Martian Atmospheric Data Assimilation with a Simplified General Circulation Model: Orbiter and Lander Networks.

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

University of Oxford, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, U.K.

Address for correspondence:

Dr. S.R. Lewis,

Atmospheric, Oceanic and Planetary Physics,

Clarendon Laboratory,

Parks Road,

Oxford OX1 3PU,

U.K.

(e-mail: s.lewis1@physics.oxford.ac.uk)

Short title: MARTIAN ATMOSPHERIC DATA ASSIMILATION

Abstract. A meteorological data assimilation scheme for the martian atmosphere has been implemented and tested, based on techniques used in the current operational scheme for weather forecasting at the UK Meteorological Office. The scheme has been interfaced with a range of simple models and with the martian GCM currently under simultaneous development at Laboratoire de Météorologie Dynamique du CNRS in Paris and at Oxford. As well as the interpretation of data from any future spacecraft, the assimilation scheme may be used for comparisons between different models, for model validation using earlier martian data, and for data impact studies to assist in planning new missions. Despite proposed new missions to Mars, observations of the atmosphere of Mars in the near future are still likely to be very sparse compared to those of the Earth (perhaps comprising a single orbiter and a few surface stations at any one time) and the scheme has been adapted with this in mind. In this paper, twin model experiments are conducted in which simulated observations are generated from a second model started from different initial conditions. Such experiments reveal the importance of surface pressure measurements (in combination with an accurate topographic map, such as will be available from laser altimetry) in the determination of the amplitude of large-scale atmospheric waves. It is shown that atmospheric temperature profiles from a remote-sensing instrument on a polar orbiting satellite combined with simultaneous surface pressure observations at a limited number of sites, as planned for the INTERMARSNET mission, is a useful scenario for data assimilation.

Introduction

Data assimilation is a useful technique for the analysis of atmospheric observations (Daley, 1991), especially where a map of synoptic-scale features is required from data taken asynchronously. A straightforward gridding analysis of observations from a single orbiting satellite, for example, can lead to ambiguities in the interpretation of observations when studying synoptic-scale phenomena which change significantly over the course of a few days. Assimilation may also be valuable where it is desirable to produce a full balanced analysis of all variables (including those not directly observed) for a global model which is physically consistent with a set of sparse measurements such as may be made from a several landers on the martian surface. Assimilation makes use of prior knowledge of atmospheric physics, embodied in a general circulation model, as well as information from previously assimilated observations summarized as an evolving model state. By examining misfits between model forecasts and observations it may also be possible to identify deficiencies in the model formulation. Data assimilation is performed operationally using observations of the Earth's atmosphere, often in the context of providing an initial state for a weather forecast (Ghil *et al.*, 1981; Lorenc, 1986). With the new atmospheric data expected from forthcoming spacecraft missions to Mars it is now possible for the first time to propose the use of this technique for another planet (e.g., Banfield *et al.*, 1995; Lewis and Read, 1995).

This paper describes assimilation experiments conducted with a relatively simple general circulation model of the atmosphere of Mars. The model solves the nonlinear, adiabatic primitive equations of meteorology on a sphere in full, but simplifies physical diabatic and frictional processes to linear forms for speed and simplicity. Artificial observations are generated from one run of the model (henceforth known as the truth run) and assimilated into a second model (the record of model states from the assimilation run is henceforth known as the analysis), which may be started from different initial conditions and have different forcing and dissipative terms. The record of states from the truth run may be compared with the analysis to assess how well the assimilation process performs. In order to avoid the problem of identical models producing overly optimistic analyses, in these experiments the assimilation model, unlike the true model, uses no thermal forcing and frictional boundary layer terms. Although the physical schemes in use in the true model are simple, the test for the assimilation scheme to work with a model where these schemes are absent is quite extreme. It is important to note that the assimilation of model-generated data may be more accurate than the assimilation of real atmospheric observations into a numerical model. A numerical model will not necessarily provide a good, unbiased representation of even the large-scale processes in such a complex system as a real atmosphere and real observations can suffer from poorly-known systematic errors. However, this test process does provide a good method to compare observing strategies, such as different scan patterns for remote sounding instruments onboard orbiting spacecraft or different numbers and locations of landers. Experiments in which data is generated from, and assimilated into, models with much more realistic representations of physical processes will be described in following papers.

The following section describes the numerical model and assimilation scheme in more detail and describes changes made to the assimilation scheme for the martian assimilations. The type of artificial data generated from independent model experiments and how its resolution and coverage relate to likely future spacecraft missions is next described. The data consists of temperature profiles, as might be obtained from a remote sounding instrument, and surface pressure measurements at a limited number of discrete sites, as might be chosen for INTERMARSNET and other future missions to the martian surface. Results are shown from two types of experiments. The first set of experiments focus upon the reconstruction of propagating planetary waves; in order to isolate the planetary waves from other phenomena, such as stationary waves, the model is run in an idealized form without topography. The second set of experiments are more realistic with greater vertical coverage, up to 80 km altitude, and with martian surface topography. They focus upon the analysis of the full global circulation. Both sets of experiments are designed to test whether assimilation with temperature profiles from a single polar orbiting instrument is feasible and whether the analysis so obtained can be improved by the addition of a limited amount of surface pressure information. Finally, the results are summarized and some future plans for the use of the scheme are outlined.

Numerical methods

Atmospheric model

The martian general circulation model used for these experiments is based on a semi-spectral primitive equations model (Hoskins and Simmons, 1975). The adiabatic primitive equations of meteorology are solved in terms of spherical harmonic expansions in the horizontal and σ levels in the vertical (where $\sigma = p(x,y,z,t)/p_s(x,y,t)$ is the pressure divided by the surface pressure, so that the sigma coordinate follows the shape of the underlying topography without cutting the surface). The horizontal resolution is determined by the triangular spectral truncation at a total wavenumber of 21 for these trials; this is roughly equivalent to a grid separated by 5.6° in latitude and longitude for calculating the nonlinear terms in the primitive equations. In the vertical, two different resolutions are used: ten equally-spaced σ levels which range from about 500 m to 30 km in altitude, with the first nine levels lying below 20 km, for the planetary wave experiments without topography (similar to the experiments shown in Lewis and Read, 1995); and a stretched grid of twenty levels which range from 80 m to 90 km, with the first five levels lying within 10 km of the surface and the spacing between levels tending to 5 km with increasing altitude, for the global circulation experiments with topography and temperature profiles covering the region 0–80 km altitude.

The model is forced with simple, linear relaxation towards a radiative equilibrium state (known as Newtonian cooling) acting with a timescale of 2 sols (a sol is a martian solar day or about $24^{\mathrm{hr}}\ 40^{\mathrm{min}}$) and by a drag proportional to the velocity (known as Rayleigh friction) acting with a timescale of 3 sols in the lowest model level to simulate the boundary layer in a very simple way. The model has been used in this form before for several studies both without topography (Collins and James, 1995) and with topography (Joshi *et al.*, 1994, 1995). Even with such simple parameterizations of physical processes, the model is still able to produce realistic zonal mean states and regimes of both regular and irregular baroclinic waves in the winter hemisphere (see Barnes, 1981, for details of observations of martian baroclinic waves).

Analysis correction scheme

Data assimilation is conducted using a modified form of the analysis correction scheme (Lorenc *et al.*, 1991), which is in current operational use at the United Kingdom Meteorological Office. The analysis correction scheme has been found to produce improved analyses in data-sparse areas compared to the previous optimal interpolation scheme in use at the Meteorological Office, primarily because observational increments are spread over larger areas (Lorenc *et al.*, 1991), and thus was expected to be more suitable for the martian problem. It is also straightforward to modify and computationally inexpensive, which were additional important requirements in the development of an operational scheme which could be run in conjunction with a spacecraft mission in the near future. The scheme is compared with some alternative proposals in the concluding section.

The analysis correction scheme involves a form of successive correction iterated between each time step of the model which attempts to minimize, in a least-squares sense, the misfit between the model and the observations weighted by their errors. Continuous (asynoptic) repeated insertion is conducted sequentially, assimilating each measurement at the time and place it is valid without prior interpolation. Observational increments to the model are spread in time and space by empirically-determined functions with scales which depend on the difference between the current model time and the time at which the data is valid. More details are given in the next section. As an observation's valid time approaches, its weight increases and the scale which it affects becomes smaller. An observation is discarded about one hour after the model has passed its valid time. The weight of each observation is scaled by the relative magnitudes of observational and forecast model error (Bell and Dickinson, 1987). In practice what is important for the martian case with only a limited number of different observations is the relative error assigned to each observation. For these tests each observation was weighted equally, although it might be more appropriate in practice to assign observational errors as a simple function of altitude where less weight is given near the top and bottom of vertical profiles where retrievals are likely to be less accurate.

Within each cycle of the assimilation scheme each model variable is analysed sequentially (surface pressure, temperature, velocity and constituents, e.g. dust and/or water vapour) followed by the derivation of multi-variate increment fields for dynamical balance. In the case of the assimilations performed here, any surface pressure information is analysed first. This is balanced with potential temperature increments which constrain, through the hydrostatic relation, the geopotential height increment at each model level to fall off with increasing altitude above the surface according to a specified correlation function ($\sigma(1+\sigma)/2$, see Lorenc *et al.*, 1991). Next, wind increments in geostrophic and thermal wind balance with the changes to the geopotential height and potential temperature are calculated, multiplied by a weighting which falls to zero in equatorial regions where these approximations are poor. Following the surface pressure analysis, the temperature observations are analysed and balanced by further thermal wind increments in the same way. This balancing procedure is conducted in order to avoid losing much of the newly introduced information to fast inertia-gravity waves (see Daley, 1991, for further discussion of the problem). In addition to the slow, synoptic-scale, meteorologically significant features, such as planetary waves, the primitive equations on which the model is based also support fast inertia-gravity waves (e.g. Gill, 1982) which carry a small, but important, fraction of the total energy. There have been attempts to demonstrate the existence of a slow manifold within the primitive equations system, which describes the evolution of the large-scale, quasi-geostrophic features alone (Leith, 1980; Lorenz, 1980). Though the slow manifold does not seem to exist in a rigorous sense and inertia-gravity waves are an inseparable part of the system (Vautard and Legras, 1986; Lacarra and Talagrand, 1988), much atmospheric energy is contained within the slower modes and atmospheric data assimilation schemes often employ normal mode initialization (Daley, 1991) to reduce the spuriously large amount of inertia-gravity waves that can be generated in a model when data is inserted in a sub-optimal manner. In contrast to a cycle of normal mode initialization followed by a forecast to the next synoptic hour, the analysis correction scheme is continuous. Therefore, increments to one observed variable are balanced by geostrophic estimates to the others at each step. The iteration of the scheme between each model timestep, with each observation being introduced about 50 times with varying weight over the six hour window in which it is valid, means that a better approximation to the full primitive equations balance can be built up. The spreading of observations in time also leads to the slower, meteorological modes being favoured.

For data assimilation the model variables are transformed from spectral components of vorticity, ζ, divergence, δ, temperature, T and log(surface pressure), $\ln p_s$, to eastward and northward velocity, (u, v), potential temperature, θ and surface pressure, p_s, on a regular, staggered grid which has the same vertical form as the model grid and is of sufficient horizontal size to recover all the information in the spectral field by an inverse transformation. The assimilation analysis is then performed and the model fields are returned to spectral form. One transform and analysis step is interleaved with every model time step, as is done for the Earth-based scheme (Lorenc *et al.*, 1991), since this introduces small increments frequently and allows the model to establish a balance between the variables. There is no reason in principle why an analysis consisting of several iterations of the scheme could not be performed in between short forecasts of several model timesteps, in a similar manner to an optimal interpolation scheme operating at synoptic hours (Lorenc *et al.*, 1991). This would avoid the necessity of performing the spectral transforms so often, but in practice the transforms take a small amount of computational time compared to the model itself and the continuous approach results in better assimilations when most of the data is asynoptic, as is the case here.

The version of the model with data assimilation is run as a simple adiabatic model driven solely by analysis increments, i.e. Newtonian cooling and Rayleigh friction are turned off to simulate a complete lack of knowledge about the physics and to prevent a falsely accurate analysis. This provides a severe test of the assimilation process, since almost any reasonable assumptions made about the radiative heating and boundary layer processes are likely to be better than setting the heating and friction to zero. With real observations it is desirable to drive the model with the best possible physical parameterizations in order to obtain the optimum analysis of the data.

Changes to the analysis correction scheme

This section summarizes the main differences between the analysis correction scheme as used for martian assimilations compared to its operational implementation in the global version used for the Earth. First the specific numerical values for some of the most important parameters are described in the context of the equations in the scheme in which they appear. Then the overall method by which these values were arrived at for Mars is outlined. A more complete account of the scheme is given by Lorenc *et al.*(1991) and the notation used here follows that paper.

The analysis correction scheme splits the analysis of increments into a vertical step (at the observation location) followed by a horizontal and temporal analysis. For temperature profiles, the vertical analysis is conducted by presenting the temperatures as mean values between a standard set of pressure boundaries, chosen to reflect the supposed vertical resolution of the data. The total model layer thickness calculated between these boundaries is adjusted to be consistent with the temperature observations, so that the vertical scale of the temperature increments depends on the observational resolution rather than the model level spacing. Thickness increments to the individual model layers may then be further smoothed in the vertical by a triangular filter.

The horizontal analysis is conducted by calculating increments to the model variables on the model grid, Δx_k, from increments at the observation locations, C_i, as follows,

$$\Delta x_k = \lambda \sum_i \mu_{ki} \tilde{Q}_i R_i^2(\delta t_i) C_i \tag{1}$$

where the index k refers to a model gridpoint and the index i to an observation. The normalization factor, $\tilde{Q}_i$, is a function of the ratio of observational to first guess error and of the local observation density around each observation location (details are given in Lorenc *et al.*, 1991).

Observational increments are spread in time by the function $R_i(\delta t_i)$, where δt_i is the time difference between the current model and observation time. The function first acquires observations five hours before they are valid and finally releases them one hour after their valid time. The function used to do this increases linearly with time from zero at $\delta t_i = -5$ hr to unity at $\delta t_i = 0$ hr and falls again linearly to zero at $\delta t_i = +1$ hr. This function shape and time scale was preserved from the scheme in use for the Earth (Lorenc *et al.*, 1991); the time scale represents a similar fraction of the planetary rotation period in each case. The asymmetry in time of the function is also ideal for the problem of the assimilation of observations from a single polar orbiting satellite since it biases the assimilation gains towards the direction of motion of the satellite ground track, which is the region of the atmosphere which has not been observed recently, while dropping observations from previous orbits which have been assimilated in the recent past. This point is further discussed in the conclusions.

In the horizontal, the observational increments are spread by a second-order autoregressive function of the distance, r_{ki}, between the observation and each model gridpoint,

$$\mu_{ki} = (1 + r_{ki}/S_i(\delta t_i)) \exp(-r_{ki}/S_i(\delta t_i)) \tag{2}$$

where the correlation scale of the function, $S_i(\delta t_i)$, is a function of δt_i which is large at $\delta t_i = -5$ hr, falls linearly to a minimum at $\delta t_i = 0$ hr and increases again until the observations are dropped at $\delta t_i = +1$ hr. Thus increments from observations with a low weight at the start of their insertion period are given a larger scale and the scale falls as the weight increases. This choice of function has some advantages over a Gaussian curve in that it has a longer tail which is of benefit in areas where data is sparse. For practical reasons the function is cut-off to zero at 3.5 correlation scales. The best results were obtained with a scale, S, which is similar to that used for the southern hemisphere on Earth; 540 km at the start of the insertion period, decreasing to 340 km at the valid time and increasing to 375 km at the end of the insertion period for the martian assimilations (compared to 360 km, 225 km and 250 km at the same times in the Earth's northern hemisphere and 601 km, 376 km and 418 km in the Earth's southern hemisphere). This correlation scale means that for a single polar orbiter, such as Mars Observer, which was planned to have a spacing of 28° latitude (about 1660 km at the equator) between adjacent ground tracks, the observational increments from each successive orbit should overlap in their tails.

The total increments from the horizontal and temporal analysis at one timestep are illustrated in Fig. 1, using the functions and parameters outlined above. For this illustration, a Mars Observer polar orbit has been assumed (12 2/3 orbits per sol) and profiles have been taken 90 times per orbit. It was assumed that the vertical analyses have already been conducted and that at this level in the atmosphere the results imply increments of +1 K at each profile location, all with equal errors. All the profiles which fall within the current time window and contribute to the analysis are marked. The shading shows the accumulated increments on the model grid, prior to weighting by the relaxation coefficient ($\lambda \sim 0.2$), i.e. the contents of the sum on the right hand side of Eqn. 1 for each gridpoint. This sum peaks at 1.0 K near the currently valid observation; the westward and forward bias compared to the current ground track is visible.

The parameter λ determines the fraction of the analysed increment that is added onto the model field. It may be related to a "nudging coefficient" (e.g. Davies and Turner, 1977), G, by,

$$G = \lambda/(1 - \lambda)\Delta t \tag{3}$$

if the model timestep is Δt (about 450 s for these experiments) and the analysis correction process is regarded as equivalent to adding an implicit relaxation term $G(x_{\mathrm{obs}} - x)$ to the right hand side of the model equation for x, with x_{obs} the analyzed observation of x. The best results were found when G was increased from its southern hemisphere value on Earth to a value of 5×10^{-4} s^{-1}, midway between the value used for the terrestrial southern and northern hemispheres. The same value is used globally for the martian experiments in this paper.

Geostrophic balance increments were scaled by a factor of 0.7 in both extra-tropical hemispheres, as used for the terrestrial southern hemisphere (Lorenc *et al.*, 1991). The factor for the geostrophic scaling reduces linearly to zero at the equator from 30° latitude. The vertical scaling for geostrophic wind increments was simplified so that the full increment is applied from $\sigma = 1$ to $\sigma = 0.05$ and then decreases linearly to zero with σ. In the scheme for the Earth, the geostrophic increment falls above $\sigma = 0.55$ and is zero at $\sigma = 0.25$ and above. This difference presumably reflects the stronger tropopause and stratosphere on Earth, though the form of this function for Mars may be subject to further tuning as discussed later.

When applied operationally for the analysis and forecasting of the Earth's atmosphere, the use of the analysis correction scheme is combined with enhanced dissipation acting on the small-scale divergent components of the wind (known as divergence damping) in the general circulation model. Most martian data is anticipated to be in the form of temperatures and pressures which are balanced by geostrophic winds (modified by functions of height and latitude), and no extra divergence damping is used for the present martian simulations. Only increments to the vorticity, temperature and surface pressure fields are transformed back from the analysis grid, with divergence increments being set identically to zero. The assimilations still develop indirectly a larger divergence than the truth runs (by a factor of 10–20% when integrated globally), because the increments are not in full balance, but this is commonly found in assimilation experiments and did not cause a numerical problem here.

As a first step towards determining the appropriate values of the parameters for the analysis correction scheme applied to Mars, all the parameters of the scheme which may vary as a function of latitude were made symmetric about the equator and set to the values used in the terrestrial southern hemisphere (Lorenc *et al.*, 1991, Appendix 2), where observations on Earth are generally sparser. The length scales were re-scaled by the relative radii of the planets and the preferred zonal wavenumbers for mid-latitude baroclinic instabilities, which gives a combined factor close to one, and the timescales by the rotation periods of the planets, again close to one. The scheme was then tuned by a long series of numerical experiments, similar to those shown in the section on planetary wave experiments here, but for a northern hemisphere winter simulation (Lewis and Read, 1995). Observations of temperature profiles (degraded in resolution and with random errors as described below) as if sensed from an orbiter making 12 2/3 orbits per sol were used. The parameters of the assimilation scheme were varied to obtain the most accurate assimilations and fastest convergence in the retrieval of a wave of the correct amplitude and phase, as described in the section on planetary waves. The form of the time correlation function and the length of time over which observations entered the scheme was kept the same as for the Earth at a quarter of a rotation period (6 hours). The correlation length scale, S, and "nudging coefficient", G, were modified in small steps over a range of an order of magnitude in each direction about their original values. The dependency of the results on these parameters was not found to be very strong within a reasonable range of the estimated value and the best values were found close to the original estimates.

That the correlation scale, S, should be similar for Mars and the Earth is not surprising since although Mars has only about half the radius of the Earth it also typically has planetary waves of lower wavenumber (3–4 on Mars, Barnes, 1981, compared to 6–8 on Earth, Palmén and Newton, 1969). Equivalently, the Rossby deformation radius for mid-latitudes (NH/f, where $N \sim 10^{-2}$ s^{-1} is the buoyancy frequency, $H \sim 10$ km the atmospheric scale height and $f = 2\Omega \sin(\text{latitude}) \sim 10^{-4}$ s^{-1} the Coriolis parameter, e.g. Gill, 1982), at which scale buoyancy and rotational forces are equal, is of order 1000 km in both atmospheres. The similarity of the "nudging coefficient", G, should also be expected since the efficiency of the assimilation is governed by the ratio of the timescale G^{-1} to the rotational timescale f^{-1} (Lyne, 1979), and Earth and Mars have a similar rotational period. Thus the empirical tuning of the assimilation parameters produces length and timescales close to what might be predicted from those in use for the Earth and which also seem appropriate for the assimilation of data from a polar orbiting satellite with a period and ground track spacing as indicated above. As discussed later, in the section on the global circulation, some further refinements to the form of the vertical weighting for the geostrophic balance increments and the introduction of latitudinal variation of the "nudging coefficient" (as in Lorenc *et al.*, 1991) may be advantageous.

Artificial observations for assimilations

Two types of artificial data were generated from truth runs of the model for the experiments which are shown in the next two sections; temperature profiles (as a function of pressure) and surface pressure observations. It is possible to assimilate other types of observations, such as surface temperatures, wind speeds, dust and water vapour concentrations, but $T(p)$ and p_s are the two variables most likely to be available routinely during forthcoming spacecraft missions and are of fundamental importance in the determination of the global atmospheric state. Temperature profiles not only give information about the thermal structure of the atmosphere but also, indirectly, the vertical wind shear, and surface pressure observations give information about the atmospheric mass distribution and barotropic flow. It is possible to contemplate using either variable in isolation, as is done for temperature profiles in this paper and surface pressure observations in our previous work (Lewis and Read, 1995). Without the temperature profiles, however, one must rely on the model radiation scheme to determine the thermal state.

Temperature profiles

Temperature profiles were taken at regular intervals 90 times per orbit following the track of a polar orbiting satellite making 12 2/3 orbits per sol at an inclination of 92.866° and crossing the equator at 2am/2pm local time; the orbit planned for the Mars Observer spacecraft. The temperature was integrated over layers (weighted by mass) between a standard set of pressure boundaries, spaced one pressure scale height (about 10 km) apart, to simulate the vertical resolution of a remote sounding instrument. Thus each profile consists of eight temperatures spanning the region from the surface to about 80 km. A limb-sounding instrument such as the Pressure Modulator Infrared Radiometer, PMIRR (McCleese *et al.*, 1992), which will fly on Mars Surveyor 98, may in fact be able to obtain better resolution than this by a factor of about two. A nadir sounder such as the Thermal Emission Spectrometer, TES (Christensen *et al.*, 1992), which will fly on Mars Global Surveyor in 1996, may have lower vertical resolution, though the resolution of TES will approach one scale height over a certain altitude range.

It is more difficult to retrieve atmospheric temperatures near the surface where there is contamination by radiation from the ground. Either information from instrument weighting functions which overlap the surface must be discarded or surface properties (emissivity, temperature and, in the case of limb-sounders, the fractional field of view which is taken up by the surface) must be estimated, which introduces additional errors. To simulate the problem of retrieving temperatures close to the surface in a straightforward way, the temperature for any layer which touches the surface was removed from the profile and flagged to the assimilation scheme as missing. Gaussian errors were added to each temperature with a standard deviation of 1 K.

Surface pressure

The temperature profiles give temperature as a function of pressure. To insert them into the model at the correct altitude, or to obtain information about the barotropic component of the flow, the surface pressure must be known. Surface pressure could be sensed remotely with an infrared instrument, but in many situations not with high accuracy compared to the amplitude of the weather variations. Two extreme cases are shown in this paper. In the first case, it was assumed that surface pressure could be sensed accurately at each profile location; these assimilations are denoted $T(p)$ & p_s in the table and figure captions which follow. In the second, it was assumed that no useful information was available about the surface pressure and that the model surface pressure has to be believed as a first guess; these assimilations are denoted $T(p)$.

Further assimilations were conducted with intermediate scenarios resembling the second case, $T(p)$, but adding some surface pressure information at a limited number of sites whose locations are marked in Fig. 2. In the first case these were chosen to be the four sites proposed as configuration 1 for the INTERMARSNET mission (Chicarro *et al.*, 1994); these assimilations are denoted $T(p)$ & $p_s(I)$. To these were added four regularly spaced sites at 40°S latitude, $T(p)$ & $p_s(I, 40°)$, and four further sites at 60°S latitude, $T(p)$ & $p_s(I, 40°, 60°)$. Attention was focussed on the southern hemisphere to complement the INTERMARSNET sites, three of which lie there for this configuration. Surface pressure was reported once an hour at each site. Gaussian errors were added to all surface pressure data with a standard deviation of 1 Pa. Although landers are probably the best method of making repeated surface pressure measurements to high accuracy over long timescales it is also worth noting that radio occultation (Tyler *et al.*, 1992) could provide some surface pressure measurements of useful accuracy, especially in high latitudes.

Planetary waves

This section describes experiments which assess the ability of the assimilation scheme to reproduce transient baroclinic waves, given observations of the type described in the previous section. The model used to generate the data was run with 10 levels in the vertical using a thermal equilibrium temperature state chosen to give flows representative of southern hemisphere winter solstice on Mars. The radiative time-scale was taken to be 2 sols and a time-scale of 3 sols was taken for Rayleigh drag in the lowest model level. The model was run first without topography, which results in waves with a much simpler spectral structure for analysis. Under this combination of radiative and frictional timescale, either a strong wavenumber 3 flow or a more zonally-symmetric state with irregular, weaker waves are possible depending upon the initial conditions (Collins and James, 1995). These states after 120 sols of spin-up are illustrated, in terms of the temperature field on the 214 Pa surface, in Fig. 3. Both were spun-up from rest under the same conditions, with the weak wave state obtained by applying a smaller initial perturbation to the surface pressure field. Both types of states are persistent under the same conditions, and have been integrated forward for at least another 500 sols without observing any spontaneous transitions between the two states.

The truth run from which data was generated was the wavenumber 3 flow, consisting of the state shown in Fig. 3(a) integrated forwards for a further 30 sols. This wave has a mean period of about 4 sols, but the phase speed fluctuates between about 20 and 40 degrees per sol over the course of the 30 sol run. The amplitude also fluctuates by about 20%. It is thus not trivial to assimilate the wave accurately. While the period does not agree precisely with waves observed on Mars by Viking Lander 2 (Barnes, 1981), or those seen in more complete models, it is in the correct range for baroclinic wave activity on Mars and the surface pressure and near-surface temperature amplitude is similar. It thus provides a reasonable test for the assimilation scheme.

The assimilations were begun from the weak wave state, as shown in Fig. 3(b), and were conducted with the thermal forcing and drag turned off so that the models were not identical. It has already been shown for assimilations of this type with orbiter data that the results are remarkably insensitive to either the precise initial conditions or to the forcing timescale (Lewis and Read, 1995). A similar procedure was followed for northern hemisphere winter in Lewis and Read (1995), but there the orbiter and surface pressure data was not used in combination and, in the case of the orbiter data assimilations it was assumed that surface pressure could be sensed. This assumption has strong implications for the accuracy of the assimilation, as shown in the present paper.

The temperature state on the 214 Pa surface after 30 sols for the truth run and assimilations is shown in Fig. 4. This surface is at about 10.8 km in altitude, but is typical of the temperature maps at any level in this experiment. As can be seen, the assimilation which uses knowledge of the surface pressure at the temperature profile sites, Fig. 4(b), produces a very accurate reproduction of the temperature on this pressure surface, whereas that with no surface pressure information, Fig. 4(c), gives a less clear wavenumber 3 flow with increased evidence of other wavenumbers mixed in. The temperature gradient around the edge of the wave in Fig. 4(c) is somewhat weaker in accordance with the smaller amplitude of the wavenumber 3 component; nonetheless, the assimilation with temperature profiles alone still always gives a strong wavenumber three signal in both the temperature and vorticity field compared to the irregular weak wave state (the vorticity field is not shown). There is also a clear phase error, visible for example in the ridge near the top of the diagram, which lies closer to the 0° longitude line in the truth run. The effect of adding surface pressure observations at a few points, $T(p)$ & $p_s(I, 40°, 60°)$, is illustrated in Fig. 4(d); the cases with $T(p)$ & $p_s(I)$ and $T(p)$ & $p_s(I, 40°)$ data in fact look remarkably similar, except for barely detectable errors in the phase, and so are not shown for reasons of space. It can be seen that even adding a few points of surface pressure information has a dramatic impact in improving the temperature assimilations of the wave, almost to the degree obtainable with full knowledge of the surface pressure.

The global root-mean-square errors in the temperature on this pressure surface are shown in Fig. 5 for the duration of the assimilation. Most of the rapid improvement, as a result of the excitation of a strong wavenumber 3 flow in the southern hemisphere, can be seen to take place in the first 2 sols, with some further improvements in the cases with more surface pressure information in the next 3 or 4 sols. The global errors reach values as low as 0.6 K (compared to the 1 K standard deviation Gaussian errors applied to the artificial observations) mainly because of the accumulated accuracy in the northern hemisphere where waves are weak.

Global root-mean-square errors in the surface pressure are shown in Fig. 6. There are very clear differences between the assimilations with different surface pressure data. While the errors fall rapidly in the first 2 sols for the cases with complete knowledge of the surface pressure, or for the case with surface pressure measurements at 60° latitude, they stay roughly constant with more limited surface pressure information and actually increase when there is no surface pressure data at all. This is because the assimilation with temperature profiles alone actually spins-up a wave with a much stronger barotropic component around the winter pole, especially in the low wavenumbers, than is present in the truth run. The surface pressure data at high latitudes acts to control the amplitude of these components as shown in the next two diagrams.

Figure 7 shows maps of the transient component of the surface pressure 30 sols after the assimilation experiments were begun for both the truth run with a strong wavenumber 3 flow, Fig. 7(a), and for a forecast model run, Fig. 7(b), in which weak wave initial conditions were integrated forwards in time for 30 sols. Here a forecast is taken to mean an independent model integration, without data assimilation, but with a complete knowledge of the Newtonian cooling and Rayleigh friction terms, i.e. an identical twin of the truth model which is thus able to make the best possible prediction from the weak initial conditions, which are different from the truth run, although equally valid for this choice of parameters. This latter forecast run remains in an irregular, weak wave state for as long as it is integrated. The transient surface pressure is defined as the surface pressure minus the mean surface pressure over the previous 20 sols, and so shows the surface pressure component associated with travelling waves. As can be seen, the wavenumber 3 flow has a rather complicated pattern of surface pressure anomalies associated with it, whereas the irregular state has stronger wavenumber 1 and 2 components close to the winter pole.

Figure 8 shows the difference in transient surface pressure between the forecast in the irregular state, the five assimilation experiments and the true transient surface pressure in Fig. 7(a) at the same point in time. It is clear that the assimilation with knowledge of the surface pressure in Fig. 8(b) is close to the true field whereas the assimilation with no surface pressure information, Fig. 8(c), actually has worse errors near the winter pole than the unperturbed irregular state, Fig. 8(a), because of the strong barotropic wave that has been spun-up there. Even the introduction of the four INTERMARSNET sites, Fig. 8(d), only three of which lie in the southern hemisphere and only one of which lies poleward of 30° where there begins to be a significant wave signal, acts to control this error to a large extent. For example, although there are still errors poleward of 60° these are several times smaller than with no surface pressure information at all, despite the fact that there are no measurements being taken poleward of 40° latitude. This is an example of one of the advantages of assimilation over simple gridding; in finding a state consistent with the incomplete set of observations which it does have available, the model can predict some aspects of features which are apparently not observed directly. Courtier and Talagrand (1987) show a clear instance of this non-local feature of assimilation in the context of a variational scheme applied to a two-dimensional vorticity equation model of the Earth's atmosphere (Talagrand and Courtier, 1987). They assimilated observations of the 500mb geopotential and wind over the northern hemisphere for one day and found that the model was able to reconstruct the Aleutian depression, despite the fact that there were no observations in that region (Courtier and Talagrand, 1987). The remaining plots here, Fig. 8(e,f), demonstrate that as more surface pressure stations are added the error is further controlled, including the region between the pole and 60° latitude where no surface pressure observations are ever made.

Table 1 summarises the global root-mean-square errors in temperature and velocity on the 214 Pa surface and in surface pressure, averaged in time over the last 10 sols of each assimilation. These are compared with the forecast model run in which the weak wave initial conditions are integrated with a perfect model. It is apparent that all the assimilations produce a better fit to the temperature state, and that introduction of surface pressure observations results in increasingly smaller temperature errors as well as improved fits to the velocity and surface pressure.

The advantage of this idealized simulation with a clear spectral peak at a single wavenumber (and its harmonics) is that it allows the possibility of diagnosing the amplitude and phase of that wavenumber as another measure of the accuracy of the assimilations (Lewis and Read, 1995). Root-mean-square errors alone can be misleading in this type of experiment, since relatively small errors in phase with a strong wave flow give global errors as large as those from a zonally-symmetric or weak wave state. Table 2 shows the amplitude and phase errors in the zonal wavenumber 3, gravest latitudinal spherical harmonic of the temperature field on the 214 Pa pressure surface. This mode is symmetric between the hemispheres; there is an equally large antisymmetric component since the wave is only present in the southern hemisphere. The total amplitude of that mode in the true case is about 4.0 K and fluctuates by about 0.5 K over one sol as the wave varies in amplitude during the course of the experiment. It can be seen that all the assimilations capture the amplitude of the component to much better accuracy than its own variability. The phase speed is also reproduced much more accurately than its variations. Furthermore, increasing amounts of surface pressure information lead to ever more accurate representations of the temperature mode in terms of both amplitude and phase, so that the assimilation with surface pressure sites at 60° S reproduces the wave as well as that with surface pressure information at every temperature profile site. It is also apparent that even the case with surface pressure measured no further poleward than at 40° S actually captures the amplitude of the wave about as accurately as the assimilations with more surface pressure information; the larger root-mean-square errors in this case come mainly from the greater error in the phase of the wave.

Global circulation

This section describes assimilation experiments which test the ability of the assimilation scheme, with similar data to that used in the previous section, to reproduce the global temperatures and circulation in a model with greater coverage in the vertical and including realistic martian topography. The model was run with a stretched grid of twenty levels in the vertical covering the surface to about 90 km altitude. This is slightly greater than the height range over which temperature profiles are likely to be retrieved from an infrared thermal sounder, perhaps up to 80 km altitude (McCleese *et al.*, 1992). With the model top raised to 90 km altitude the highest three levels of the model (at 80 km and above) had an increasingly strong Rayleigh friction term which acted on eddy components alone in order to limit reflection of waves from the top of the model. For these experiments a topographic map was included, based on the Digital Terrain Model for Mars (U.S. Geological Survey, 1991). The timescales for Newtonian cooling and for Rayleigh friction were as used before (2 and 3 sols respectively) but the equilibrium temperature state which was used for the heating was based on a time-mean over 100 sols centred on the southern winter solstice from a martian general circulation model employing the same dynamical schemes but a more complete set of physical parameterizations (Hourdin *et al.*, 1993), including a radiation scheme for dust heating and CO_2 cooling. The reason for generating data from the simpler model as a first step, rather than the more complete model directly, was to avoid some of the complications arising from the lack of various physical schemes, such as convective adjustment, in the simple model used for assimilation.

The initial state used for the model was a horizontally-uniform vertical temperature profile, approximately equal to the global mean of the thermal equilibrium state, with the atmosphere at rest and the surface pressure in hydrostatic balance with a state of no motion. For the truth run this state was integrated forwards in time for 500 sols to establish a statistically-steady circulation. Data was then taken over the next 50 sols. The initial state for the assimilations (begun at sol 500 compared to the truth experiment) was at rest, as used to initialize the truth experiment spin-up, with globally-averaged temperatures on pressure surfaces. Assimilations were conducted for 50 sols (sol 500–550 of the truth experiment) with the same combinations of data as used in the previous section, although the temperature profiles extended up to about 80 km. Once again the model used for assimilation had no Newtonian cooling or Rayleigh friction.

Figure 9 shows the zonally-averaged temperature and zonal velocity from the last ten sols (540–550) of the truth run. All the assimilations were able to reproduce the zonal-mean temperature structure to a high degree of accuracy over most of the atmosphere and so error maps only are shown for two experiments; Fig. 10 shows the error in zonal-mean temperature and zonal velocity for an assimilation with temperature profile information but no surface pressure observations, and Fig. 11 shows the same maps for an assimilation using the same temperature profiles but including surface pressure observations at the four INTERMARSNET sites once an hour. The assimilation scheme is able to create a temperature field with strong, and realistic, latitudinal gradients despite the fact that there is no thermal forcing by radiation in the model used for assimilation.

In both assimilations, the errors in zonal-mean temperature over the bulk of the atmosphere are much less than 1 K; the root-mean-square errors in the full three-dimensional temperature field are about 1.3 K in the atmosphere above 10 km altitude. The only area of significant error is close to the surface, especially near the equator where errors may reach 10 K very close to the ground. This is largely related to the fact there are few observations here and to the consequent poor representation of the mean meridional circulation, as discussed later. Many of the standard pressure layers for observations cut the ground and the corresponding observations are discarded. There are few valid observations for the lowest 10 km of atmosphere, therefore, especially in the southern hemisphere and near the equator where the topography is high, and near the equator the observations are more widely separated between orbits than they are nearer the poles. In the case of assimilation of data from a model with more complete physics (such as Hourdin *et al.*, 1993), or more importantly in the case of assimilation of real observations, into a simpler model, such as that used here, one might also expect errors near the surface as a result of the model's inability to represent accurately the planetary boundary layer and exchanges of radiation with the surface.

The zonal winds are also reproduced accurately over much of the atmosphere. For example, over most of the strong eastward jet in the winter hemisphere between 30°S and the pole, the errors are less than 2 m s^{-1} compared to a peak wind speed of greater than 110 m s^{-1}. Since the jet amplitude and position can be obtained accurately, even from temperature observations alone, the winds there must be closely balanced with the temperature structure. The area of significant error in the zonal-mean winds is clearly in the high equatorial regions, where thermal wind balance breaks down, and near where the winter jet meets the ground. This corresponds to the regions where errors in the temperature gradient are large near the ground.

A further diagnostic of the assimilation is the extent to which it is able to recover the correct mean meridional atmospheric circulation. Mass streamfunctions for the truth run and the assimilations with and without surface pressure information are shown in Fig. 12. The Hadley cell which straddles the equator in Fig. 12(a) is poorly represented in both assimilations. In the case of the assimilation with temperature profiles only, Fig. 12(b), this cell is reduced by a factor of at least three in strength, is compressed to lower altitudes and is split with peaks on both sides of the equator. The much weaker circulation in this region is associated with the larger temperature errors near the surface and also perpetuates them; the weak overturning motion means that low-level air is never carried up to higher altitudes where its potential temperature might be corrected by assimilation increments. Errors in the zonal wind over the equatorial region also must be related to the weaker meridional transport of angular momentum in the assimilation. Interestingly, the assimilation with surface pressure observations added at the INTERMARSNET sites alone, Fig. 12(c), improves the recovery of the mean meridional circulation. Although the Hadley cell is still substantially in error, it is stronger and less confined in the vertical. This may help to explain why the velocity and temperature errors are slightly smaller in this case, Fig. 11.

It is to be expected that the mean meridional circulation is a very difficult test for the assimilation procedure, especially when the only observations are of temperature. For example, the Hadley cell is about three times stronger and more extensive in both latitude and altitude in the model with more complete physical schemes which was used to generate the equilibrium temperature field for the Newtonian heating in this experiment, despite the fact that the temperature fields appear remarkably similar; it is also stronger at a similar season in another martian model (Haberle *et al.*, 1993, Fig. 10). It is known that the Hadley circulation is sensitive to the form of the model boundary layer scheme, the presence or otherwise of convection and of divergent flow from one polar cap to the other. Thus, it is difficult to constrain the Hadley cell by temperature measurements alone, however accurate, if the model physics is incorrect. The difference in model physics between the run used to generate the true state and the assimilation is the presence of Newtonian heating and of Rayleigh friction in the lowest layer of the truth experiment. The assumption of zero drag in the lowest layer in the assimilation is a fairly strong one and so some additional assimilations of temperature profiles were tried with Rayleigh friction with a timescale of 3 sols, as in the true case, and 1 sol, which is a stronger drag though not unreasonable. In both cases the mean meridional circulation appeared almost identical to that in Fig. 12(b), despite the fact that the details of the low level cross-equatorial flow were different, being dominated by inertial effects in the low drag cases and frictional effects in the high drag case (Joshi *et al.*, 1994, 1995). The temperature errors were very similar to those in Fig. 10, and there was some minor reduction in the velocity errors. The correct form of boundary friction, therefore, offered almost no improvement over the assumption of zero friction, in contrast to the addition of very few surface pressure measurements, Fig. 12(c), which were beneficial.

These results indicate that the large temperature errors at the ground and the related poor mean meridional circulation are primarily a result of the missing data at the bottom of the temperature profiles and the lack of other near-surface observations. The temperature measurements higher in the atmosphere are insufficient to set up the mean meridional circulation on their own. The initial conditions of no motion are a poor guess at the atmospheric state, and in practice a better initial state would already have a Hadley cell, though its strength might be poorly constrained by the assimilation directly. This situation could be improved by assimilation of actual surface temperatures from a remote sensing instrument, in combination with a better model boundary layer scheme, and the results with a few surface pressure measurements here give encouragement that surface pressure and near-surface temperatures and winds from landers might also be helpful.

The larger wind errors high in the atmosphere, especially over the equator, might be also related to the fact that the balancing wind increments in the assimilation scheme are calculated using geostrophic balance, the weighting of which is reduced to zero between 30° latitude and the equator. The greatest errors often lie over the equator and at altitudes above 30 Pa where the geostrophic increments are reduced with pressure over the extra-tropical region as well (see earlier section). Experiments are currently being conducted to improve the balancing procedure in this region, perhaps by extending the geostrophic increments to lower pressures everywhere and by implementing an improved balance relationship near the equator. Another improvement may be to reduce the strength of the nudging coefficient for the insertion of assimilation increments in the tropical region; a constant value was used for the experiments here, whereas for Earth it is often varied over different latitude bands (Lorenc *et al.*, 1991). Lyne (1979) shows that if the nudging coefficient, G (Eqn. 3), becomes too large compared to the Coriolis parameter, $f = 2\Omega \sin(\text{latitude})$, then assimilation of wind data can be hindered.

It is generally true that the global wind errors in the assimilation with partial surface pressure data, even at only four points, Fig. 11, are slightly smaller than in that with none, Fig. 10. For example, the error in the zonal wind at the ground at 40°S is about 10 m s^{-1} in Fig. 10 compared to about 6 m s^{-1} in Fig. 11; this may be due in part to an improved representation of the descending branch of the Hadley cell, Fig. 12. There may also be a smaller error in the barotropic component of the winds globally, through an improved surface pressure field. Figure 13 shows the global root-mean-square surface pressure error for each experiment over the first 5 sols. The errors are systematically lower with increasing amounts of surface pressure data in the assimilation, although the differences are less spectacular than in the planetary wave case in the previous section. Even the experiment with no surface pressure information shows some improvement in the surface pressure field when there is topography, presumably due to the re-arrangement of atmospheric mass as a more realistic thermal state is built up. Even though most of the surface pressure sites lie between the equator and 60° S, the errors in the $T(p)$ & $p_s(I, 40°, 60°)$ assimilation are about half way between those with no surface pressure information and those with surface pressure at every profile site (1140 different sites per sol, spread over the whole globe) and show a less dramatic initial increase in global error. Thus even limited surface pressure information may be very valuable in constraining the assimilation as a whole.

Conclusions

The experiments described here show that the concept of assimilation of spacecraft data into a martian atmospheric model is potentially feasible and worthwhile. Useful information may be extracted about planetary waves with periods and amplitudes comparable to those in the martian atmosphere. Temperature soundings from a single polar orbiting satellite may be used to reconstruct the thermal state of the atmosphere and indirectly the velocity field. The process is robust; temperatures and zonal velocities over much of the atmosphere were recovered to a reasonable degree of accuracy despite the severe test of no thermal forcing and zero boundary layer friction in the assimilation model, the lack of temperature measurements near the surface and the consequent errors in the mean meridional circulation. Simultaneous surface pressure measurements, from even a small number of sites, have been shown to enhance substantially the quality of such analyses. It therefore seems desirable that temperature soundings from an orbiting satellite and meteorological observations from surface stations, such as

INTERMARSNET, be obtained simultaneously if possible.

It is worth noting that in all experiments performed here the number of surface pressure measurements is tiny compared to the number of such measurements that are routinely assimilated into a numerical weather prediction model for the Earth. For the Earth these measurements often have the greatest impact on the analysis, and it is thus a stringent test of the whole assimilation approach to attempt to find a beneficial impact upon the model as a result of so few, widely-distributed observations. Attention here has been concentrated upon southern hemisphere winter assimilations and, of course, more surface pressure sites would be needed to characterize fully the global circulation over the course of a full martian year. Experiments, such as those shown here, could help in making decisions such as whether it is better to concentrate on one hemisphere and characterize the planetary waves there accurately, or to accept a lower degree of accuracy but attempt to cover the whole globe, within practical constraints. The scientific rationale for a larger network, which only employs pressure sensors, is described by Haberle and Catling (1996). Furthermore, a low-cost, global network mission appears technically feasible (Merrihew *et al.*, 1996). The additional value of having even a few surface pressure measurements at mid-latitudes and poleward, for the analysis of planetary waves, is clear from the assimilation experiments in the present paper.

The analysis correction data assimilation scheme is particularly suitable for the martian problem because the iterative technique places no requirement to cast all the observations to synoptic hours when they have, in fact, been made continuously. This is in contrast to many optimal interpolation schemes in use for weather forecasting on the Earth (Lorenc, 1986), and may also produce improved results where observations are sparse because increments are spread over relatively large areas (Lorenc *et al.*, 1991). The analysis correction scheme also offers the important advantage of being computationally inexpensive and reasonably straightforward to implement and change, especially compared to more sophisticated techniques such as four-dimensional variational assimilation (Talagrand and Courtier, 1987) which offer some formal advantages in principle. In practice, one iteration of the analysis correction scheme takes less time than one model time step, compared to the need to integrate a forward and adjoint model many times (of order several tens) for a variational scheme. Thus a more sophisticated, high resolution model may be used to conduct an assimilation with the analysis correction scheme in real time given finite dedicated computing resources. Whether this is a real advantage depends upon how well the simpler analysis correction scheme may be tuned to approach the optimum analysis.

An alternative to the four-dimensional variational approach is the Kalman filter (Ghil *et al.*, 1981) which, in its optimal form, is prohibitively expensive for atmospheric general circulation models because it must advance in time the prediction error covariance matrix, which is of dimension N^2 for a model with N variables. A steady-state approximation to the Kalman filter has been made (Banfield *et al.*, 1995) for a variety of simple models, in the case where the observations are taken in a regular pattern following the ground track of a polar-orbiting satellite. This procedure is computationally inexpensive compared to the full Kalman filter, and closely approaches the optimal gains for a single, regular data set, but is perhaps less flexible than the analysis correction scheme for more complicated data sets which may contain different types of observations, changes in scan pattern, missing data points and odd observations randomly distributed in time and space. Interestingly, although the analysis correction scheme used here was tuned independently of the steady-state Kalman filter work, both schemes produce accumulated gain functions of similar scale and longitudinal shape in mid-latitudes for the case of a single polar-orbiting satellite. The gain function is asymmetric and biased towards the west of the current observation. In the case of the analysis correction scheme this is because of the asymmetric function used to spread the observations in time; the steady-state Kalman filter does not spread observations in time, but inserts each observation once only with a pre-computed gain function of constant shape (see Banfield *et al.*, 1995). Thus a difference between the schemes is that with the analysis correction scheme the westward bias is produced by including observations from future times centred on their correct position in space, whereas with the steady-state Kalman filter the bias is produced by spreading observations valid at the current time to greater distances in space. In both cases this westward bias may be understood as a consequence of the satellite ground track moving westwards with successive orbits; the location to the east was observed one orbit ago, and should not need to be further updated, whereas the location to the west was last observed almost one day ago and can benefit from an update with more recent increments. The temporal function used here implies that, at any one instant, data from two orbital tracks to the west of the current satellite ground track may be entering the scheme, whereas that from the track to the east will already have been released, for typical polar orbiter periods of around two hours as in the case of Mars Observer.

Further publications with results from the analysis correction scheme will include assimilations with data obtained from, and inserted into, a more realistic martian model which uses sophisticated schemes to parameterize various physical processes. Use of the more complex model to generate the data directly, combined with assimilation into simpler models, such as that used here, will allow the sensitivity of the results of the assimilations to specific features of the models to be further tested. As well as the introduction of new problems arising from, for example, increased divergent flow associated with the diurnal cycle, such parameterization schemes may actually help to improve the assimilations in some instances. For example, it is possible to assimilate surface temperatures which can be measured from orbit. A boundary layer and vertical mixing scheme could then help to reduce temperature errors low in the atmosphere, such as those seen here (Figs. 10,11) where near-surface atmospheric measurements are missing. Additional data may also be assimilated in future, in particular distributions of dust loading and water vapour concentrations, once transport schemes have been introduced. The dust loading inevitably has a large impact on the radiative scheme, and assimilations of dust and water will allow schemes for sources, transport and sinks to be tested against observations as well as providing additional constraints upon the model winds.

It is clearly desirable to use the most realistic model possible when assimilating real observations from forthcoming spacecraft missions; the analysis correction scheme interfaced to the dynamical code of the martian general circulation model as presented here will allow this process to be conducted with the flexible introduction of further physical parameterization schemes. The results presented here, even with the almost purely dynamical model, suggest that data assimilation will be a useful technique for the analysis and interpretation of real observations of the type expected from future missions such as Mars Surveyor and INTERMARSNET.

Acknowledgements. We thank Drs. A.C. Lorenc and R.S. Bell at the United Kingdom Meteorological Office for their advice and for making the original code for the analysis correction data assimilation scheme available to us. This work was supported by a grant from the United Kingdom Particle Physics and Astronomy Research Council.

References

Banfield, D., Ingersoll, A.P. and Keppenne, C.L., A steady-state Kalman filter for assimilating data from a single polar-orbiting satellite. *J. Atmos. Sci.* **52,** 737–753, 1995.

Barnes, J.R., Midlatitude disturbances in the martian atmosphere: A second Mars year, *J. Atmos. Sci.* **38,** 225–234, 1981.

Bell, R.S. and Dickinson, A., The Meteorological Office operational numerical weather prediction system, *Sci. Paper No. 41,* U.K. Meteorological Office, 1987.

Chicarro, A., Scoon, G. and Coradini, M., INTERMARSNET — An international network of stations on Mars for global martian characterisation, *ESA Journal* **18/3,** 207–218, 1994.

Christensen, P.R., Anderson, D.L., Chase, S.C., Clark, R.N., Kieffer, H.H., Malin, M.C., Pearl, J.C., Carpenter, J., Bandiera, N., Brown, F.G. and Silverman, S., Thermal Emission Spectrometer experiment: Mars Observer mission, *J. Geophys. Res.* **97 (E5),** 7719–7734, 1992.

Collins, M. and James, I.N., Regular baroclinic transient waves in a simplified global circulation model of the martian atmosphere, *J. Geophys. Res.* **100 (E7),** 14421–14432, 1995.

Courtier, P. and Talagrand, O., Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results, *Quart. J. R. Met. Soc.* **113,** 1329–1347, 1987.

Daley, R., *Atmospheric Data Analysis,* Cambridge University Press, Cambridge, 1991.

Davies, H.C. and Turner, R.E., Updating prediction models by dynamical relaxation: an examination of the technique, *Quart. J. R. Met. Soc.* **103,** 225–245, 1977.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K. and Isaacson, E., Applications of estimation theory to numerical weather prediction, in *Dynamic meteorology: Data assimilation methods* (edited by L. Bengtsson, M. Ghil and E. Kallen), pp. 139–224, Springer-Verlag, New York, 1981.

Gill, A.E., *Atmosphere-Ocean Dynamics*, Academic Press, San Diego, 1982.

Haberle, R. M. and Catling, D. C., A micro-meteorological mission for global network science on Mars: Rationale and measurement requirements, *Planet. Space Sci.*, this issue, 1996.

Haberle, R.M., Pollack, J.B., Zurek, R.W., Leovy, C.B., Murphy, J.R., Lee, H. and Schaeffer, J., Mars atmospheric dynamics as simulated by the NASA Ames general circulation model: 1. The zonal-mean circulation, *J. Geophys. Res.* **98 (E2),** 3093–3123, 1993.

Hoskins, B.J. and Simmons, A.J., A multi-layer spectral model and the semi-implicit method, *Quart. J. Roy. Meteor. Soc.* **101,** 637-655, 1975.

Hourdin, F., Le Van, P., Forget, F. and Talagrand, O., Meteorological variability and the annual surface pressure cycle on Mars, *J. Atmos. Sci.* **50,** 3625–3640, 1993.

Joshi, M.M., Lewis, S.R., Read, P.L. and Catling, D.C., Western boundary currents in the atmosphere of Mars, *Nature* **367,** 548–551, 1994.

Joshi, M.M., Lewis, S.R., Read, P.L. and Catling, D.C., Western boundary currents in the martian atmosphere: Numerical simulations and observational evidence, *J. Geophys. Res.* **100 (E3),** 5485–5500, 1995.

Lacarra, J.F. and Talagrand, O., Short-range evolution of small perturbations in a barotropic model, *Tellus* **40A,** 81–95.

Lewis, S.R. and Read, P.L., An operational data assimilation scheme for the martian atmosphere, *Adv. Space Res.* **16 (6),** 9–13, 1995.

Leith, C.E., Nonlinear normal mode initialization and quasi-geostrophic theory, *J. Atmos. Sci.* **37,** 958–968, 1980.

Lorenc, A.C., Analysis methods for numerical weather prediction, *Quart. J. R. Met. Soc.* **112,** 1177–1194, 1986.

Lorenc, A.C., Bell, R.S. and Macpherson, B., The Meteorological Office analysis correction data assimilation scheme, *Quart. J. Roy. Meteor. Soc.* **117,** 59–89, 1991.

Lorenz, E.N., Attractor sets and quasi-geostrophic equilibrium, *J. Atmos. Sci.* **37,** 1685–1699, 1980.

Lyne, W.H., Data assimilation by repeated correction of model fields — theoretical considerations, *Met. O. 11 Tech. Note No. 130*, U.K. Meteorological Office, 1979.

McCleese, D.J., Haskins, R.D., Schofield, J.T., Zurek, R.W., Leovy, C.B., Paige, D.A. and Taylor, F.W., Atmosphere and climate studies of Mars using the Mars Observer Pressure Modulator Infrared Radiometer, *J. Geophys. Res.* **97 (E5),** 7735–7757, 1992.

Merrihew, S. C., Haberle, R. M. and Lemke, L. G., A micro-meteorological mission for global network science on Mars: A conceptual design, *Planet. Space Sci.*, this issue, 1996.

Palmén, E., and Newton, C.W., *Atmospheric Circulation Systems,* Academic Press, New York, 1969.

Talagrand, O. and Courtier, P., Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quart. J. R. Met. Soc.* **113**, 1311–1328, 1987.

Tyler, G.L., Balmino, G., Hinson, D.P., Sjogren, W.L., Smith, D.E., Woo, R., Asmar, S.W., Connally, M.J., Hamilton, C. and Simpson, R.A., Radio science investigations with Mars Observer, *J. Geophys. Res.* **97 (E5),** 7759–7779, 1992.

U.S. Geological Survey, Topographic map of Mars, *U.S.G.S. Misc. Inv. Series Map I-2160*, 1991.

Vautard, R. and Legras, B., Invariant manifolds, quasi-geostrophy and initialization, *J. Atmos. Sci.* **43,** 565–584, 1986.

Fig. 1. A plot of the total accumulated increments from the analysis correction scheme at one level in the vertical, assuming that all observations imply an equal increment of +1.0 K and that all have equal errors. The small crosses mark the locations of the observations which are included in the analysis, taken following the ground track of a polar orbiting satellite which makes 12 2/3 orbits per sol. The currently valid observation is outlined by a square at (0°,0°) and the satellite ground track moves westwards with time. The grey shading shows the total increment on the model grid calculated as described in the text. This would be weighted by a relaxation increment, $\lambda \sim 0.2$, before being added to the model temperature field at the end of a timestep. White areas in the plot are not influenced by the assimilation at this timestep; the four shades of grey correspond to increments of below 0.25 K, 0.25–0.5 K, 0.5–0.75 K and 0.75–1.0 K in order of increasing darkness. The increments peak at 1.0 K near the currently valid point.

Fig. 2. A map of Mars showing the positions of sites where surface pressure was measured for the assimilation experiments. Longitudes are given in degrees East. The four triangles mark a possible set of sites for the INTERMARSNET mission, at (-78.2°,40.1°), (-73.2°,-9.7°), (-127.5°,25.0°) and (125.0°,-43.0°) where the coordinates are given as (longitude, latitude). The eight squares are additional sites, used for experiments in this paper, regularly-spaced in longitude at 40°S and 60°S latitude. The contours show the topography at 2 km intervals from the USGS Digital Terrain Model. In this and future contour maps positive contours are solid lines, the zero contour is dotted and negative contours are dashed.

Fig. 3. Polar stereographic projections of the southern hemisphere (south pole at the centre of the diagrams, the equator at the edges, grid interval 30°) showing temperature on the 214 Pa pressure surface (about 10.8 km altitude) for the initial conditions for the assimilation experiments; **(a)** a strong wavenumber three flow, **(b)** an irregular, weak wave flow. Both are the result of 120 sol spin-ups of the model with the same parameters but different initial conditions.

Fig. 4. Polar stereographic plots of the temperature on the 214 Pa pressure surface 30 sols after beginning assimilation. The south pole is at the centre of each plot and the edges are here at 30°S in order to focus on the wave in southern winter mid-high latitudes. Plot **(a)** is the true state, the rest are assimilations; **(b)** $T(p)$ & p_s, **(c)** $T(p)$ and **(d)** $T(p)$ & $p_s(I, 40°, 60°)$. Assimilations with 4 and 8 surface pressure sites appear very similar to **(d)**, though with slightly larger errors in phase, and so are not shown.

Fig. 5. The global (area weighted) root-mean-square temperature error on the 214 Pa surface as a function of time after the start of assimilation for five experiments. All assimilations used the same temperature profiles and varying surface pressure information; **(a)** $T(p)$ & p_s, **(b)** $T(p)$, **(c)** $T(p)$ & $p_s(I)$, **(d)** $T(p)$ & $p_s(I, 40°)$ and **(e)** $T(p)$ & $p_s(I, 40°, 60°)$.

Fig. 6. The global (area weighted) root-mean-square surface pressure error for five assimilation experiments labelled **(a)**–**(e)** as in Fig. 5.

Fig. 7. The transient surface pressure after 30 sols (the surface pressure at sol=30 minus the mean surface pressure over the previous 20 sols) for **(a)** a strong wave three flow (the true state) and **(b)** an irregular flow (obtained by integrating the weak wave initial conditions over 30 sols with no assimilation). The plot is a polar stereographic projection of the southern hemisphere. Contour interval 2.5 Pa.

Fig. 8. The transient surface pressure error (assimilated state minus true state) after 30 sols of assimilation experiments. For comparison **(a)** shows a forecast from the initial conditions but with the use of a perfect model. The remaining plots are assimilations with temperature profiles and varying surface pressure information; **(b)** $T(p)$ & p_s, **(c)** $T(p)$, **(d)** $T(p)$ & $p_s(I)$, **(e)** $T(p)$ & $p_s(I, 40°)$ and **(f)** $T(p)$ & $p_s(I, 40°, 60°)$. The positions of landers are marked with triangles and squares as in Fig. 2. Contour interval 5 Pa.

Fig. 9. The true zonal mean state, time averaged over the last 10 sols of a twenty level experiment with topography: **(a)** temperature, contour interval 10 K; **(b)** zonal velocity, contour interval 10 m s^{-1}. The heavy line at the bottom of the contour map is the zonal mean topography.

Fig. 10. Errors in the zonal mean state over the last 10 sols of a 50 sol assimilation (corresponding to the period in Fig. 9) using only $T(p)$ data (assimilated mean state minus true mean state): **(a)** temperature, contour interval 1 K, errors greater than 1 K are shaded; **(b)** zonal velocity, contour interval 2 m s^{-1}, errors greater than 5 m s^{-1} are shaded.

Fig. 11. Errors in the zonal mean state over the last 10 sols of a 50 sol assimilation (as Fig. 10) using $T(p)$ & $p_s(I)$ data (assimilated mean state minus true mean state): **(a)** temperature, contour interval 1 K, errors greater than 1 K are shaded; **(b)** zonal velocity, contour interval 2 m s^{-1}, errors greater than 5 m s^{-1} are shaded.

Fig. 12. The zonal mean meridional circulation, time averaged over the same period as Figs. 9–11, shown as a mass streamfunction with a contour interval of 10^8 kg s^{-1}: **(a)** the true state as in Fig. 9; **(b)** from an assimilation using $T(p)$ data as in Fig. 10; **(c)** from an assimilation using $T(p)$ & $p_s(I)$ data as in Fig. 11. The contours are labelled in units of -1.0×10^8 kg s^{-1}, with the negative sign used so that the dominant Hadley Cell is plotted with solid lines following the convention used here, with the zero contour dotted and negative contours dashed. The sign of the circulation is such that there is rising motion in the summer, northern hemisphere and descent in the winter, southern hemipshere. Only the lower half of the model atmosphere is shown; flow in the upper half is relatively weak and similar in all three cases.

Fig. 13. Global root-mean-square errors in the surface pressure field over the first 5 sols of assimilation into the simple model with topography (Fig. 9). All assimilations used the same temperature profiles and varying surface pressure information; **(a)** $T(p)$ & p_s, **(b)** $T(p)$, **(c)** $T(p)$ & $p_s(I)$, **(d)** $T(p)$ & $p_s(I, 40°)$ and **(e)** $T(p)$ & $p_s(I, 40°, 60°)$.

Table 1. Global root-mean-square errors in surface pressure, Δp_s, and in temperature, ΔT_{214}, and velocity, Δu_{214}, on the 214 Pa surface (about 10.8 km altitude). The first row is a forecast made from the initial conditions with a perfect model; the remaining rows are for assimilation experiments detailed in the text. A mean has been taken over sol 20–30 of the assimilation.

Experiment	Δp_s (Pa)	ΔT_{214} (K)	Δu_{214} (m s^{-1})
Forecast	20.0	2.20	12.1
$T(p)$ & p_s	1.8	0.53	3.0
$T(p)$	29.0	1.30	12.0
$T(p)$ & $p_s(I)$	20.0	1.10	9.3
$T(p)$ & $p_s(I, 40°)$	20.0	0.96	9.3
$T(p)$ & $p_s(I, 40°, 60°)$	4.0	0.62	4.1

Table 2. Error in the amplitude and phase of the zonal wavenumber three spherical harmonic of the temperature field on the 214 Pa pressure surface (mean over sol 20–30). The assimilations are as in Table 1. The forecast experiment gives much larger amplitude errors (by a factor of about 10), and essentially random phase errors, and is not shown.

Experiment	Amplitude error (K)	Phase error (longitude)
$T(p)$ & p_s	-0.08	3.7°
$T(p)$	-0.18	8.7°
$T(p)$ & $p_s(I)$	-0.12	8.5°
$T(p)$ & $p_s(I, 40°)$	-0.08	7.2°
$T(p)$ & $p_s(I, 40°, 60°)$	-0.07	3.9°

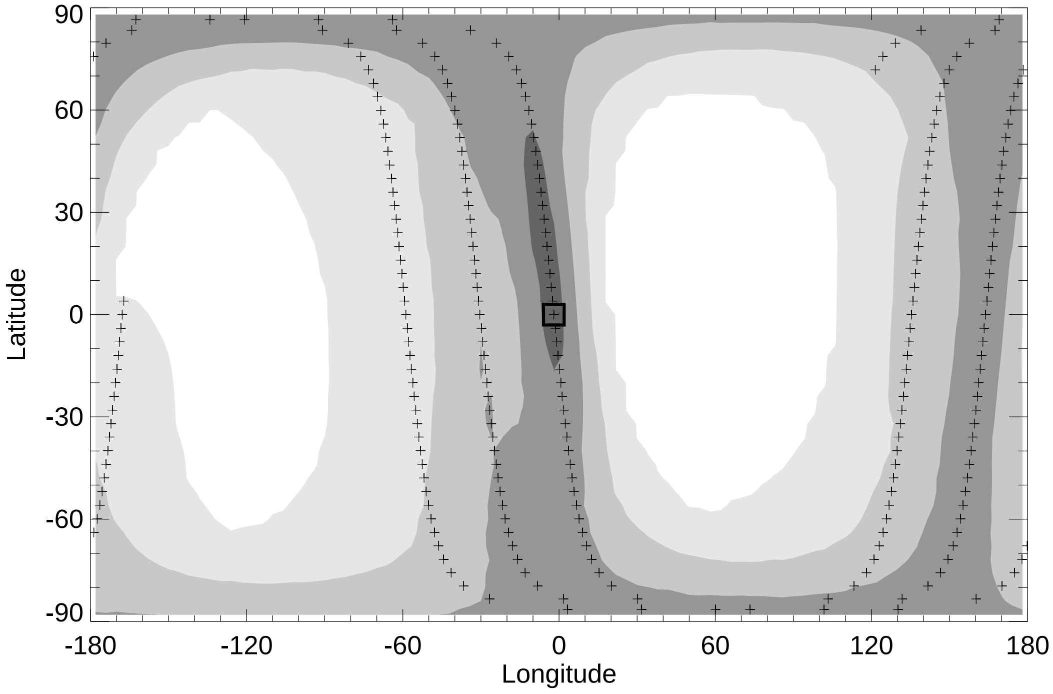


Fig. 1

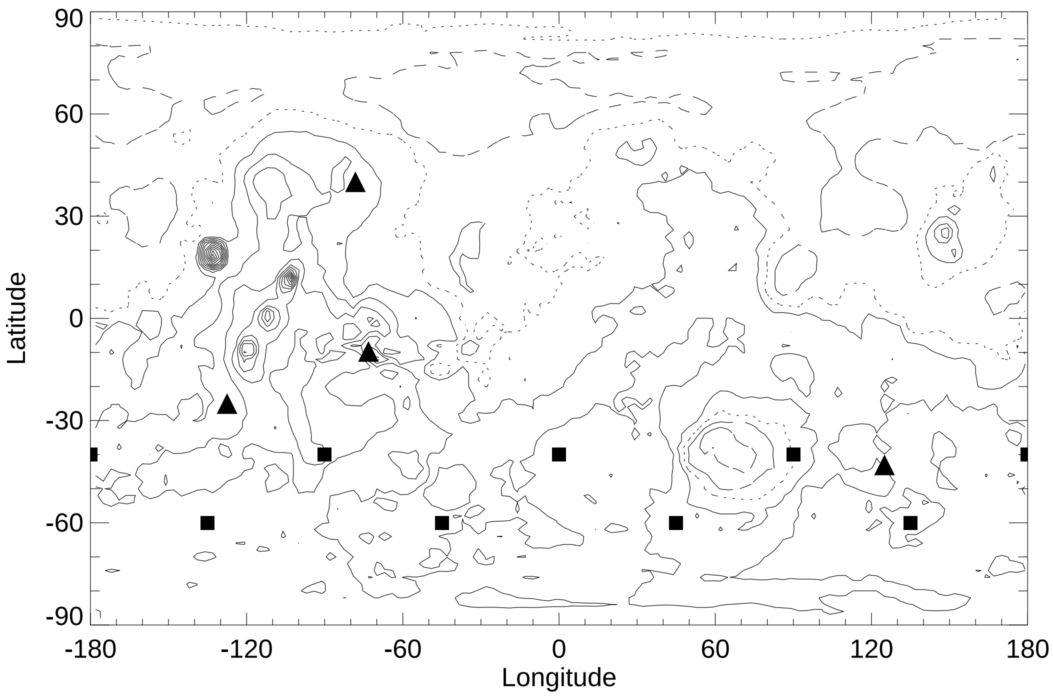


Fig. 2

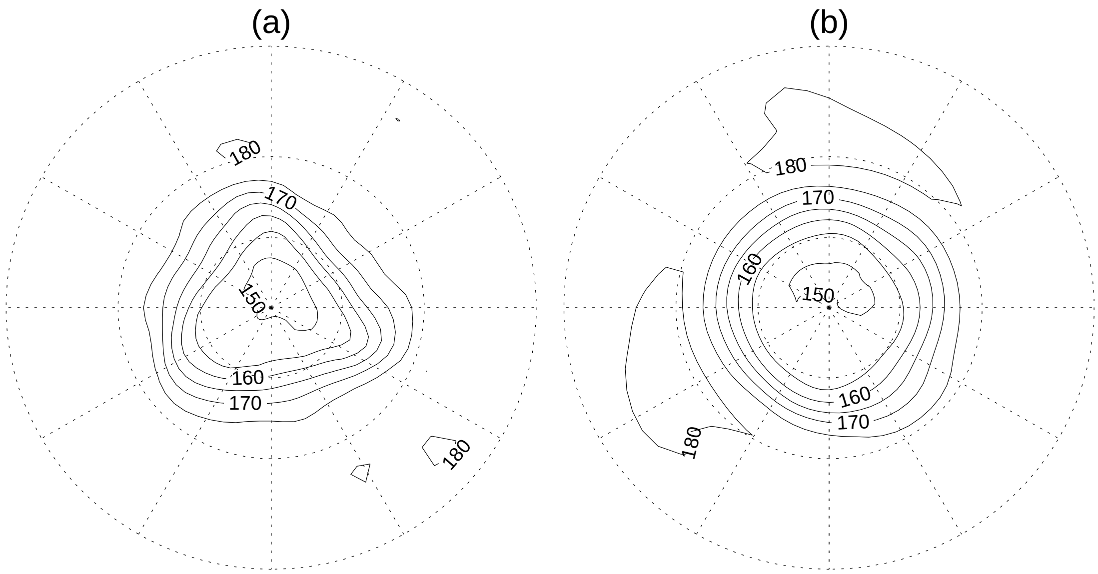


Fig. 3

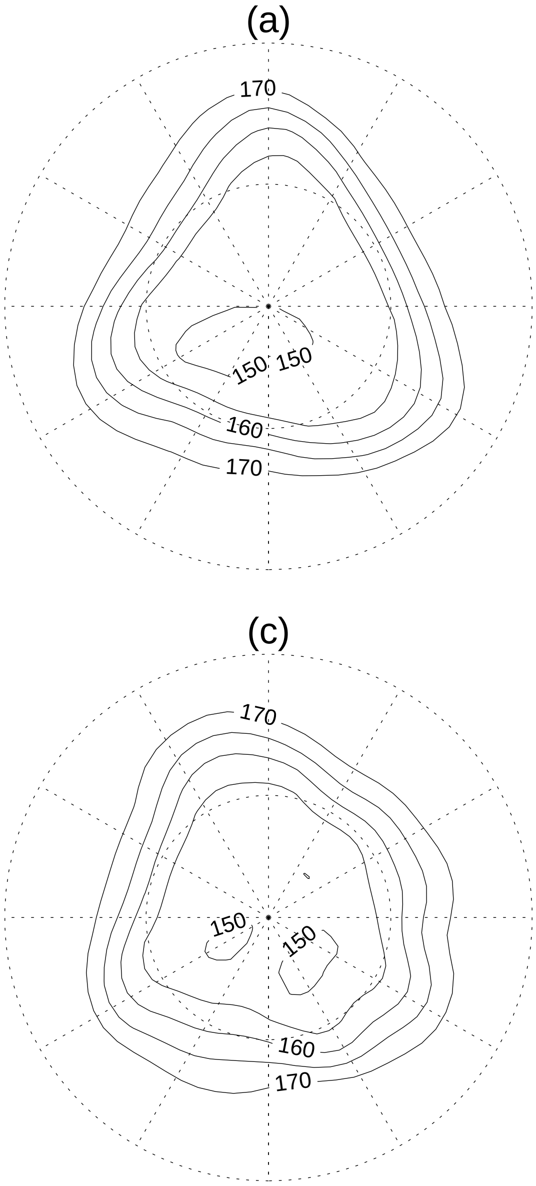
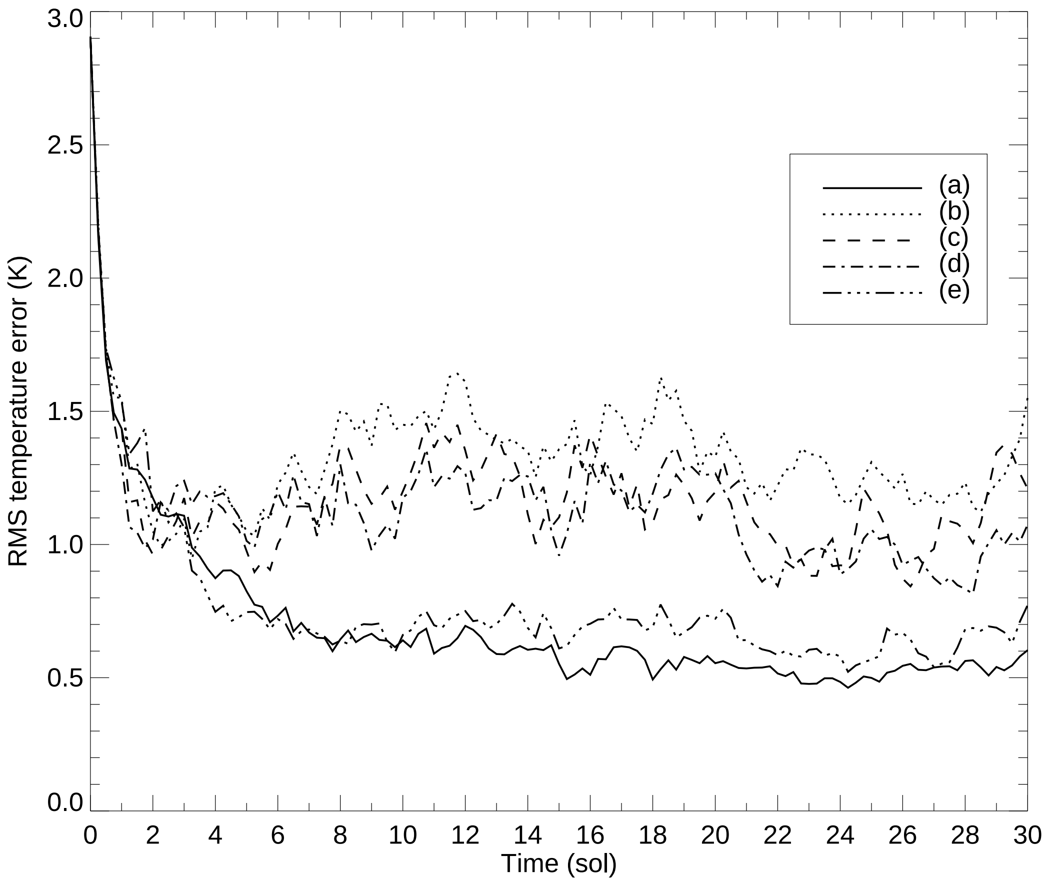
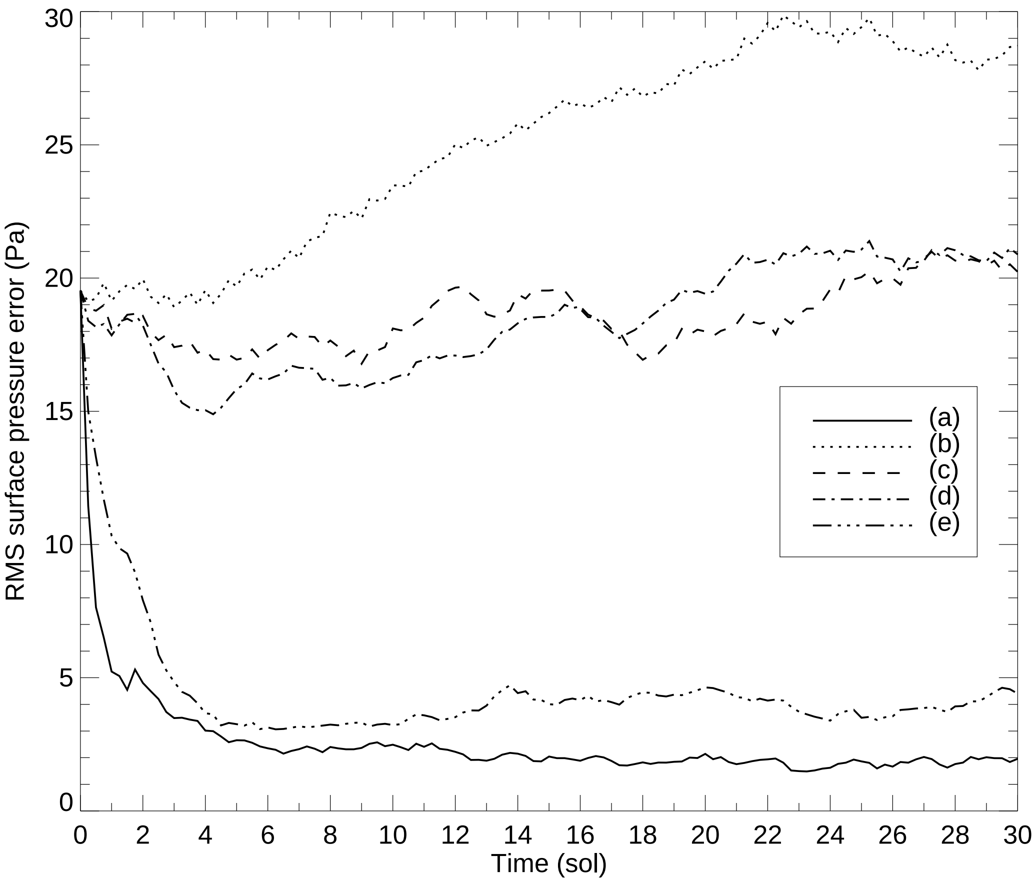

(a)
170
150
150
160
170
(b)
170
160
150
160
170
(c)
170
150
150
160
170
(d)
170
160
150
160
170

Fig. 4



3.0
2.5
2.0
1.5
1.0
0.5
0.0
RMS temperature error (K)
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Time (sol)
(a)
(b)
(c)
(d)
(e)

Fig. 5



RMS surface pressure error (Pa)
Time (sol)
0
5
10
15
20
25
30
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
(a)
(b)
(c)
(d)
(e)

Fig. 6

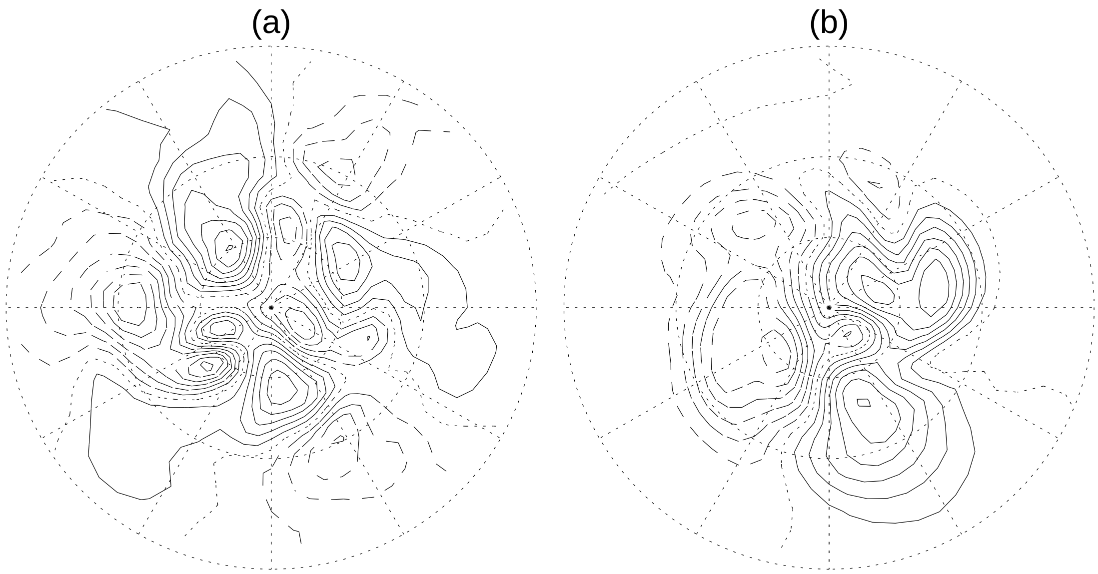


Fig. 7

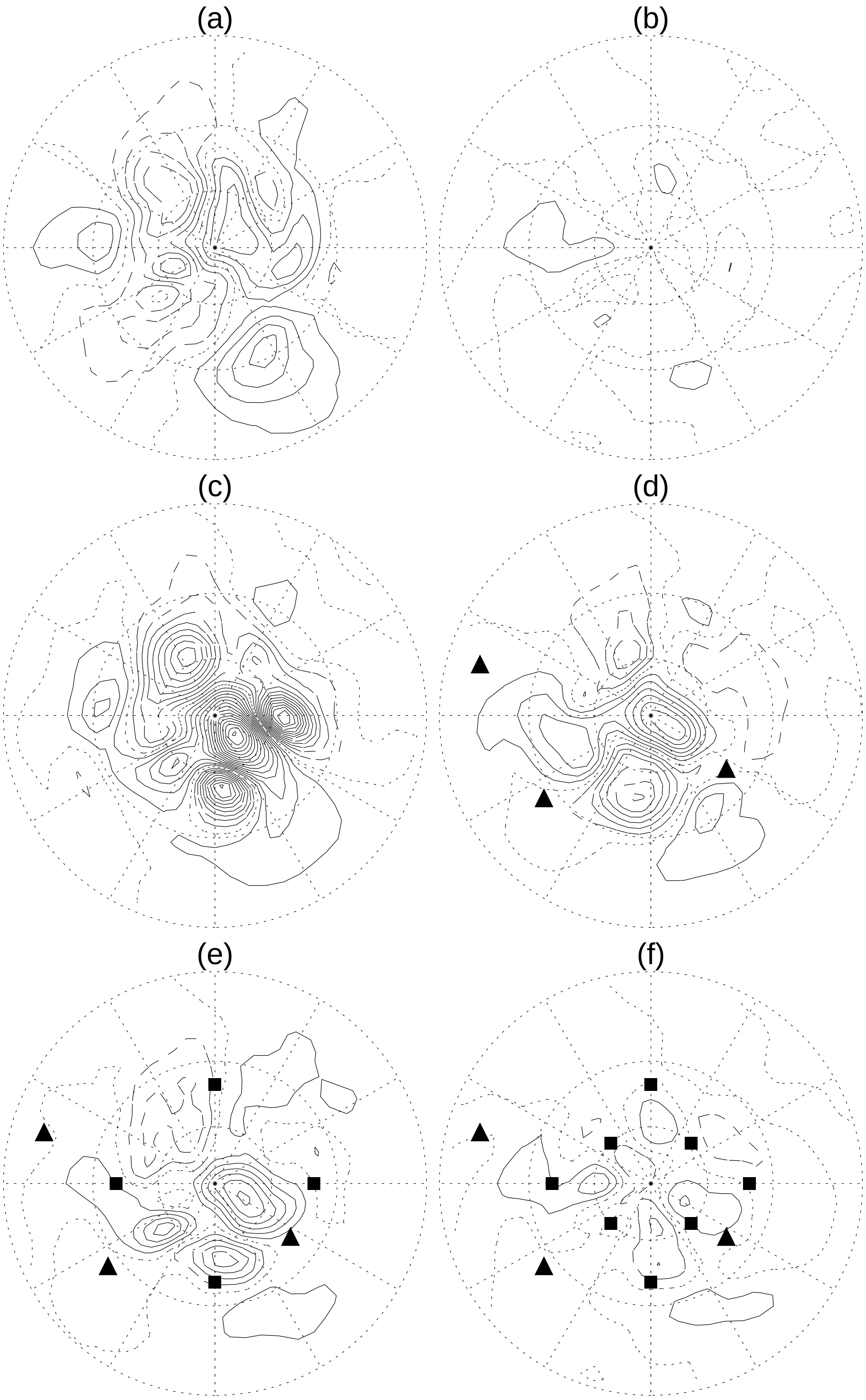


Fig. 8

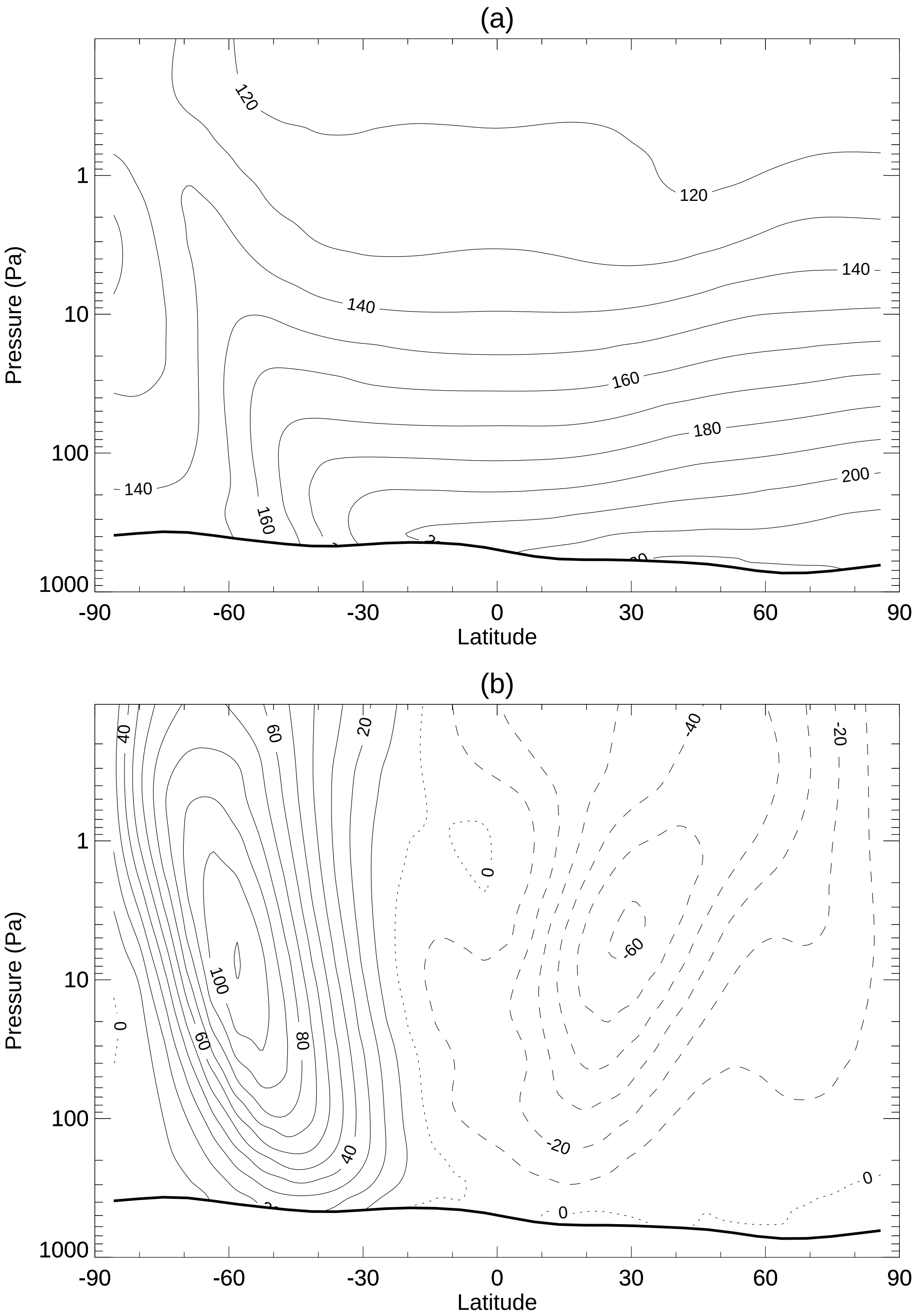

(a)
120
120
140
140
140
160
180
200
140
160
Pressure (Pa)
1
10
100
1000
-90
-60
-30
0
30
60
90
Latitude
(b)
40
60
20
-40
-20
0
-60
100
0
60
80
40
-20
0
0
Pressure (Pa)
1
10
100
1000
-90
-60
-30
0
30
60
90
Latitude

Fig. 9

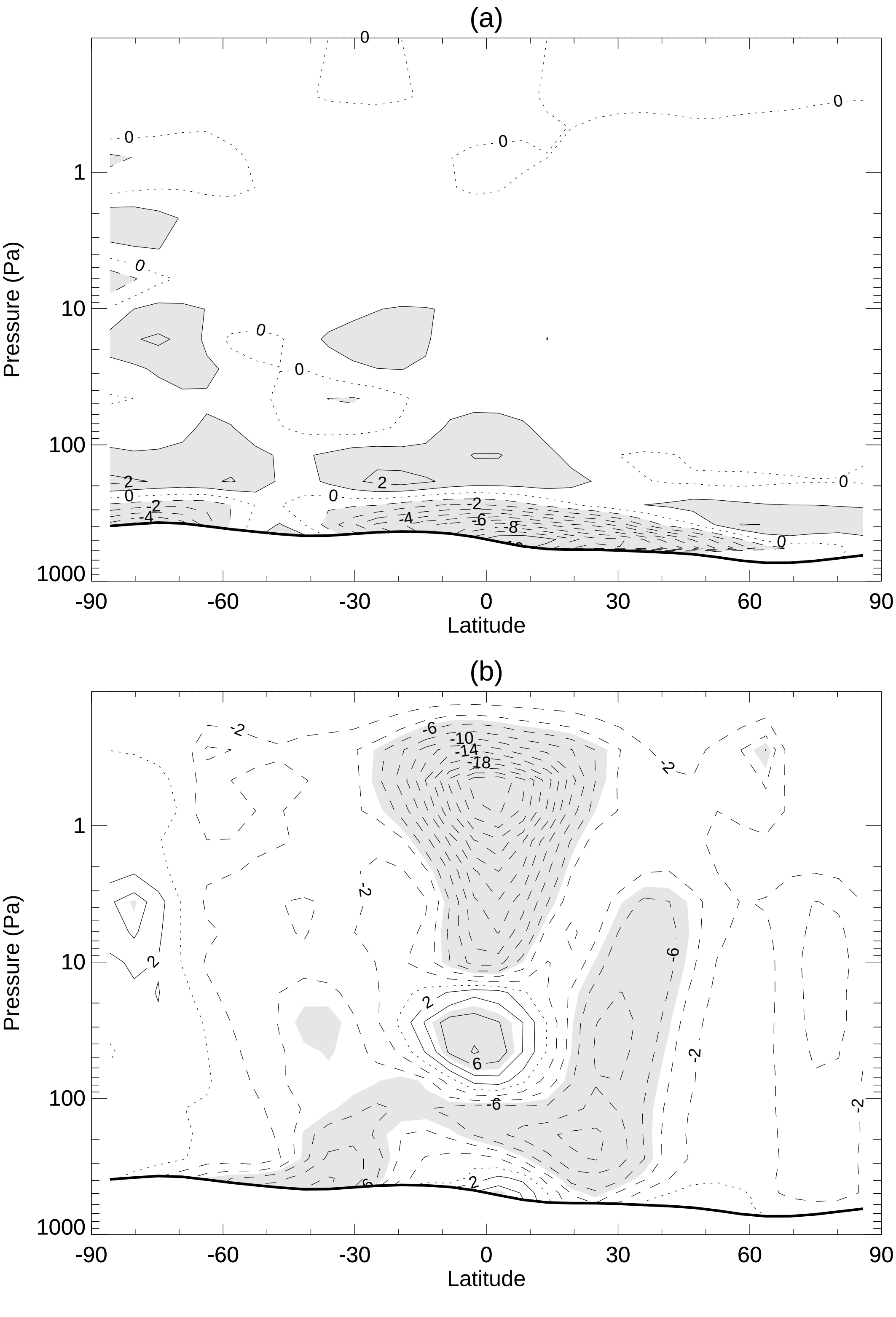

(a)
Pressure (Pa)
Latitude
0
2
-2
-4
-6
-8
1
10
100
1000
-90
-60
-30
0
30
60
90
(b)
-2
-6
-10
-14
-18
6
1
10
100
1000

Fig. 10

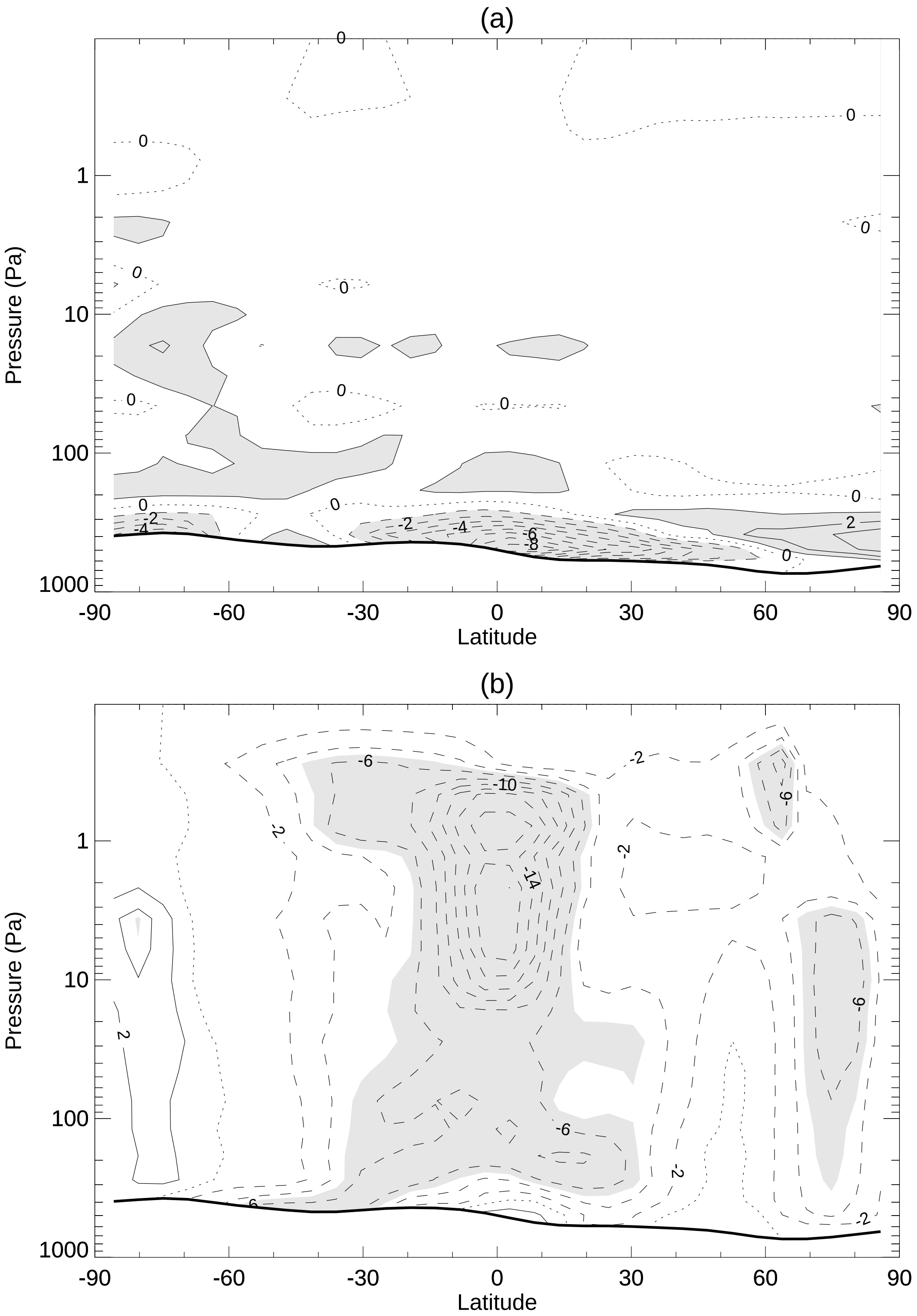


Fig. 11

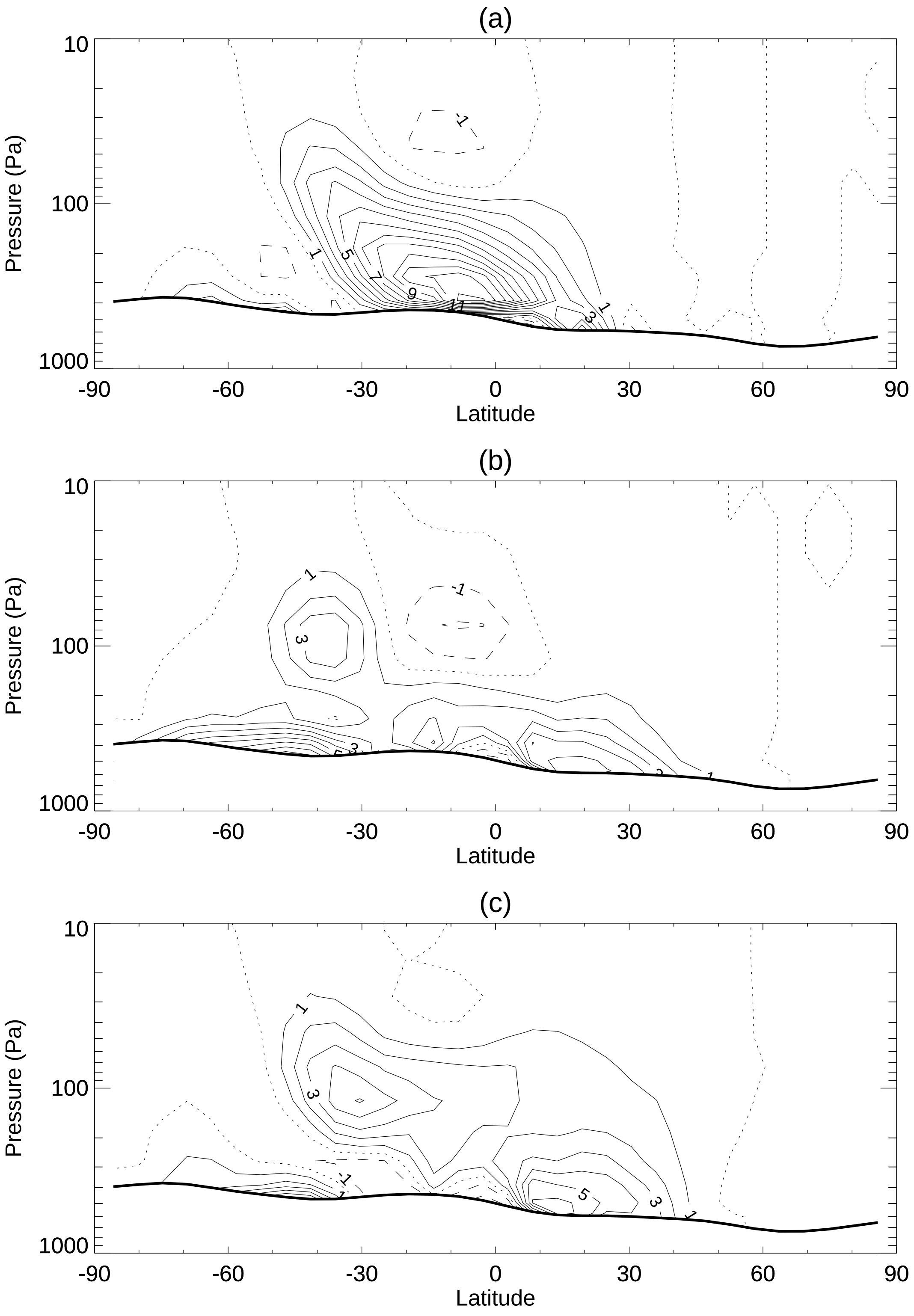

(a)
Pressure (Pa)
10
100
1000
-90
-60
-30
0
30
60
90
Latitude
-1
1
5
7
9
11
3
1
(b)
Pressure (Pa)
10
100
1000
-90
-60
-30
0
30
60
90
Latitude
1
-1
3
3
1
(c)
Pressure (Pa)
10
100
1000
-90
-60
-30
0
30
60
90
Latitude
1
3
-1
5
3
1

Fig. 12

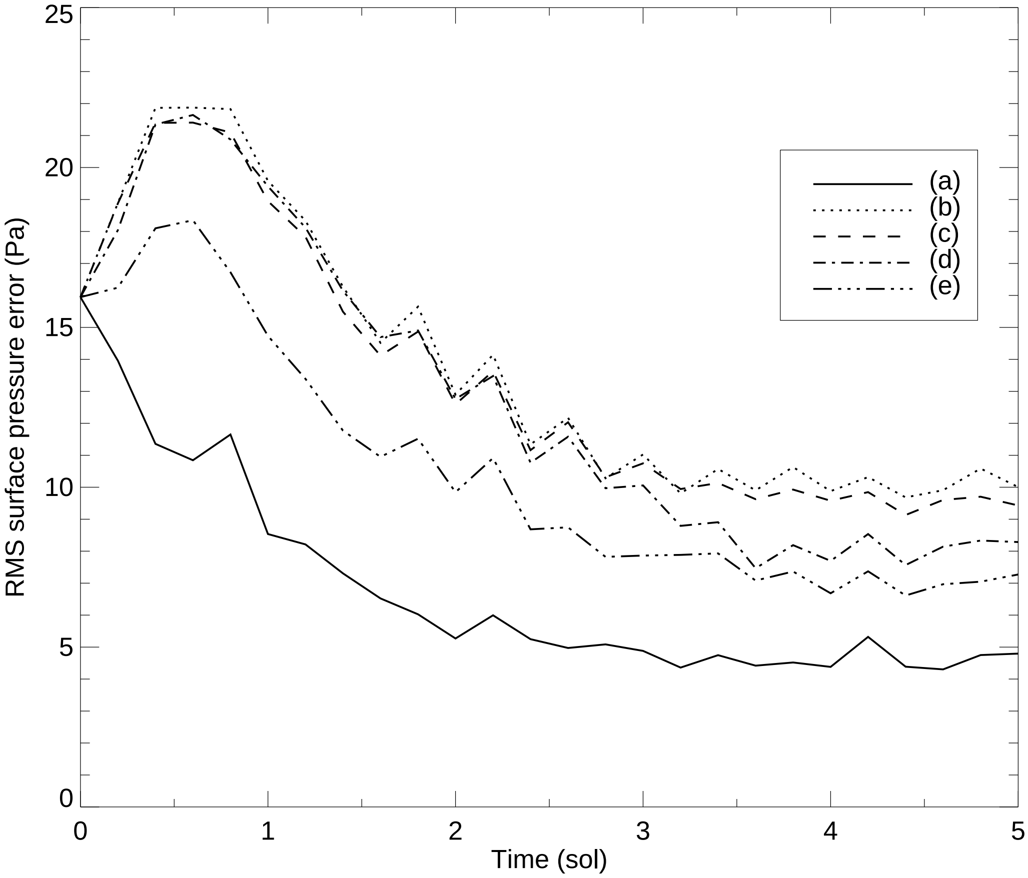

RMS surface pressure error (Pa)
Time (sol)
(a)
(b)
(c)
(d)
(e)
0
5
10
15
20
25
0
1
2
3
4
5

Fig. 13