time m is then

$$p'_{mi} \; = \; ps_{mi} + \Phi pv_{mi} \tag{10}$$

where Φ is a normally distributed random variable with unit standard deviation.

3 Examples

In this section the new variability model is illustrated with some examples. The longitude-height EOFs have been calculated from one year of a Viking scenario run (see e.g. MCDv1.0 Detailed Design Document) using the AOPP version of the GCM with data saved once a day on a coarse 20° longitude × 20° latitude grid.

3.1 Variance Capture

In version 1.0 of the variability model only 6 EOFs per location were retained in the series (1) for each of the 12 seasons to limit the amount of data storage required. Because, in the improved version, a single set of EOFs are calculated for the whole year, 72 EOFs can be retained in the series with no greater storage overhead, to give a potentially greater overall variance capture. Figure 2 shows the fraction of variance captured by retaining 72, 36, 18 and 9 EOFs at different latitudes. By retaining 36 EOFs approximately 90% or greater of the variance is captured and by retaining 72 EOFs approximately 95% or greater of the variance is captured. Including more EOFs gives relatively little increase in variance capture but increases the storage requirements considerably. Thus 72 EOFs seems to be a good compromise between variance capture and data storage.

3.2 Density Cross-Sections

Density is one of the most important variables used by engineers in mission planning (e.g. when calculating entry trajectories or aerobraking manoeuvres). Hence

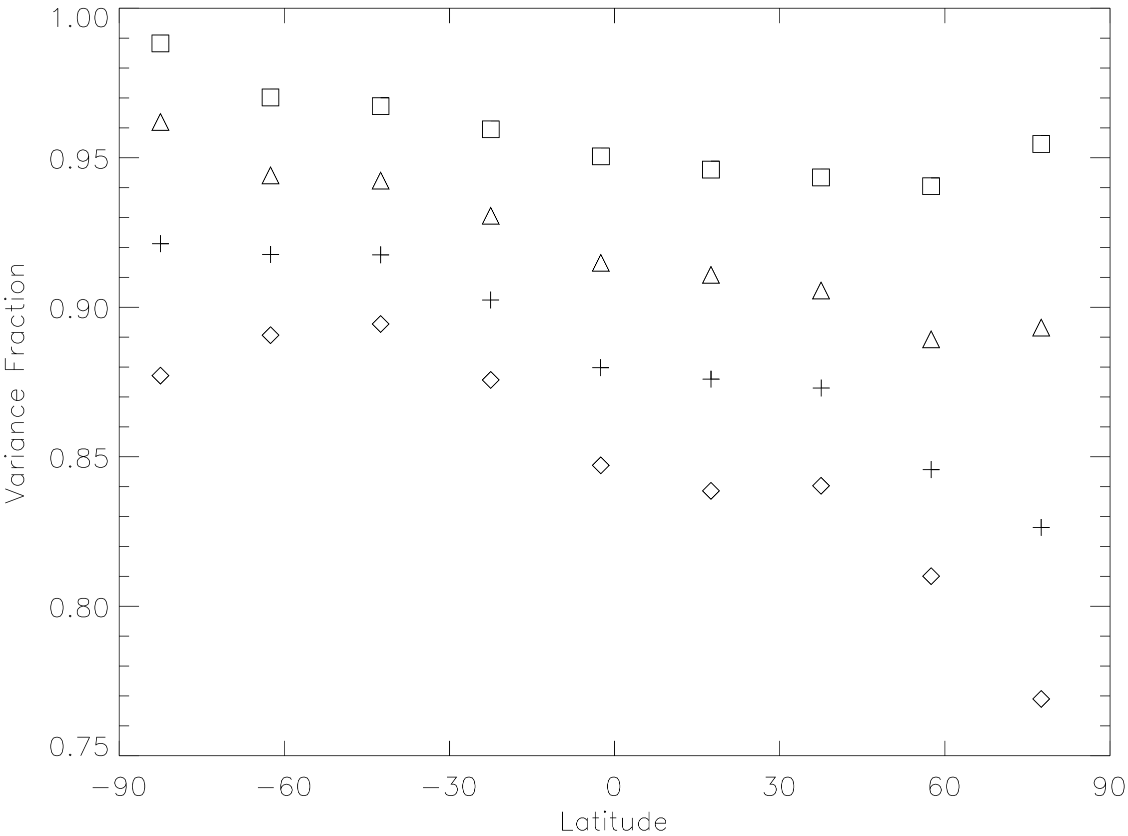

Figure 2: The fraction of variance captured by 72 EOFs (squares), 36 EOFs (triangles), 18 EOFs (plus signs) and 9 EOFs (diamonds) at different latitudes for the Viking scenario run. 72 EOFs capture above 95% of the variance at most latitudes.

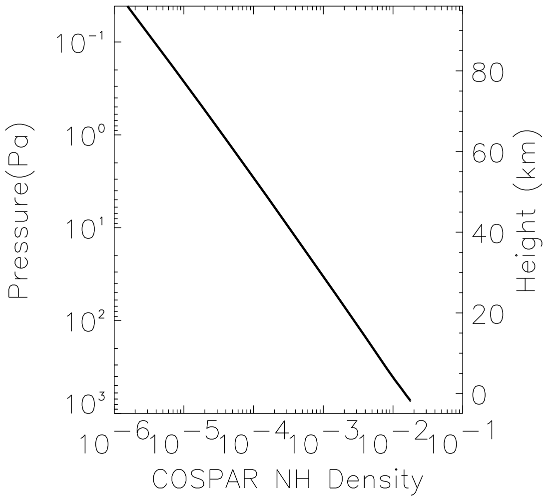

Figure 3: COSPAR Northern Hemisphere mean density for Mars. Note the logarithmic scale on the abscissa.

the new variability model is illustrated using this field. Because density varies exponentially with height it is useful to compare any density profile with some 'standard' profile. Figure 3 shows the northern hemisphere COSPAR standard atmospheric density profile from the surface to approximately 90 km. All density profiles are referenced to this profile by expressing the density as a percentage difference, d, from the COSPAR profile. i.e.

$$d = 100 * \frac{\rho_{COSPAR} - \rho}{\rho_{COSPAR}} \tag{11}$$

Figures 4 and 6 show a selection of instantaneous longitude-height fields of density (referenced to the COSPAR profiles) taken from the Viking scenario run at $L_s = 0°$, $L_s = 90°$, $L_s = 180°$ and $L_s = 270°$. The variability model simulations of these fields are shown in figures 5 and 7. On a large scale the variability model simulates the large seasonal cycle of density well. Small differences between the GCM and the variability model simulated fields are due to the random nature of the PC model (equation 10). Such synoptic variability is illustrated in figure 8 which shows several density profiles taken from the GCM next to several simulated profiles using the new variability model. The general large scale variability is captured well, though, in the first case, the simulated profiles appear to show less small scale variance (especially above 30km) than the actual profiles. Further work will be conducted to investigate whether this is significant in general.

4 Summary

This document describes an updated version of the large scale variability model used in the MCD. The model adds a series of functions (EOFs), statistically derived from the GCM simulations, to seasonally averaged fields taken from the MCD. The new variability model (version 2.0) improves on the original model (version 1.0) by including longitudinal correlations of variables (two-dimensional rather than one-dimensional EOFs) and by increasing the statistical stability of the EOFs via increased sampling. These improvements are achieved without requiring additional storage space for the EOF data and with a negligible increase in processor time for the end user. The only significant additional computation required to calculate the two-dimensional EOFs occurs when the database is initially compiled. The resulting model gives a convincing simulation of large scale variability.

References

[1] M. Collins, S. R. Lewis, P. L. Read, N. P. J. Thomas, O. Talagrand, F. Forget, R. Fournier, F. Hourdin, and J-P Huot.
A climate database for the martian atmosphere.
ESA Special Publication, SP-392:323–327, 1996.

[2] M. Collins, S. R. Lewis, and P. L. Read.

Mars Climate Database v1.0: Detailed design document.
ESA Technical Report, 1997.

[3] K.C. Mo and M. Ghil.
Statistics and dynamics of persistent anomalies.
J. Atmos. Sci., 55:877–901, 1987.

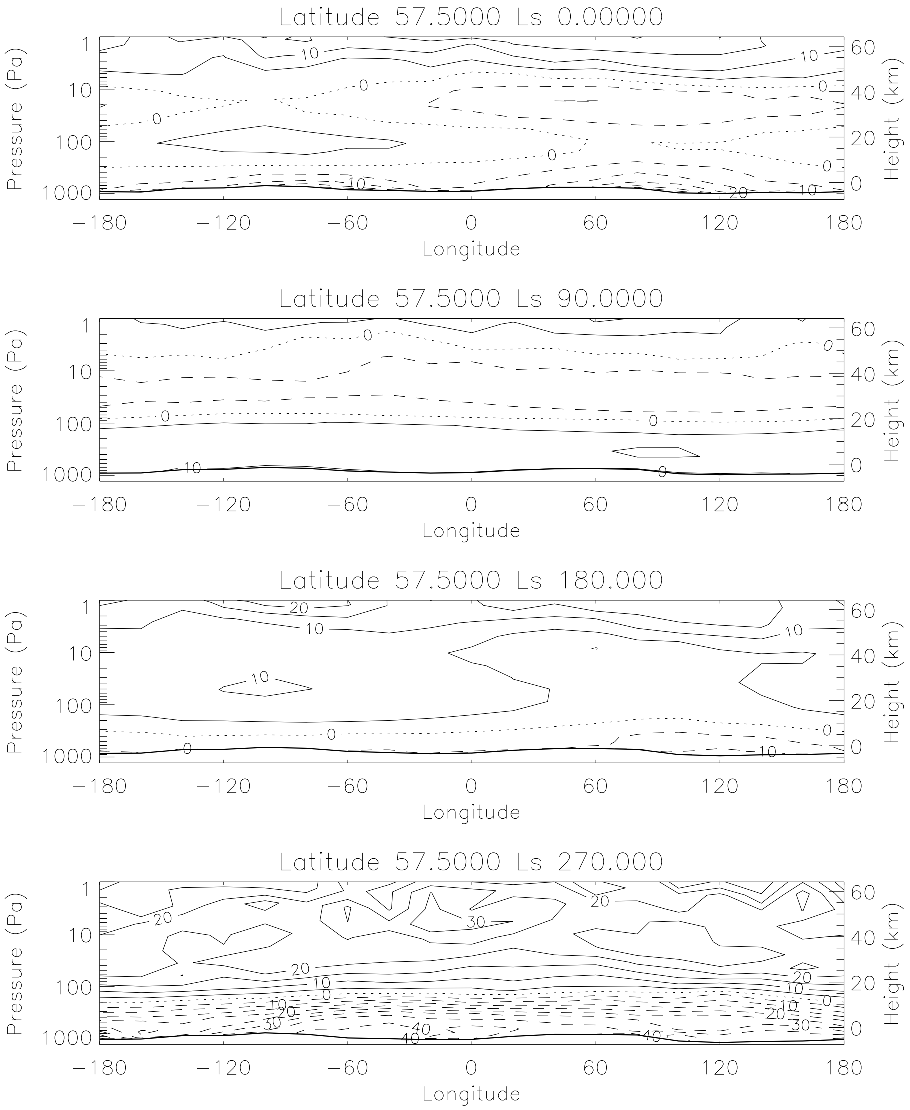

Figure 4: Longitude-height cross sections of density at 57.5°N expressed as a deviation from the COSPAR northern hemisphere reference profile at several seasonal dates. The fields are instantaneous and have been taken from the Viking scenario model run.

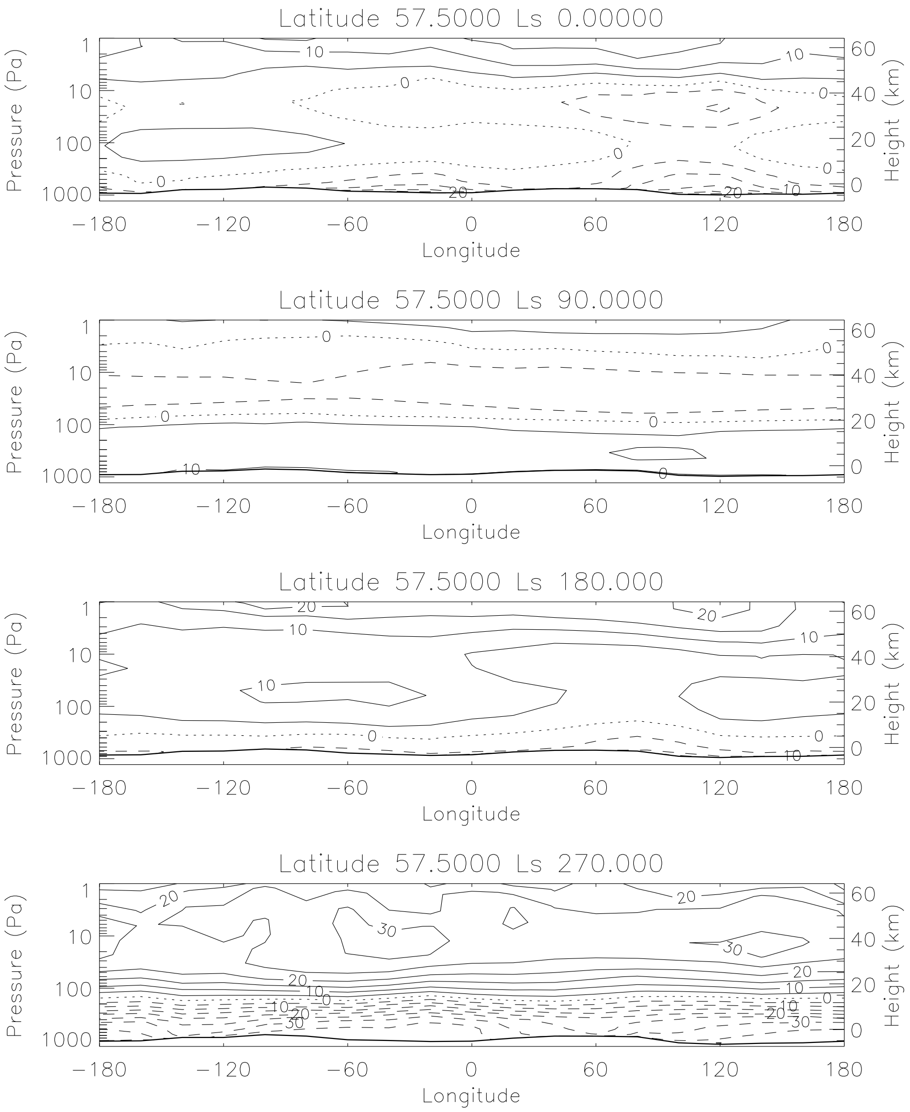

Figure 5: Longitude-height cross sections of density at 57.5°N simulated using the new variability model and expressed as a deviation from the COSPAR northern hemisphere reference profile.

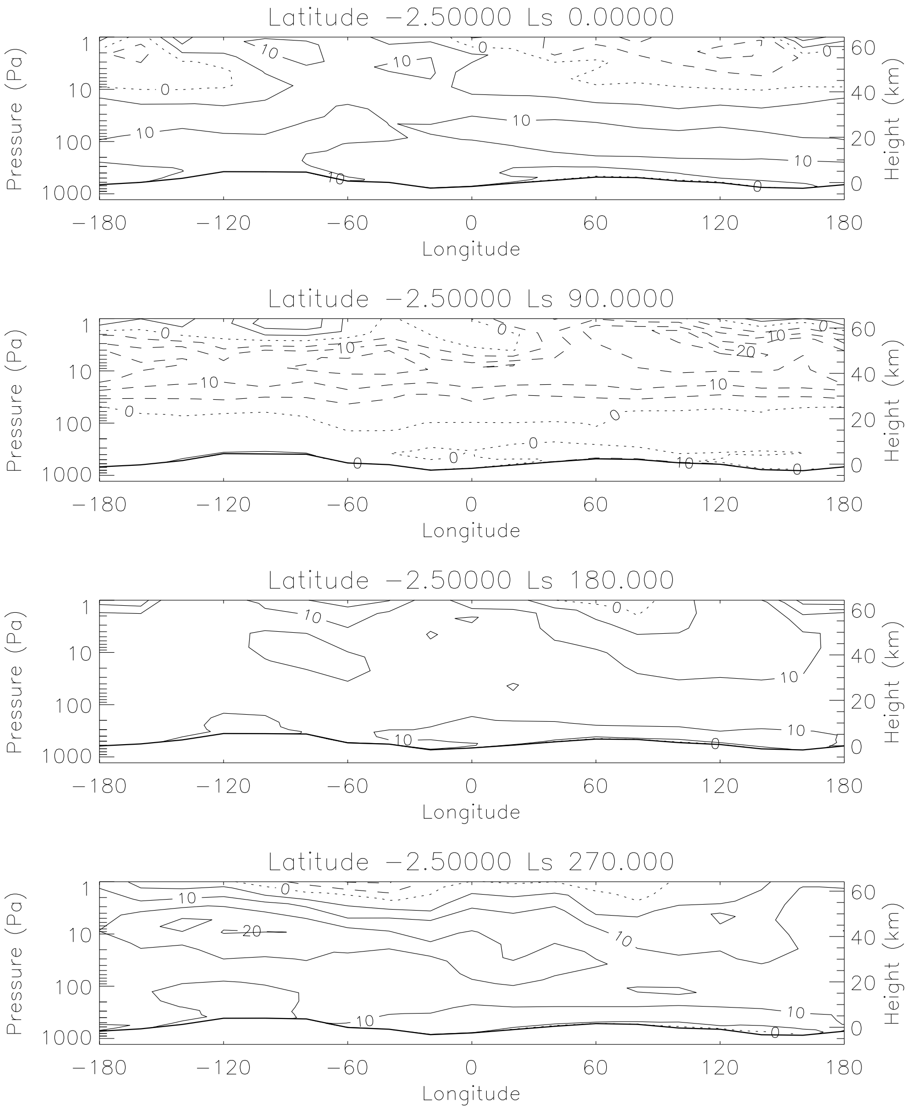

Figure 6: Longitude-height cross sections of density at 2.5°S expressed as a deviation from the COSPAR northern hemisphere reference profile at several seasonal dates. The fields are instantaneous and have been taken from the Viking scenario model run.

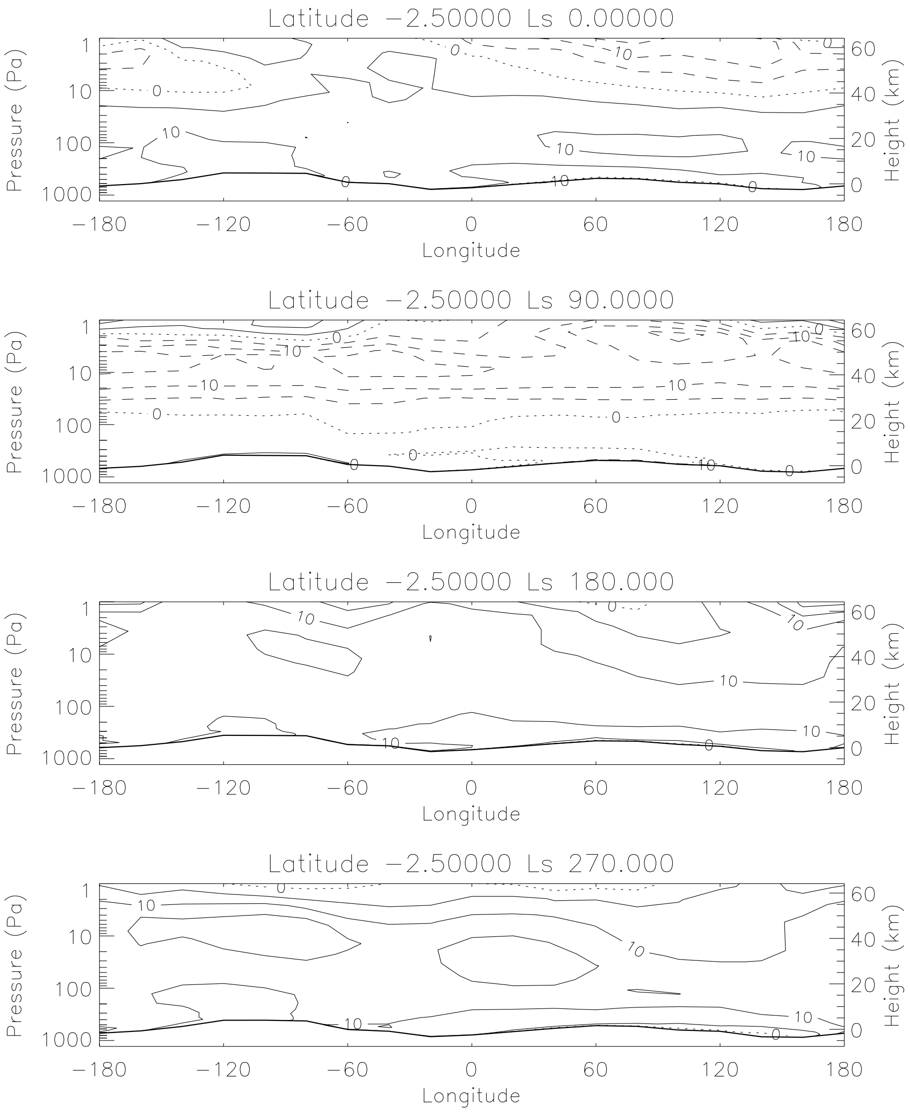

Figure 7: Longitude-height cross sections of density at 2.5°S simulated using the new variability model and expressed as a deviation from the COSPAR northern hemisphere reference profile.

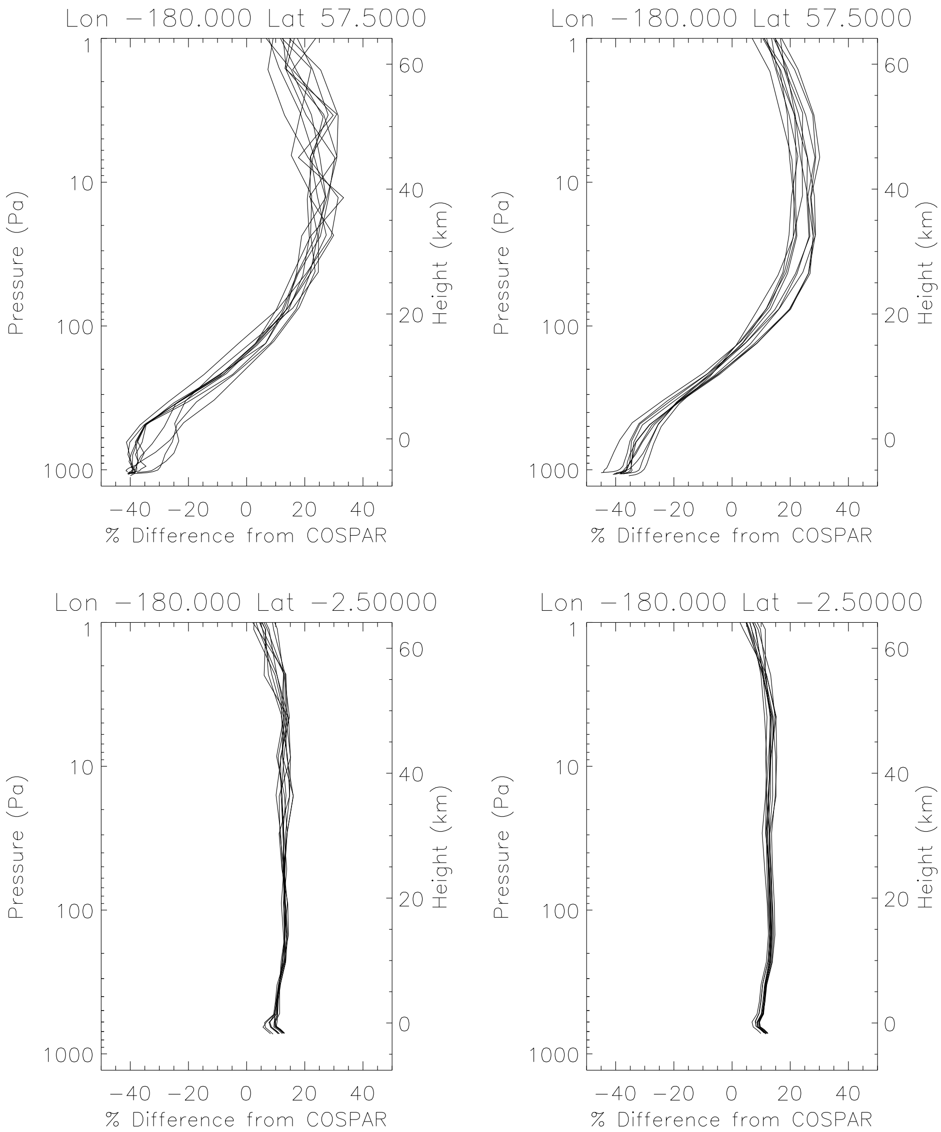

Figure 8: Left: 10 profiles of density (expressed as a deviation from the COSPAR profile) taken from $L_s = 270°$ at two points of the GCM simulation. Right: Simulations of these profiles using the new variability model.