Change in the Martian Atmosphere

Thesis by

David M. Kass

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted February 25, 1999)

Acknowledgements

The six and a half years I have spent here at Caltech seem to have passed in a blink of an eye. It seems like it was just yesterday that I got here and started learning about planetary science. Overall it has been a very exciting time. There was the disappointment at the loss of Mars Observer, then the excitement of Galileo and Mars Pathfinder and finally the Mars Global Surveyor. On top of the excitement of the missions have been the many discoveries and new results, both from the missions as well as from theoretical and other observational work. This includes the discovery of extra solar planets, the discovery of Kuiper belt objects and the possible discovery of fossils from Mars. While exciting, there have also been frustrating moments as well as those intense times where the amount of work seemed overwhelming. But overall, it has been an enjoyable, mind stretching time.

I would never have gone this direction without the encouragement of many individuals who helped convince me that I could go on and tackle science with a capital S. This includes Miss Pontius who encouraged me to look beyond the immediate situation and challenge myself. The University of Dayton Honors program, and especially Dr. Palermo and my fellow honors students helped me realize that there was more to university life than just classes. Prof. Lang and Prof. Smith in computer science who encouraged me to consider computing just a foundation.

Throughout my education, I have had the invaluable support of my parents and family (despite the occasional comments about a real paying job). Their understanding and acceptance of my decision to come to Caltech has been especially important. Their encouragement and good cheer from home have been invaluable in getting through many of the intense and frustrating moments.

While here, I have found how important a community is to doing scientific research. In addition to the general community of Caltech, the division and planetary science department has been an important part of successfully completing my thesis.

This includes insightful discussion of ideas and projects as well as the home away from home for Thanksgivings from my advisers Andy Ingersoll and Yuk Yung. Without them, neither of these projects would have been completed. Beyond that, there was invaluable input and discussions with all the faculty.

In addition to the faculty, the graduate students are a critical part of the community. Many are the nights I spent working, discussing, and going forward with the planetary science graduate students; as well as others in the division. I'd especially like to thank all of you who entered with me in 1992 and shared the joys and pains of Orals (as well as classes and research) together. One very important group are Bruce, Hari, Jim, Ashwin, Uta, Liz, Lori, Charlie, and Ben, all my officemates over the years.

Of course, no mention of support at Caltech would be complete without Mike, Kay, Irma and Tammie who keep everything running and allowed me to concentrate on research.

Beyond Caltech, there is the greater community of planetary science that has helped define and refine so many of these ideas. Special thanks to Janet Luhmann and Bob Johnson for those critical numbers that make it possible to actually model sputtering. The data assimilation procedure is built on the huge foundation put together by Andy, Don Banfield, and Christophe Dumas. Also, my thanks to John Pearl, Barney Conrath, Phil Christensen and the rest of the TES team for the TES observations without which the data assimilation would never have occurred. The other half of the assimilation process, the Mars GCM, was very kindly supplied by Bob Haberle along with lots of help in understanding how it works.

And, finally there are the many people I spent my few (precious) moments away from research with. This includes the entire Friday Social hour crowd over the years–a key to allowing me to relax at least one evening a week. Dave, Dave, Dave, Justin and Cy helped me escape with great days spent playing games. Lee Silver and Jason Saleeby showed me the charms of Hawaii.

Abstract

There are several lines of evidence that the atmosphere of Mars has significantly evolved over the history of the planet. Because Mars does not have a strong intrinsic magnetic field, the atmosphere is eroded by interactions with the solar wind. Conditions in the early solar-system significantly enhanced this loss, notably the component due to solar-wind induced sputtering away. Modeling indicates that, integrated over the last 3.5 billion years, 0.8 bars of CO_2 have been sputtered. This is accompanied by the loss of 50 m of water. The loss of CO_2 is a significant loss when compared to the estimates of the thickness of the early atmosphere. A simple model of the behavior of the atmospheric $\delta^{13}C$, based on the expected Martian carbon "cycle" constrains the size the current CO_2 reservoirs by putting the sputtering loss into the context of the evolution of the atmosphere. In order to balance the isotopic effects of the sputtering loss, it is necessary for there to be $\sim$ 100 mbars of CO_2 trapped in the planet. This is quite reasonable given the ability of the regolith to hold adsorbed CO_2.

Using a modified form of Optimal Interpolation, it is possible to assimilate Thermal Emission Spectrometer (TES) observations from the Mars Global Surveyor (MGS) spacecraft into the Ames Mars General Circulation Model (MGCM). The method is optimized for the assimilation of the irregular data obtained during the aerobraking phase of the mission. Based on 25 sols of data at $L_s \approx 200$, the assimilation process reveals several interesting features of the Martian atmosphere. The assimilation indicates that the lower atmosphere (up to $\sim$ 0.1 mbar) in the northern polar regions is very cold—probably at or close to the CO_2 condensation temperature. Furthermore, the data imply that the midlatitudinal westerly jets extend poleward of the indicated MGCM locations. In addition to correcting the phasing of the Northern baroclinic storm belt, the data indicate that the amplitude of the waves are stronger than expected and possibly with a lower zonal wavenumber. Thus the data and assimilation allows the MGCM to create the 1997 Martian Fall instead of a random Fall.

Contents

List of Figures

List of Tables

Chapter 1 Preface

Mars is the fourth planet outward from the sun. In many ways it is the planet the most similar to the Earth. It has a solid, rocky surface over a differentiated interior, covered by a thin atmosphere. The spin rate is almost identical to that of the Earth, giving it a very similar length of day. Furthermore, the obliquity, at least at present, is also similar to that of the Earth, so it has a comparable seasonal solar forcing. There is evidence for significant erosion as well as volcanic activity at some point in the planet's history.

While Mars is similar to the Earth, there are also significant differences. This is especially true when looking at the atmosphere of the planet. First, the orbit of Mars is elliptical. This causes large variations in the total solar forcing over the course of the Martian year (which, due to a semi-major axis of 1.5 AU is about twice the length of the Terrestrial year). Furthermore, many of its orbital parameters undergo large chaotic changes due to forcing from Jupiter and the other planets. Also, being further from the sun, coupled with the thin 7 mbar atmosphere, Mars is a colder planet. This results in several large effects. First, the cold temperatures cause the atmosphere to be very dry, which mirrors a very dessicated surface. The low temperatures also result in the freezing out of up to a third of the atmospheric CO_2 at the winter pole as a seasonal polar cap. But since CO_2 is the major constituent (95%) of the atmosphere, this causes large oscillations in the total pressure over the course of the Martian year.

The planet is significantly smaller than the Earth (a radius of 3390 km compared to 6380 km for the Earth). The smaller size results in a much lower escape energy, which makes it much easier for objects, including the atmosphere, to escape from the planet. A second effect of the smaller size is a faster cooling and smaller core size, resulting in the lack of a global magnetic field at present (and presumably for much of its history). This allows the solar wind to directly interact with the upper atmosphere creating a very different plasma environment than at the Earth.

One of the important features of the Martian atmosphere is that it is variable and changes on many different time scales. Fast changes include the diurnal cycle which causes the atmosphere to vary over the course of the day as the sun crosses the sky. There are even changes on shorter time scales caused by local phenomena such as dust devils and fogs. At slightly longer periods, there are weather systems which propagate around the planet. Then over the course of the year, there is the progression of the seasons, and accordingly a change in the response of the atmosphere. The inter-annual variability is driven by the irregular appearance and size of dust storms. Depending on the dust storm, it may cover part, or even all of the planet with optical depths reaching 5 or even higher. On geological time scales, the orbital parameters of Mars change. This results in variability on scales as short as 10^5 years with chaotic changes on even longer scales, ranging up to 10^7 years or longer. The changes in the orbital parameters force the atmosphere into different regimes. The atmosphere also changes on time scales as long as the age of the planet due to the loss and evolution of the atmosphere since Mars accreted. Thus change is fundamental to the behavior of the Martian atmosphere.

The available observations are not really sufficient to study the change in the atmosphere of Mars except in restricted cases. In order to be able to study the change in a broader context, I am relying on computational numerical models. This allows many of the processes to be described physically while covering large periods of variability. The numerical models are based on theoretical concepts of the processes involved, but are capable of looking at the evolution of the processes as well as steady-state issues. In order to actually make the model results relevant to Mars, it is necessary to use observations to constrain the models and to evaluate the model results. This requires a wide range of observations, ranging from photographic interpretation to spectral measurements, to in-situ measurements, to Terrestrial based measurements as well as spacecraft measurements.

In my investigation of the Martian atmosphere, I examine how the atmosphere changes on several different time scales. These range from the longest time scales for change to some of the shortest ones. The next four chapters can be grouped into

two pairs of chapters. Chapters 2 and 3 deal with the change and evolution of the atmosphere over the last 3.5 billion years. They concentrate on the processes that affect the atmosphere over the age of the solar system. Chapter 2 models a process called solar-wind induced atmospheric sputtering that is efficient at removing the bulk atmosphere to space where it is swept away. Chapter 3 then puts the process in the context of the evolution of the Martian atmosphere. It studies the various processes that have affected the atmosphere and how they may have acted together to create the atmosphere seen today.

Chapters 4 and 5 look at atmospheric change on daily to seasonal time scales. In order to study these time scales on a global basis, I combined an MGCM (Mars General Circulation Model) with actual observations of the atmosphere by the TES (Thermal Emission Spectrometer) on the MGS (Mars Global Surveyor) spacecraft. Chapter 4 describes the data assimilation technique developed to merge the data and observations and the testing to show how effective the process is. Chapter 4 then covers the assimilation of the actual data. The assimilation model results are used to investigate the weather in the MGCM as well as the mean seasonal behavior of the atmosphere.

Chapter 2 Sputtering Model

2.1 Introduction

There are several lines of evidence that indicate that the Martian atmosphere has undergone significant change over the history of the planet. The first is the Martian geomorphology, notably the channels, which appear to indicate that the planet once had significant quantities of water at or near the surface. This implication is also supported by the significant enrichment in deuterium (relative to the Earth) in the atmosphere, implying a more humid atmosphere. In order to get a humid atmosphere and water near the surface, it is necessary to warm the surface from its current mean temperature of $\sim$ 220K. The easiest way of doing this is with a thicker atmosphere and an atmospheric greenhouse effect. The fractionation of several other stable isotopes also support the idea of a thick atmosphere during the early history of Mars.

Although there are questions about how much a greenhouse effect could raise the temperature [*Kasting,* 1991, *Haberle,* 1998], most current models require an atmosphere of at least 0.5 bar of CO_2 in order for liquid water to be near the surface (for example, McKay and Davis [1991], Kieffer *et al.* [1992], or Haberle [1998]). Because the current Martian atmosphere has only 7 mbar of CO_2 and only a small amount of H_2O (10 pr-μm, or if it were all precipitated out, a layer 10 μm deep [*Carr,* 1996, *Kieffer et al.,* 1992]), an important question is the fate of the early water and CO_2.

There are two major possibilities: Either the early atmosphere is sequestered somewhere in the planet [*Fanale et al.,* 1982] or it has been lost to space. McElroy [1972] pointed out that the water loss might have been determined by dissociative recombination of $O_2^+ + e$, because H loss by Jeans escape is easily accomplished. However, neither Jeans escape, nor dissociative recombination, is capable of removing significant amounts of CO_2 over the lifetime of the planet's "contemporary" atmosphere.

One alternative mechanism for loosing the atmosphere to space is atmospheric sputtering (erosion by impact of energetic atoms and ions). Initial investigations concentrated on sputtering by solar wind particles [*Tombrello,* 1982, *Watson,* 1980]. While capable of removing H, the solar wind particles are estimated to only sputter the equivalent of the current atmosphere of CO_2 over the age of Mars. The mass difference between the solar wind particles and the bulk atmosphere makes momentum and energy transfers very inefficient. Energetic exospheric O^+ ions were considered a possible source of heavier particles [*Tombrello,* 1982, *Watson,* 1980], but the actual fluxes where not determined until Luhmann and colleagues [*Luhmann and Kozyra*, 1991, *Luhmann et al.,* 1992] modeled them. This process is called indirect solar wind induced sputtering, and it is much more effective than direct sputtering at removing CO_2 and other heavy species.

2.2 Theory

Indirect solar-wind induced sputtering occurs on planets that do not have a significant intrinsic magnetic field. In this case, the solar wind is not stopped by the magnetic field and interacts directly with the upper atmosphere. Sputtering can be broken into two steps (see Figure 2.1). In the first step, the solar wind interacts with the planetary ionosphere. The Martian ionosphere is predominantly populated by O^+ created primarily through dissociation by extreme ultra-violet (EUV) solar radiation. The interplanetary magnetic field, carried along by the solar wind is draped over Mars and penetrates into the ionosphere since Mars does not have a strong intrinsic field. Through plasma interactions, the solar wind and solar EUV transfer energy to the ionosphere. The draping of the interplanetary magnetic field results in a complex field geometry that, due to temporal and spatial variability, further accelerates the ions. These particles then follow helical trajectories defined by the interplanetary magnetic field. Some of the trajectories miss the rest of the atmosphere and the ion is swept away by the solar wind as a pickup ion. But some of the particles impact the upper atmosphere after being accelerated by the combined solar wind and interplanetary

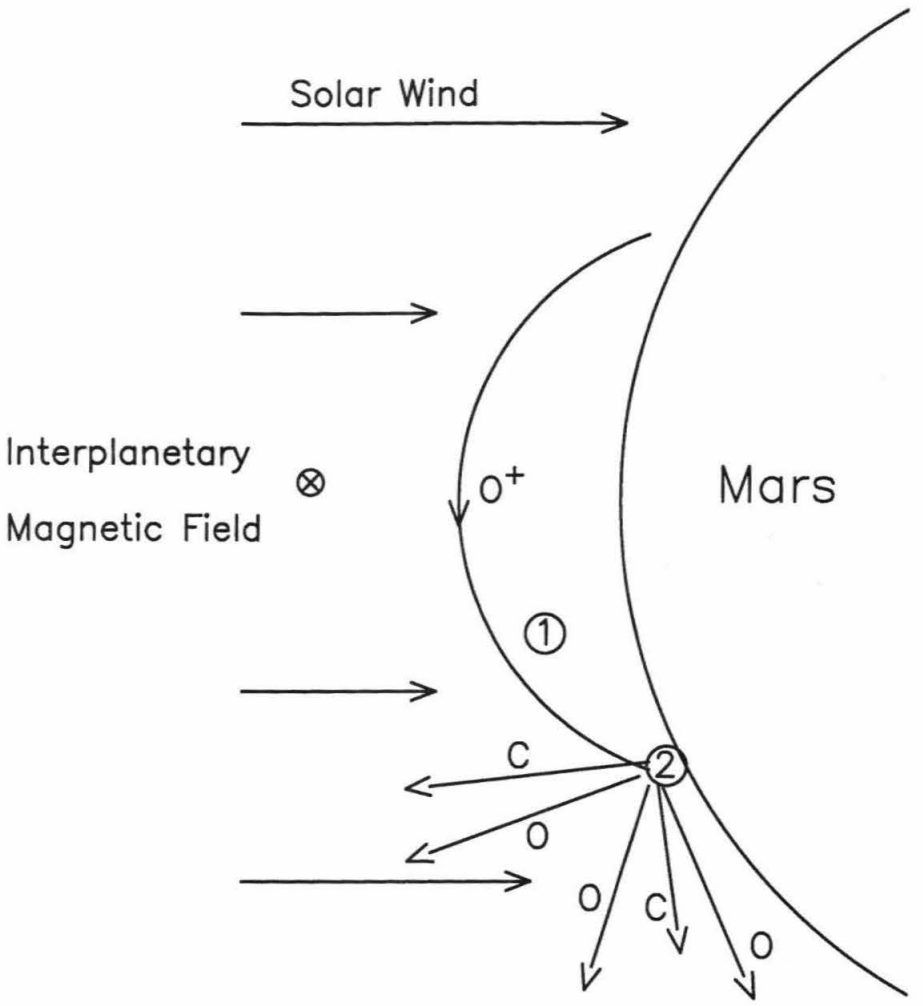

Figure 2.1: Sputtering Process
In step 1, ions are accelerated by the interactions of the ionosphere, solar EUV and solar wind. In step 2, the ions impact the upper atmosphere and collisionally transfer energy and momentum, ejecting part of the atmosphere.

magnetic field. These particles will have a significant amount of energy (around 1 keV total) compared to the escape energy (1.5 eV for a carbon atom).

When these accelerated ions impact the upper atmosphere, they collisionally transfer energy and momentum to the neutral component and can give other atoms sufficient energy (and an upwards velocity) to escape. This is the second step of the sputtering process (and the actual sputtering). A useful way of considering the process is that each impacting ion will cause a certain number of other particles to escape the planet. This is the sputtering yield. Thus in order to calculate the flux of particles lost by this type of sputtering, it is necessary to calculate the flux of impacting particles and the sputtering yield.

In their work, Luhmann *et al.* [1992] used an analytical model to determine the sputtering yield for the second part of the process. Their work indicated that the sputtering could remove a small amount of atmosphere (about .14 bars of CO_2), but was not capable of removing sufficient atmosphere to create the greenhouse to warm the planet. I then created a Monte-Carlo model of the sputtering process itself to calculate the yield. Combining this with the impacting fluxes calculated by Luhmann *et al.* [1992] allows the total amount of atmosphere lost to be estimated.

2.3 Model

The Monte-Carlo model operates by firing individual ions at the atmosphere and then determining what happens in the atmosphere. As each ion impacts the atmosphere, the model determines the number of particles that leave the top of the atmosphere with sufficient energy to escape. By analyzing the results of a large number of ions impacting the atmosphere, the model can calculate the mean number of atoms that are lost: the sputtering yield.

The sputtering yield is modeled at three different epochs, the present (or 1 EUV), 2.5 Ga ago (3 EUV), and 3.5 Ga ago (6 EUV). Each epoch is defined by the strength of the extreme ultraviolet (EUV) flux as a multiple of the current value. These epochs correspond to the times for which Luhmann *et al.* [1992] have calculated the

impacting fluxes. For each epoch, the calculation uses a model neutral atmosphere as the target of the ions. These atmospheres can be divided into three parts, a homosphere, a diffusive region, and an exosphere. The homosphere is the part of the model where the atmosphere is well mixed, and ends at the homopause, above which each species assumes its own mass dependent scale height. The exosphere (which starts at the exobase) is the region of the atmosphere where the particles can escape without colliding. Both the homopause and exobase are more zones than sharp layers and actually extend over several kilometers of altitude; this is especially true when comparing the various species.

The atmosphere for the present is from Nair *et al.* [1994] (see Figure 2.2) and was expanded by isostatically calculating an exosphere for each species. Argon was also added to the atmosphere by assuming a constant mixing ratio up to the homopause and then letting it assume its mass dependent scale height. The model has O, CO_2, CO, N_2, N, H_2, H, and Ar in the atmosphere. It has a 2 km vertical resolution. While the isostatic assumption is fairly good for most of the species, it probably fails for H_2 above 300 km due to dissociation and loss to space. The H_2 in this region does not significantly impact the results of interest, so the uncertainty is not a problem. The atmospheres at 3 EUV and 6 EUV are from Zhang *et al.* [1993] and have also been extended vertically. They have only CO_2 and O with a resolution of 10 km. Since the sputtering occurs at very low number densities (basically at and slightly below the exobase), none of the atmospheres actually reach the surface. The structures are controlled by the local temperatures and by photochemistry and are very insensitive to the total atmospheric pressure. This makes the results independent of the actual atmospheric pressure (at least for the major species); the actual altitude of the processes will change, but the pressures will remain constant. All three atmospheres are assumed to be plane parallel and all of the particles are assumed to be at rest. These assumptions greatly simplify the calculations for the individual collisions and appear to be reasonable for the regimes of interest.

The Model fires 10,000 ions with an impact angle of $\sim$ 124 degrees and an energy of 1 keV, the mean parameters derived by Luhmann and Kozyra [1991] for O^+ ions

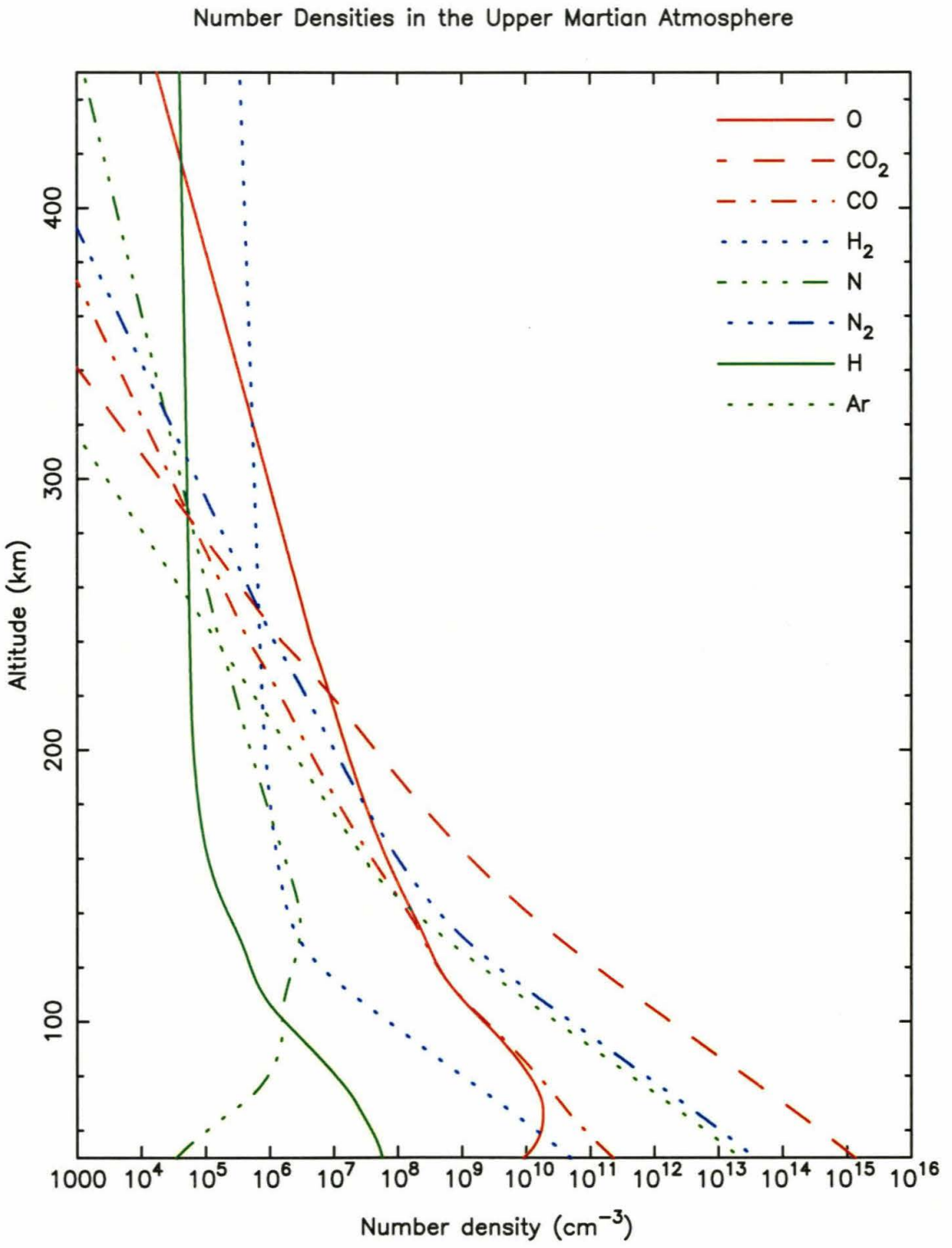

Figure 2.2: Present (1 EUV) Atmosphere Model
Number densities for the various species in the present atmosphere [*Nair et al.*, 1994]. This has been extended isostatically.

impacting the upper atmosphere of Mars. For each impacting particle, the model determines the altitude and target of its first collision. After resolving that collision, it puts all of the resulting particles in a queue. The model then pulls each particle out of the queue one at a time and has it collide with its next target. The results are again put back into the queue. If a particle reaches the top of the model atmosphere ($\sim$500 km) with sufficient energy to escape the planet, it is counted as an escaped particle. If it doesn't have enough energy, the particle is counted as being on a ballistic trajectory and is ignored. Particles that have less than 1.5 eV are also removed from the calculation. They have too little energy to allow an atom of interest here to escape, and their removal significantly speeds up the model execution. The calculation of a collision and the storing of its products continues until the queue is empty. At that point, all the statistics for that impacting particle are tallied and the next particle is started.

For each particle's flight and collision(s), the process is divided into two parameterized steps. The first step determines where the collision occurs and what molecule or atom is the target. Then a second step determines the results of the collision. The important part of the results are the direction and kinetic energy of each particle (atom or molecule) resulting from the collision.

In determining the location and target for the next collision, the model uses a table of energy dependent cross sections. Each cross section (see Figures 2.3 through 2.7) measures the effective area of each atom or molecule in the atmosphere. The cross sections depend on the two species potentially involved in the collision as well as the energy of the impacting particle. The energies are measured in the atmospheric frame of reference. For the case of two atoms colliding, the cross section is a single value (depending on the energy) and if the moving atom passes within the given area surrounding the target atom, they collide. For poly-atomic species, the cross section is a bit more complex (for example, Figure 2.5) since the molecule can dissociate, and possibly do so in multiple patterns. There is a total cross section (the solid line) representing the area where some type of collision between the two species will occur. This is then broken into individual cross sections for each possible dissociation path.

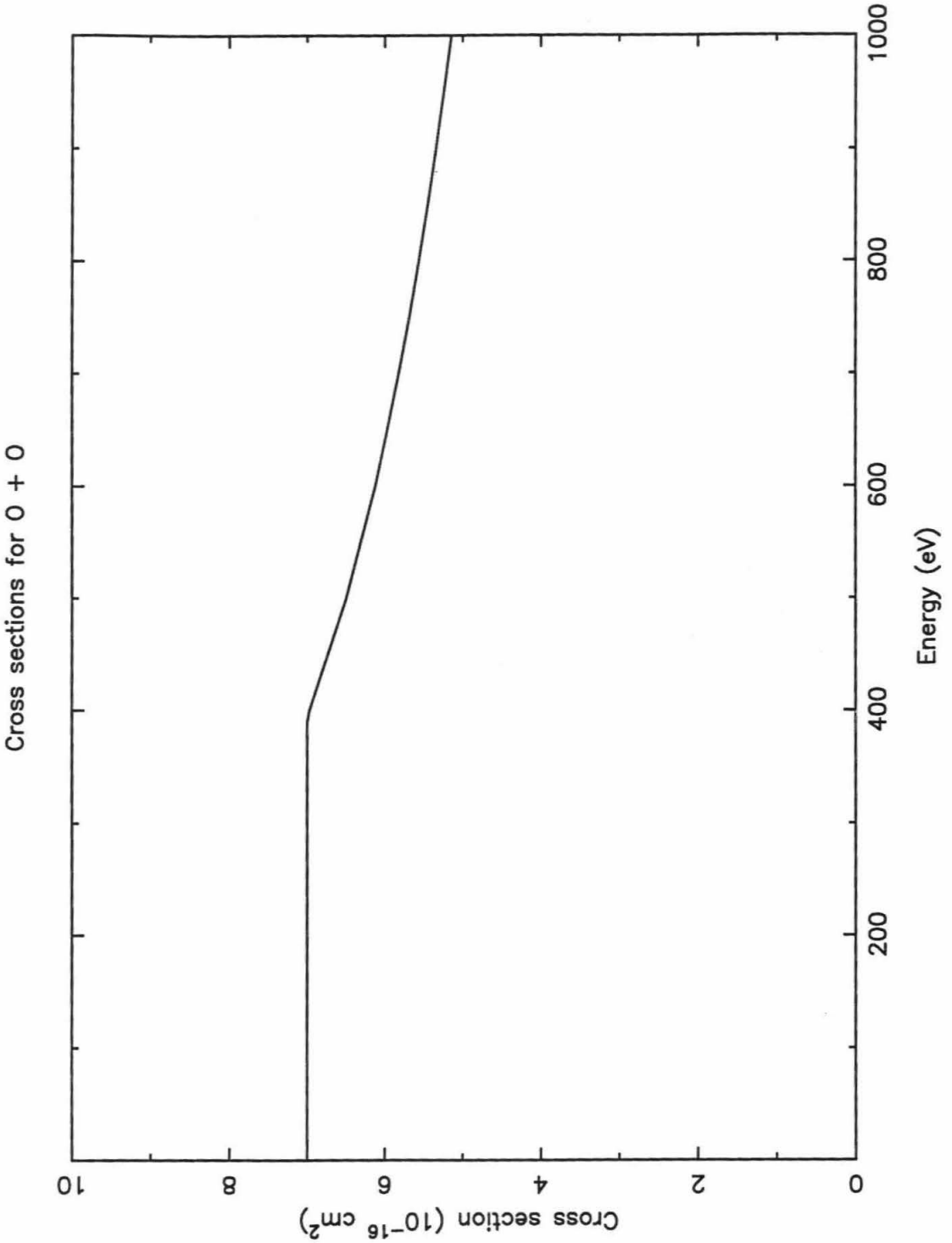

Figure 2.3: O + O Cross Section

[*Ishimoto et al.*, 1986]

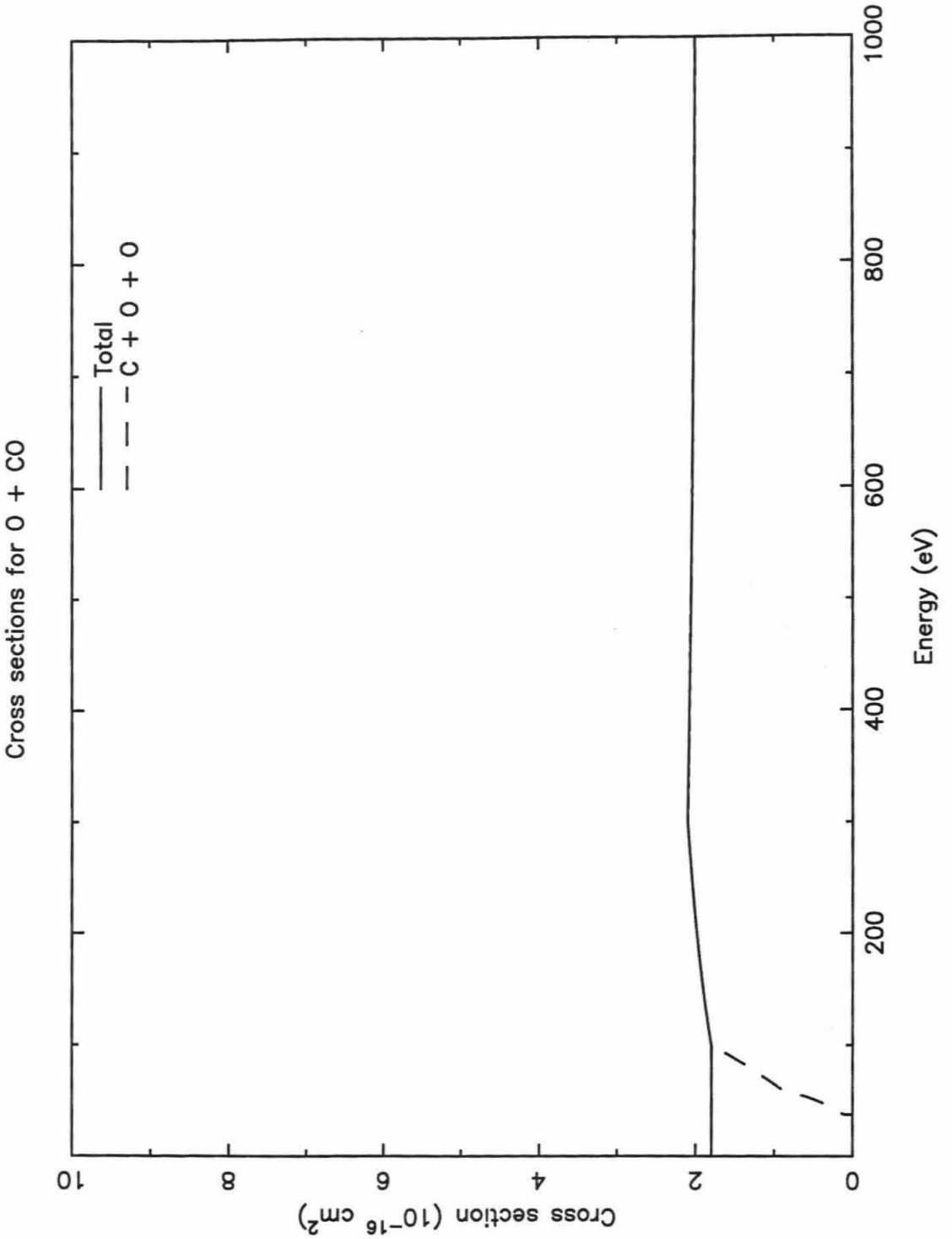

Figure 2.4: O + CO Cross Section
[*Johnson and Liu,* 1998]

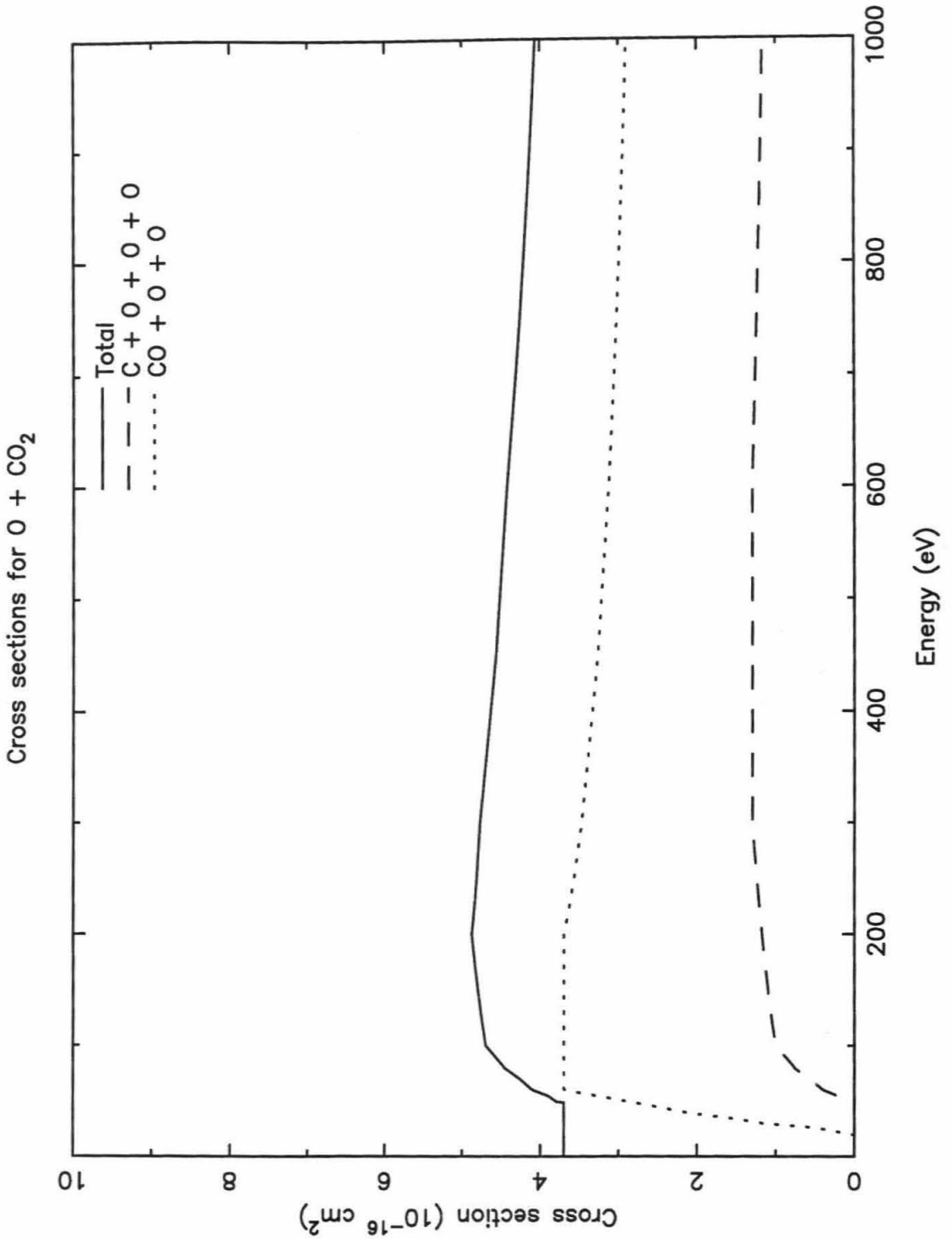

Figure 2.5: O + CO_2 Cross Section
[*Johnson and Liu*, 1998]

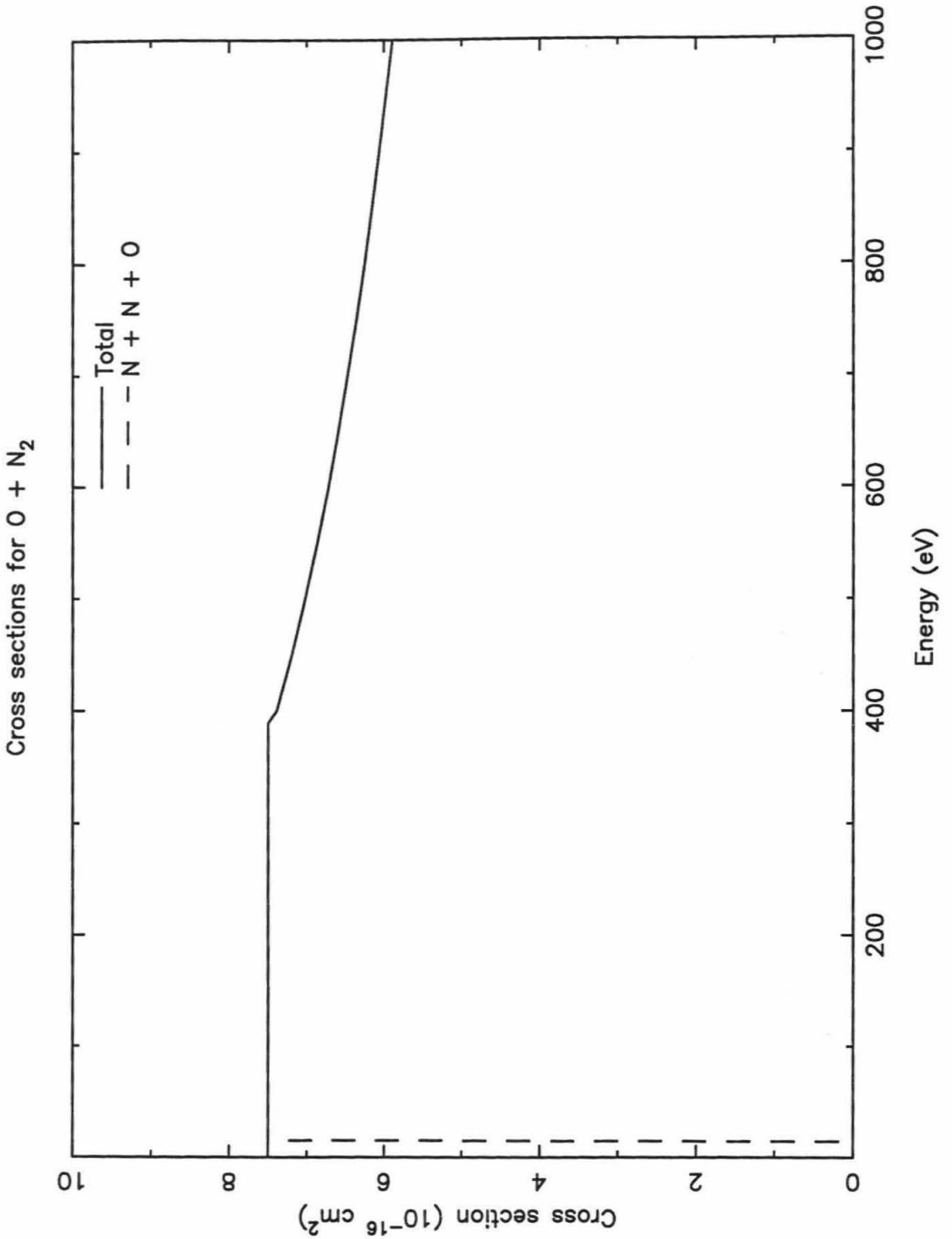

Figure 2.6: O + N_2 Cross Section

[*Ishimoto et al.*, 1986]

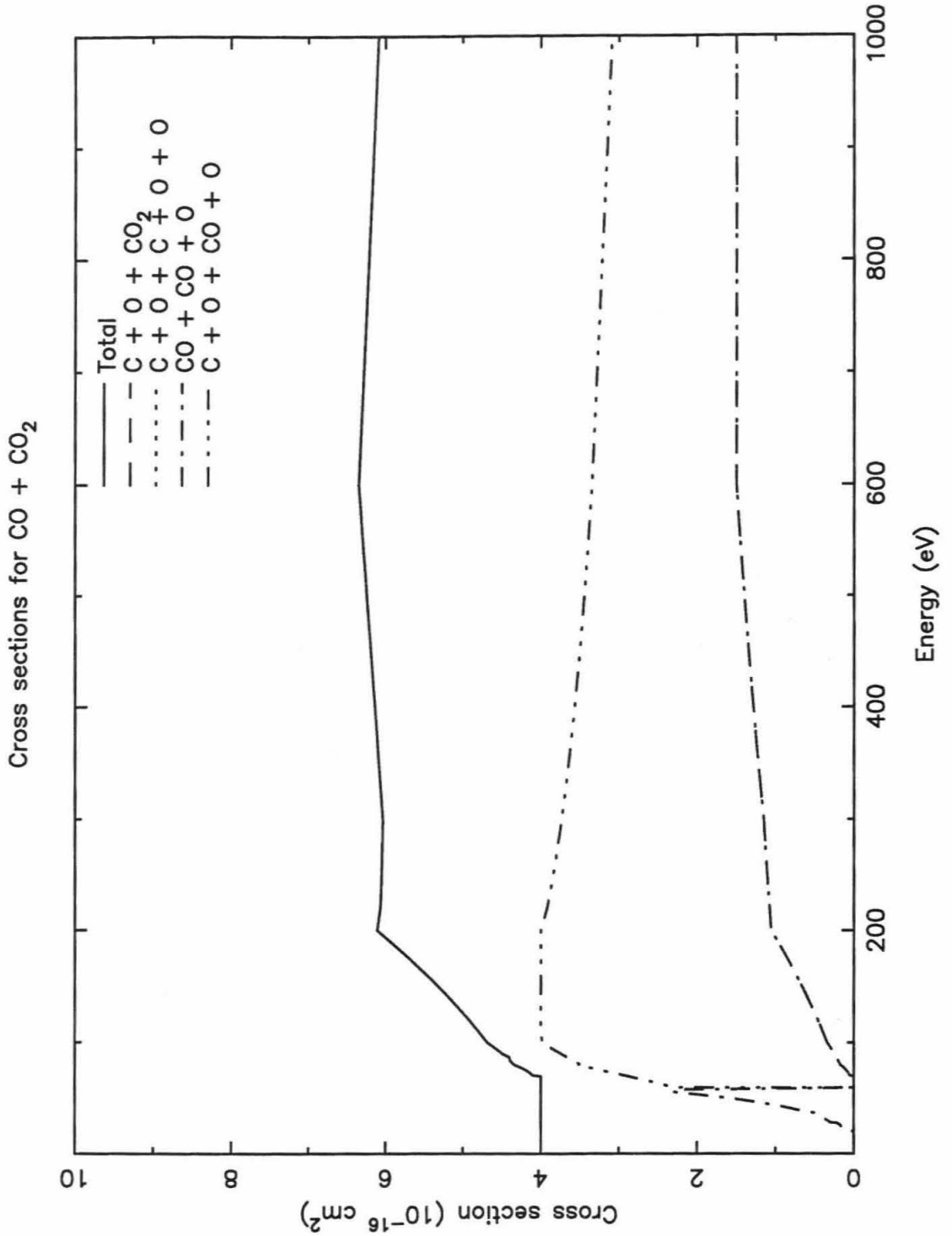

Figure 2.7: CO + CO_2 Cross Section
[*Johnson and Liu,* 1998]

Thus CO_2 can dissociate into CO + O or into C + 2O, and each possibility has its own cross section. The case where the CO_2 does not dissociate is the remainder when all the dissociation paths are subtracted from the total cross section. The cross sections used in the model come from many different sources. The O + O, O + N_2, and N_2 + N_2 cross sections are from Ishimoto *et al.* [1986]. The O + H cross section at low energies from Hodges [1993] was extrapolated out to 1 keV. O + O^+ and O + H^+ cross sections are from Stebbings *et al.* [1964]. The cross sections for H + H^+ came from an extrapolation of the literature survey of Smith and Bewtra [1978]. Many of the ones involving CO_2 (O + CO_2, O + CO, CO + CO_2 and CO_2 + CO_2), which are among the most important cross sections to the model, are from molecular dynamical calculations by Johnson and Liu [1998]. Finally, where no experimental or laboratory values were available, the cross section was calculated by taking the molecular and atomic radii of the species [*Forsythe,* 1954, *Kaye and Laby,* 1995] and using $4\pi(r_a + r_b)^2$, where r_a and r_b are the radii of the two species involved. For cases where one of the species can dissociate, it was assumed to dissociate as long as the center of mass energy exceeded the dissociation energy. Excluding CO, the current set of cross sections do not produce fast poly-atomic species, and thus many collisions involving fast poly-atomic moving particles are ignored (for example, there is no case for N_2 + N_2 yielding 4 N).

It is convenient to define an integrated cross section ($\sigma^i = N \times \sigma(e)$, where N is the local number density, and $\sigma(e)$ is the energy dependent cross section). This can be done both for the total cross sections as well as for the cross sections for each set of products. Summing the total integrated cross sections gives the integrated atmospheric cross section for the layer ($\sigma = \sum_{\text{species}} \sigma^i_{\text{total}}$).

Determining where and what the target will be is done by first randomly choosing the path length ($\tau = -\log(rnd)$ where $0 < rnd < 1$ is a random number) to the next collision for that particle. Since $\tau = l \times \sigma$, where l is the distance traveled, the actual location of the collision can be determined using the integrated atmospheric cross section for each layer of the atmosphere. The angle of the trajectory (with respect to the vertical—no need for a full three–dimensional direction) is used to determine

the effective thickness of each layer. If the particle exits the top of the top layer of the model atmosphere, it is assumed to be on a free trajectory and counted as either escaping or as following a ballistic trajectory (depending on the amount of energy it has). As the particle travels through the atmosphere, it looses energy by exciting the surrounding atmosphere. The energy lost is calculated using the curves from Johnson and Liu [1998] (see their Figure 6, curves for S_e^{O} and $S_e^{CO_2}$). Once the model knows where the collision occurs, it determines the species of the target. This is done by randomly selecting a target species and a resulting set of particles, each possibility weighted by the individual integrated cross section.

With the location and target species for the collision, the model determines the results of the collision. The model transfers the collision to the center of mass frame of reference (CM), to simplify the algebra. This is a simple transformation since the target is assumed to be at rest. With the temperatures and particle sizes of interest in the model, the target will have less than 10% of the energy of the impactor in all collisions of interest and thus the assumption that it is stationary is reasonable. The model uses a coordinate system (see Figure 2.8) that has the +x axis in the direction of the impacting particle, the +z in the upward direction in plane formed by the x axis and the atmospheric vertical. In the CM frame, the model calculates the total energy. If one (or both) of the particles will dissociate, the energy required to dissociate is subtracted from the total energy available for the collision.

The model uses the conservation of energy and of momentum to determine the results of the collision. Since there are three unknowns per resulting particle (the model uses the three vector components of the velocity) and only four conservation equations, the model uses a Monte-Carlo method to solve the system. For each additional parameter needed, the model selects the value randomly using an appropriate distribution. The exact parameters set randomly are determined by the number of particles resulting from the collision. For simplicity, physical relevance and available data, most of the parameters selected randomly are the scattering or phase angles. For each particle resulting from a collision, there are two relevant angles. The first is θ (or χ in the CM frame), the phase angle (see Figure 2.8). It is measured from the

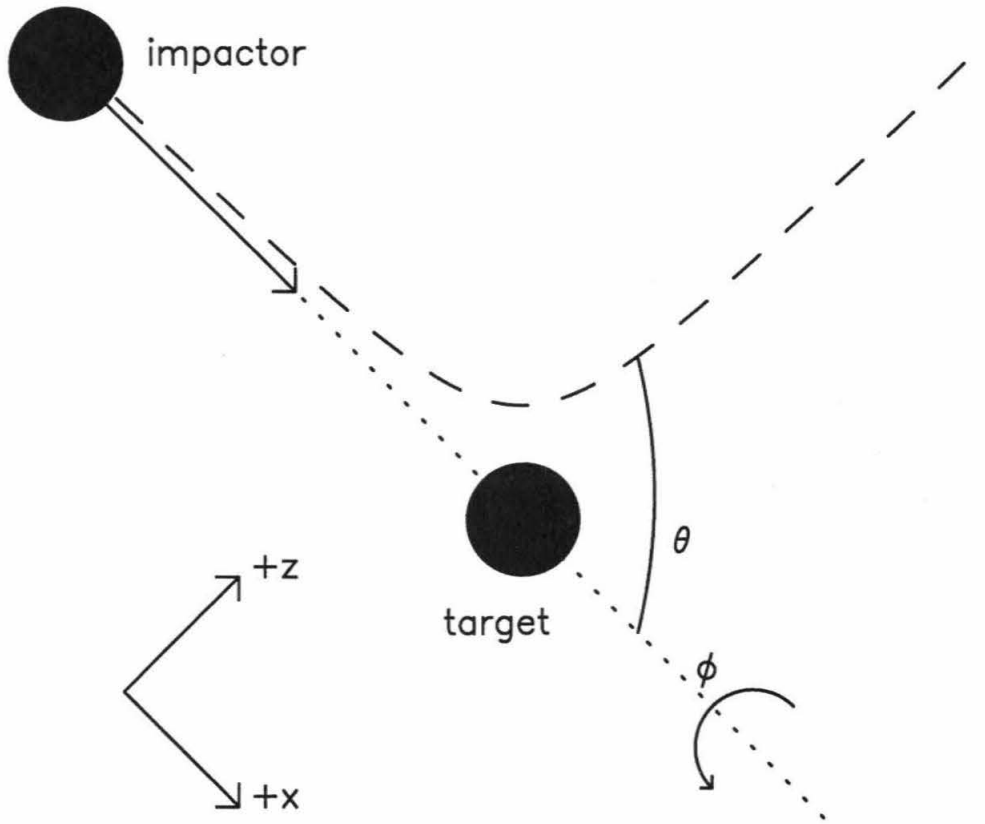

Figure 2.8: Coordinate System for Collisions
Reference coordinate system for collisional calculation. The indicated axes are the ones for the center of mass reference system. θ is the phase (or scattering) angle. It becomes χ in the CM frame. ϕ is the azimuthal scattering angle.

direction of the incident particle (due to the simplicity of the algebra in a coordinate system with that direction as one of the axes). The second angle is ϕ, the azimuthal angle, measured between 0 and 2π. Again for convenience, it is measured from the + z axis. ϕ is always selected isotropically since the collision is symmetric around the impact direction. The model always uses χ since all of the calculations are done in the CM frame. χ depends on the species involved in the collision, the specific particle in question and, for some cases, the energy of the impactor (measured in the atmospheric frame of reference). For computational simplicity, the phase functions are expressed as Henyey–Greenstein (HG) functions:

$$hg(g,\chi) = 0.5\left(\frac{1-g^2}{1+g^2-2g\cos(\chi)}\right) \tag{2.1}$$

where $-1 < g < 1$ is the parameter defining the type of scattering. Negative values represent backscattering (at least in the CM frame); positive ones are forward scattering and zero indicates isotropic scattering. In order to improve some of the scattering functions used, they were fitted with double Henyey–Greenstein functions [*Goody and Yung,* 1989]:

$$dhg(a,g_1,g_2,\chi) = a\,hg(g_1,\chi) + (1-a)\,hg(g_2,\chi) \tag{2.2}$$

with three parameters. Many of the values used in the model (Table 2.1) come from Johnson and Liu ([1998], and personal communication). The curves for O + O come from Johnson *et al.* [1999]. The curve for O + H comes from Gurwell and Yung [1993]. All of the cases not listed are either single HG functions with $g = 0.5$ for collisions without dissociation or isotropic for cases with dissociation.

Because of the method used to represent the collisions, they can be classified based on the number of resulting particles. For collisions resulting in two particles (i.e., neither dissociates), only two angles are needed. Furthermore, the collision can be performed in a plane (the one containing both outgoing particles). Thus in this case, the first angle is the phase angle of the impacting particle (chosen by

Scattering Functions

		Henyey–Greenstein parameters			
		Single	Double		
		g	a	g_1	g_2
O + O	60 eV		0.04	0.25	0.99
	200 eV		0.015	0.25	0.99
	600 eV		0.003	0.25	0.99
O + H			0.4	0.18	0.95
O + CO_2					
→ O + CO_2					
	60 eV	0.72			
	200 eV	0.85			
	600 eV	0.85			
→ O + O + CO					
O*	60 eV	−0.23			
	200 eV	0.25			
	600 eV	0.54			
O†	60 eV	0.2			
	200 eV		0.52	−0.43	−0.02
	600 eV		0.16	−0.93	−0.5
CO	60 eV	−0.37			
	200 eV	−0.74			
	600 eV	−0.9			
→ O + 2 O + C					
O*	60 eV	−0.48			
	200 eV	0.07			
	600 eV	0.49			
O	60 eV	0.2			
	200 eV		0.52	−0.43	−0.02
	600 eV		0.16	−0.93	−0.5
C	60 eV	0.51			
	200 eV	−0.02			
	600 eV	−0.35			

* Function for impacting fast molecule

† Due to this resulting in 3 particles, the scattering angle is not determined for this O atom.

Table 2.1: Selected Scattering Functions

The Henyey–Greenstein parameters (either for a single HG function or for a double HG function) used for the indicated collisions [*Johnson and Liu,* 1998, *Johnson et al.,* 1999, *Gurwell and Yung,* 1993].

convention). The second angle is the azimuthal angle of the collision plane. In a collision that results in three particles, five angles are needed. For simplicity, the model randomly determines all three azimuthal angles and the phase angle for two of the fragments. Usually they are the two fragments from the target (if an impacting CO dissociates, the fragment whose phase angle is determined is random). In the case where the collision results in four particles, the phase and azimuthal angles for each particle are determined randomly. In cases resulting in five particles (which can only occur in the model when CO impacts CO_2 and both dissociate completely), both angles are determined for each fragment. Furthermore, the energy of one C atom is selected using the probability distribution from Johnson and Liu [1998]. In this case the equations are actually quadratic and may occasionally need to be calculated with several sets of random parameters to actually get a physical solution.

Once the model uses the random variables and the conservation laws to calculate the velocity of each fragment in the CM frame, particle's velocity is transformed into the atmospheric frame and the energy and direction of the particle is stored. Each particle is evaluated to determine if it can still have an effect (i.e., has enough energy to either escape or allow another atom to escape). Those that are still relevant are stored in the queue of particles to be processed.

As each cascade is finished, the statistics (number of particles escaping as well as source location) are stored. Once all 10,000 incident ions have been run, the model determines the total yield for each species (and fragment). One advantage of using the Monte-Carlo model with the parameterizations is that it runs very fast. This allows the model to run enough particle to gather meaningful statistics on the yields of even trace atmospheric species. It also has the advantage of making the model flexible and allowing it to gather statistics on other relevant quantities (although, some quantities will require more particles to insure a sufficiently large statistical sample). The model also allows one to look at the actual effects of the various processes and determine what are the important processes.

2.4 Model Results

Since I am focusing on change and evolution in the Martian atmosphere, there are three primary results produced by the model. First is the yields (particles escaping per impacting ion). This allows the impacting fluxes to be converted into the total amounts lost. The second result that I have the model calculate is the effective loss altitude for each species. This is the mean source altitude for each escaping atom. The third result is the partitioning of the initial energy.

The yields for each species for each atmosphere are listed in Table 2.2. The H and H_2 results (from the 1 EUV atmosphere) are not included, primarily because their escape energies are well below the 1.5 eV cutoff threshold and the model values are at best a useless lower limit. Furthermore, other escape processes (notably Jeans escape) are much more efficient at removing hydrogen. All of the results in the upper part of the table are the results for the specific species, thus the row for C is for atomic carbon escaping, as opposed to carbon in CO or CO_2. In the lower part of the table are several rows that calculate the total C, O and N yield. These are useful for doing mass balance on the atmospheric loss processes. The final row of the table contains the total yield (number of escaping atoms, regardless of their species) for each atmosphere (the 1 EUV total yield does not contain the contribution from H or H_2). Each escaping carbon is assumed to represent an escaping CO_2 molecule whose O atoms escape otherwise. Any extra O is considered to come from H_2O dissociated in the lower atmosphere [*Nair et al.*, 1994].

The column in Table 2.2 labeled pure O is a test atmosphere with only atomic oxygen. It was calculated for comparison against modeling work by Johnson and co-workers ([*Johnson and Luhmann,* 1998], personal communication, 1998). The O yield of 2.4 is well below their theoretical upper limit of 6 and agrees reasonably with their numerical model yield of ~ 3.

The differences between the 6 EUV and 3 EUV atmosphere are due to different mixing ratios as well as slightly different vertical profiles. The variations in total yield are due to the sensitivity of the sputtering process to the phase functions for

Model Yields

Species	Atmosphere			
	6 EUV	3 EUV	1 EUV	Pure O
O	4.63	4.71	5.95	2.41
CO_2	0.03	0.04	0.03	
CO	0.80	0.93	0.48	
C	0.32	0.43	1.48	
N_2			0.02	
N			1.11	
Ar			0.05	
Total O	5.49	5.72	6.55	2.41
Total C	1.15	1.40	1.99	
Total N			1.15	
Total Yield	6.64	7.12	9.74	2.41

Table 2.2: Model Yields for Each Atmosphere
The yield (number ejected per incident ion) for each species in the four atmospheres used. Blank entries indicate species that were not in the model atmosphere. The "Total" lines represent the total number of the given atom that are sputtered. The last line is the total number of atoms ejected by each incident ion. The fourth atmosphere is not realistic but is used for comparisons.

the various collisions. One of the limiting steps in the sputtering process is the need to change or transfer the momentum from a downwards particle to one traveling upwards. This is primarily done by scattering through large angles (especially cases where one of the particles scatters backwards). Because of this, the total yield from a pure O atmosphere is not related to the yield for more realistic atmospheres. Due to the way the collisions occur, those that dissociate molecules are actually very efficient at producing an escaping flux. Since the momentum (and energy) is partitioned among several particles, it is easier to "kick" one backwards.

It is interesting to note that while the carbon comes out both as elemental C as well as CO, very little comes out as CO_2. This is because most collisions that can transfer sufficient energy to allow something to escape also dissociate the CO_2. Most of the CO_2 molecules that dissociate completely to allow atomic C to escape are dissociated over several collisions. A CO_2 will be dissociated into CO and O by an impact. The CO will then undergo subsequent collisions, some of which will finish

Source Altitudes

Species	Atmosphere		
	6 EUV	3 EUV	1 EUV
O	206	180	159
CO_2	248	218	191
CO	211	191	178
C	224	200	181
N_2			203
N			192
Ar			193
Mean O	207	182	161
Mean C	215	195	181
Mean N			192

Table 2.3: Mean Source Altitude for Sputtering
The mean altitude that each species is lost from. The last three lines contain the mean altitude that each type of atom is lost from. These are calculated using weighted averages based on the yields. All altitudes are in km.

dissociating it, allowing the atomic C to escape.

Figures 2.9–2.11 show the distributions of altitudes that the various species come from. Each figure is for one of the epochs and contains the number of each species escaping from each altitude bin. Each species is normalized to its total yield. The dashed line in each plot represents the nominal homopause for the model (this is the CO_2 homopause). There is a significant range of altitudes over which most of the species originate. Note that the altitudes of the various epochs are not directly comparable, since the temperature profiles are significantly different. A better measure is to look at the distribution relative to the homopause.

The mean escape altitude for each species is shown in Table 2.3. This is actually calculated directly by the model and does not use the binned data. The lower part of the table contains the mean altitude for specific elements obtained calculating a weighted mean based on the yield of each species where that atom escapes.

The energy of the impacting ion is distributed into four categories as the model runs (Table 2.4). Some of the energy is lost to electronic excitation. As the various accelerated particles travel through the atmosphere, the have close passes with other

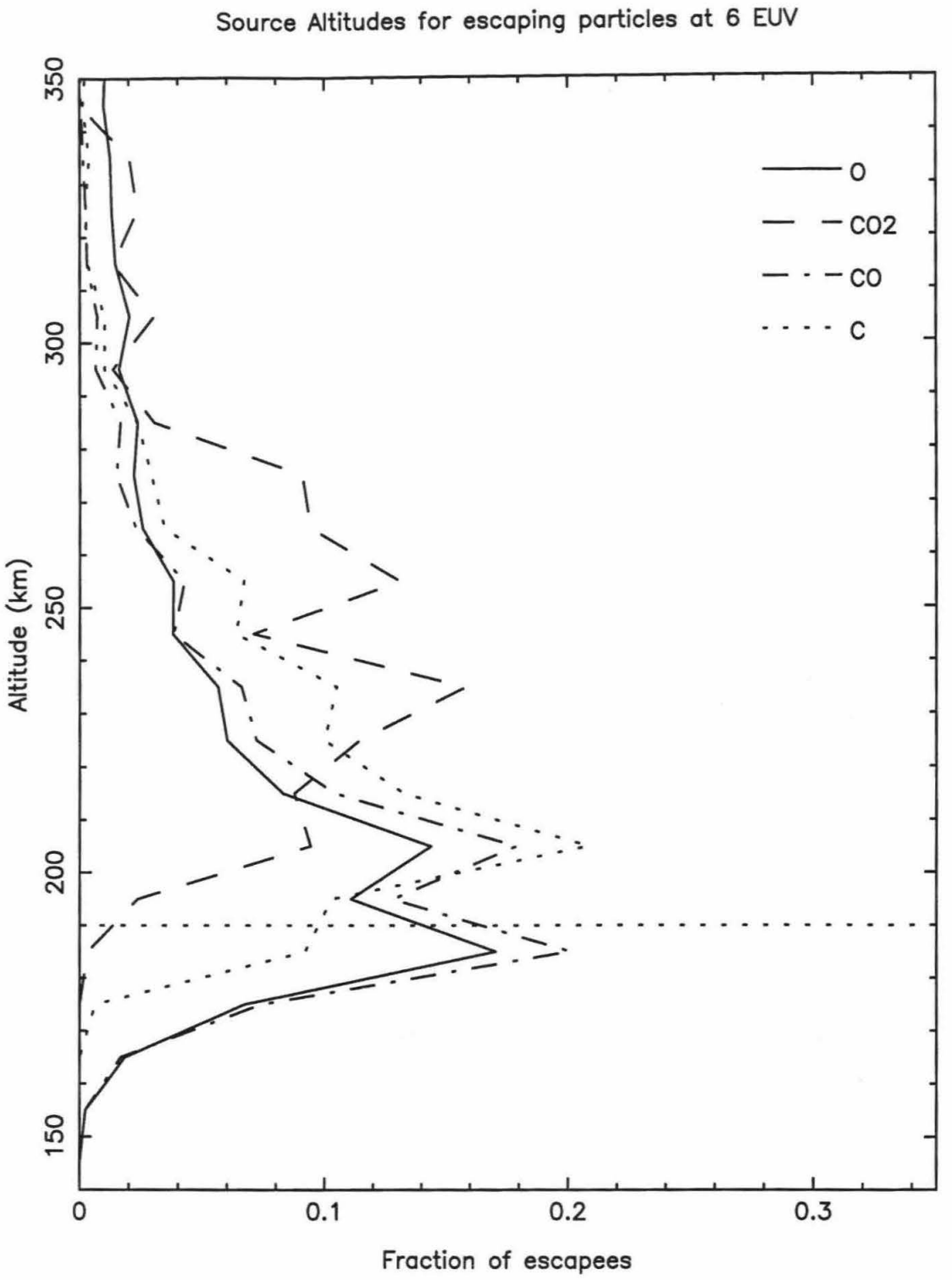

Figure 2.9: Source Altitudes for Escaping Particles at 6 EUV
The altitude that each escaping particle is accelerated from (it may have further collisions before actually escaping). Each curve is plotted as a fraction of all the molecules of that species to escape the atmosphere. The horizontal dashed line is the CO_2 homopause for the atmosphere.

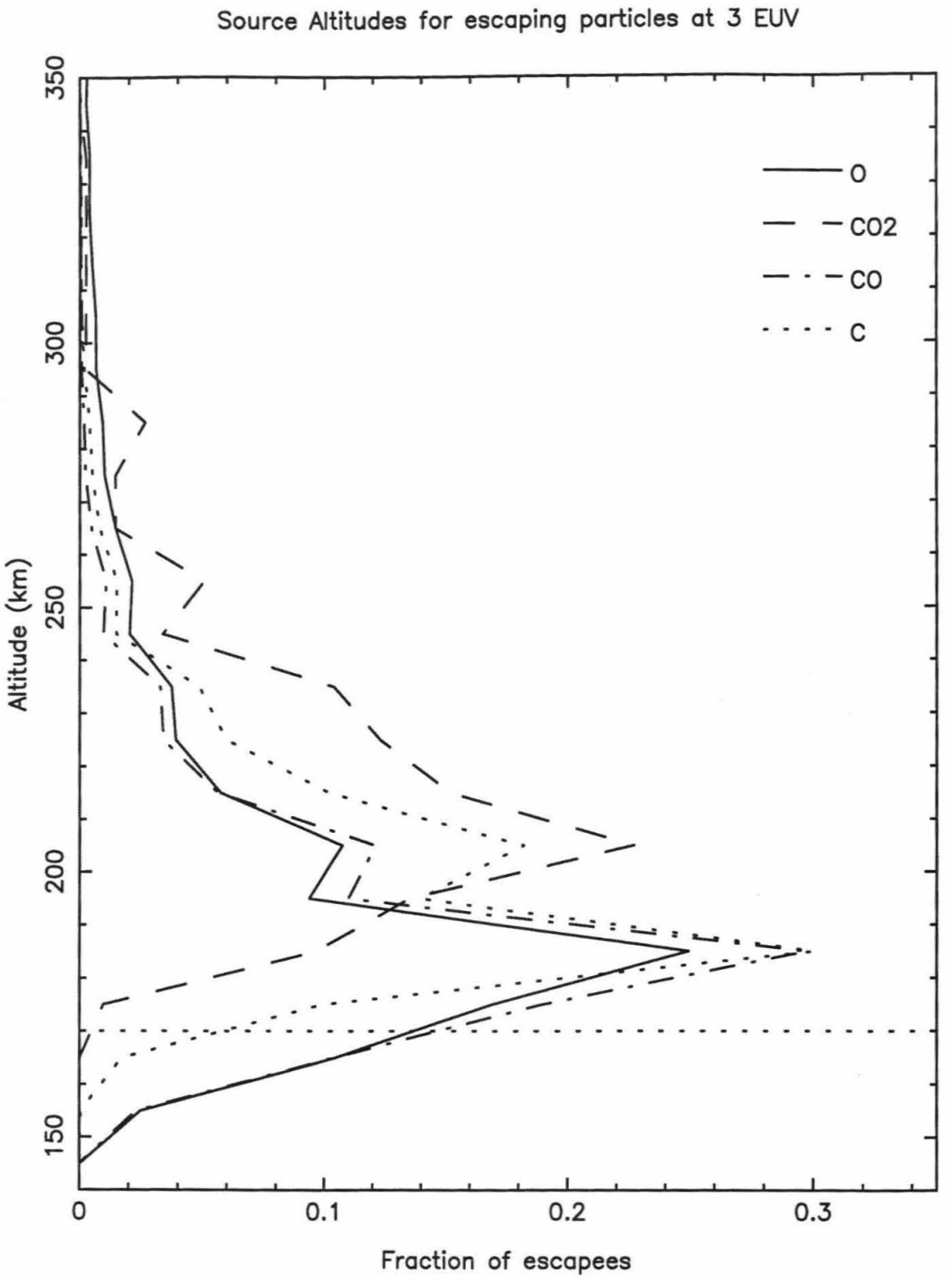

Figure 2.10: Source Altitudes for Escaping Particles at 3 EUV
Each curve is the fraction relative for the total of that species. The horizontal dashed line is the CO_2 homopause.

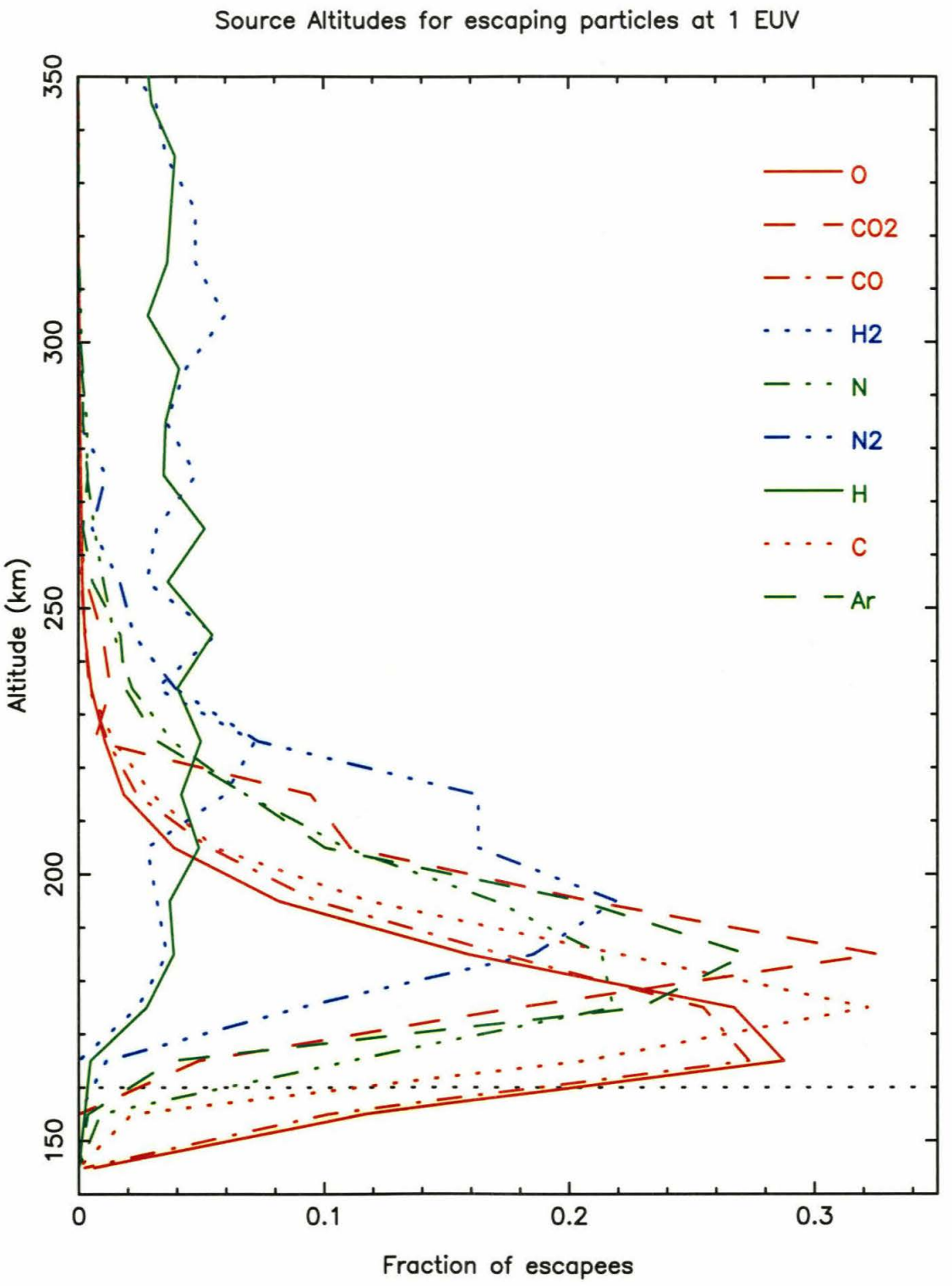

Figure 2.11: Source Altitudes for Escaping Particles at 1 EUV
Each curve is the fraction relative for the total of that species. The horizontal dashed line is the CO_2 homopause.

Energy Distribution

Energy Sink	Atmosphere 6 EUV	3 EUV	1 EUV
Electronic	580	565	401
Dissociation	152	158	252
Thermal	102	101	146
Escaping	166	176	201

Table 2.4: Model Energy Distribution
Ultimate location of the ion's initial keV of energy. These are the mean values for all the particles run in the model. "Electronic" refers to loss to electronic excitation; "Dissociation" is energy used to dissociate atmospheric molecules. "Thermal" is kinetic energy no longer sufficient to allow particles to escape—it also includes particles on ballistic trajectories. "Escaping" is the total energy of the particles that leave Mars. All energies are in eV.

molecules and atoms that they electronically excite. This is the largest sink for the energy and in the simpler ancient atmospheres, more than half of the energy goes into electronic excitation. Energy is also used to dissociate molecules during collisions. While this absorbs energy, it actually improves the sputtering yield because it is much easier for the fragments to escape than for the whole molecule. The third energy sink is in thermal motion of the atmosphere. These are particles that still have modest amounts of kinetic energy, but do not have enough to escape (or to allow anything other than hydrogen to escape). For convenience, the energy in ballistic particles, only a few eV, is also included in this category. At Mars, the ballistic component is very small (since the cutoff for thermalizing particles is close to the escape energies). Finally, some of the energy is in escaping particles. These are the ones of prime interest in the model, but they only use a small part of the total energy carried by the impacting ion.

2.5 Discussion

In order to obtain the actual fluxes of particles escaping Mars, it is necessary to multiply the yields by the total impacting flux. The plasma modeling results of

Loss Fluxes

	6 EUV	3 EUV	1 EUV	Total Loss
Sputtered O	1.0×10^{28}	6.6×10^{26}	2.3×10^{24}	1.4 bar
Exospheric O	1×10^{27}	5×10^{26}	8×10^{25}	0.3 bar
Pickup O^+	3×10^{27}	4×10^{26}	6×10^{24}	0.5 bar
Sputtered CO_2	2.2×10^{27}	1.6×10^{26}	6.9×10^{23}	0.8 bar
Escaped H_2O	1.0×10^{28}	1.2×10^{27}	8.7×10^{25}	50 m

Table 2.5: Integrated Losses
The integrated loss flux, in particles per second, for each epoch. The "Sputtered O" line contains all O atoms, regardless of the actual species. "Exospheric O" and "Pickup O^+" refer to two non-sputtering loss processes [*Luhmann et al.,* 1992]. "Sputtered CO_2" assumes that each carbon atom lost represents the loss of a CO_2 molecule (and is balanced by two O atoms). "Escaped H_2O" assuming that any O not needed to balance a C atom represents the loss of a water molecule. The last column is the integrated loss over the period modeled.

Luhmann *et al.* [1992] indicate impacting fluxes of O^+ ions of 2.6×10^9 cm^{-2} s^{-1} at 6 EUV, 1.6×10^8 cm^{-2} s^{-1} at 3 EUV and 4.8×10^5 cm^{-2} s^{-1} at 1 EUV. Using these fluxes and the model yields gives the loss fluxes indicated in Table 2.5. The table also includes estimates for the two other major loss processes for O atoms [*Luhmann et al.,* 1992] and the net water loss rate obtained by combining the three oxygen loss processes (and removing enough O to account for the loss of CO_2).

By taking the loss fluxes (Figure 2.12) and integrating over time, it is possible to estimate the total amount of atmosphere lost from Mars over the last 3.5 Gyr. Doing this yields a loss of ~0.8 bars of CO_2 and a loss equivalent to a ~50 m thick layer of water covering the planet.

While the 50 m of water is less that the geological estimates [*Carr,* 1996], it is still a significant increase over previous estimates of 3 m [*Yung et al.,* 1988]. Furthermore, it is quite likely that some of the water is still stored in the planet. The 0.8 bars of CO_2 probably represent sufficient CO_2 to create significant surface warming. It is also a significant increase over the estimate of 0.14 bars by Luhmann *et al.* [1992]. In this case, both estimates use the same impacting fluxes, but differ in the model calculation of the sputtering yield. The Monte-Carlo model improves the previous

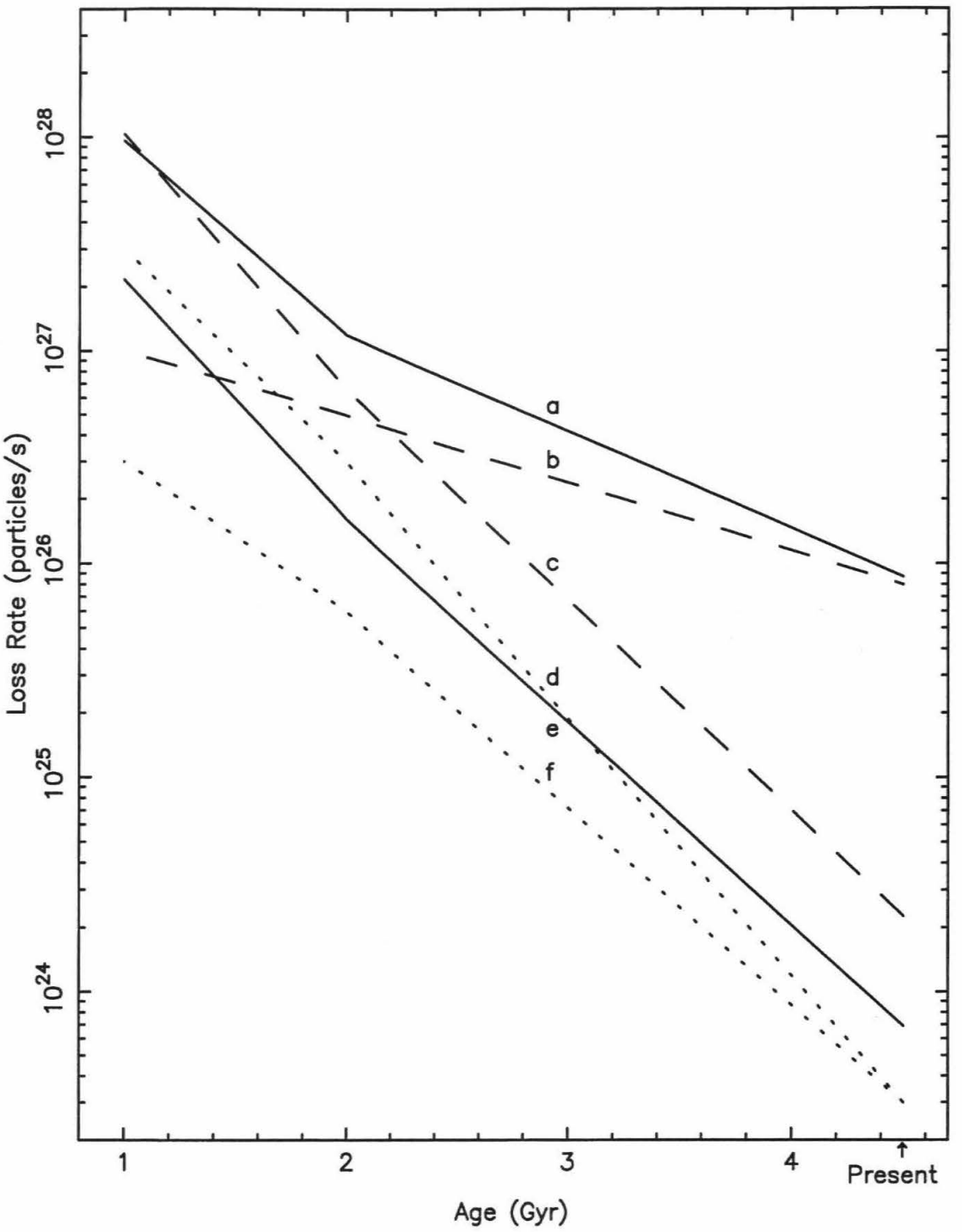

Figure 2.12: Integrated Loss over Time

Loss fluxes (particles per second) integrated over Mars. "a" is the total H_2O loss rate. "b" is the Exospheric loss rate. "c" is the O sputtering flux from the model. "d" is the O sputtering flux from Luhmann *et al.* [1992]. "e" is the CO_2 sputtering flux from the model. "f" is the CO_2 sputtering flux from Luhmann *et al.* [1992].

work by always treating CO_2 as a molecule and allowing it to dissociate. First, this allows the energetic O to impact any of the three atoms and transfer momentum to the entire molecule. Secondly, the dissociation during the full cascade creates many CO and C fragments that can escape. The fragments need less energy to escape since they are light, and more importantly, have smaller cross sections and will travel further before colliding again. This allows more collisions to occur in regions where the particles can potentially escape. Note that the actual difference in source altitudes for escaping CO_2, CO, and C reflect more than this effect. It is also caused by the energy the particle has after being accelerated and the number of collisions it can undergo and still escape. CO_2 usually has much less energy since larger energy transfers result in dissociation.

There are several significant improvements in this Monte-Carlo model over the previous one [*Kass and Yung* 1995]. The current model also also corrects the issues raised by Johnson and Liu [1998] by using the updated cross sections and scattering functions from their work. First, the model now incorporates energy loss due to electronic excitation. Secondly, the model does not assume that CO_2 always dissociates fully. And finally, it uses many more cross sections and scattering functions from molecular dynamical calculations. Of the three effects, allowing CO_2 to partly dissociate causes most of the reduction from the original estimate of $\sim$ 3 bars of CO_2 lost to the current result of $\sim$ 0.8 bars lost. The changes in the cross sections and especially the scattering function caused most of the change between the original estimate of $\sim$ 80 m of water lost and the current $\sim$ 50 m lost. While these results are closer to the original calculations by Luhmann *et al.* [1992], the two results are significantly different. This is primarily because the original calculations simplified the treatment of CO_2, but is also due to the better cross sections and scattering functions that have been calculated since then and incorporated into the current model.

The model indicates that the average energy per escaping atom is $\sim$ 25 eV for the 6 EUV and 3 EUV atmosphere. For the 1 EUV case, the average energy is $\sim$ 21 eV. Part of this difference is the effect of the H_2 in the present atmosphere. In general, the H atoms are too light to effectively transfer momentum to another species and

allow it to escape. Thus the H_2 effectively acts as an energy sink since any energy transferred to it or used to dissociate it is not available to eject other atoms. It is not a large sink and thus there was no attempt to put H_2 in the past atmospheres. Some experiments in doing so indicated that it has about a 5% effect on the yield.

Since the mean escape energies are much higher than the minimum escape energy for the atmospheric species involved the sputtering process does not directly fractionate the atmosphere. With over 20 eV of energy, the mass difference between the various isotopes will not have a noticeable effect. While the sputtering process itself does not fractionate, most of the escaping molecules come from above the homopause. Because of this, the isotopic composition of the source region differs from the bulk atmosphere (since the heavier isotopes will have a small scale height than the lighter ones). The net result is that atmosphere lost by sputtering does fractionate the bulk atmosphere.

There are three sources of error in the calculation of the total amount of atmosphere lost. The error in the Monte-Carlo model, the error in the model parameters and the error in the calculation of the impacting fluxes. Being a fast Monte-Carlo model, it was run a sufficient number of times to make the inherent model error small, but the statistical uncertainty depends on the parameter being studied. The most uncertain quantity is the escape altitude for CO_2, N_2 and Ar, but even these values should be good to 20% and the yields for the abundant species are much better. The largest error in the model results comes from the input parameters. This includes the model atmospheres, the cross sections, the scattering functions and the initial conditions. For the minor species in the present atmosphere, the parameters associated with the species will primarily affect itself. Unfortunately, the major species (CO_2 and O and their fragments) will affect all the results. Most of the important cross sections and scattering functions are now derived from molecular dynamical calculations or laboratory experiments and should be fairly good. Overall, the model results are probably valid to within 50%. Unfortunately, the calculations of the impacting fluxes are based on models of the young sun and its solar wind [*Zhanle and Walker,* 1982]. Because of this, they are much less certain and may have errors as large as an order of

magnitude. Thus the estimates for the total losses are probably only good to within a factor of ten.

In addition to the above sources of error, there is one more. Sputtering, by the process modeled here, only occurs on planets with small or no intrinsic magnetic fields. If a planet's magnetic field is strong enough to keep the solar wind from interacting with the planet's ionosphere, the sputtering will not occur. During the early history of Mars, there is evidence for a strong magnetic field [*Acuña et al.*, 1998, *Stevenson et al.*, 1983, *Kieffer et al.*, 1992]. If this field existed during the early part of the sputtering history being considered here, it would have stopped much of the sputtering [*Hutchins et al.*, 1997]. Sputtering could still have occurred during field reversals or other periods when the field was weaker or did not exist. It has also occurred since the field disappeared. Thus while a magnetic field may have kept sputtering from removing the full $\sim$ 0.8 bars of CO_2, sputtering has still played a role in the evolution of the Martian atmosphere.

2.6 Conclusion

The Monte-Carlo model results indicate that Mars has lost $\sim$ 0.8 bars of CO_2 and $\sim$ 50 m of water over the last 3.5 Gyr. This represents a significant change in the volatile inventory of the planet. The CO_2 represents a significant fraction, if not all, of the CO_2 needed to warm Mars sufficiently to allow for liquid water near the surface. Both the water and the CO_2 are large volumes compared to the known reservoirs on the planet. This indicates that sputtering is a significant process in the evolution of the atmosphere of Mars.

Because most of the major sputtering occurred early in the history of Mars, it is difficult to test the model results. In situ measurements of the modern escape fluxes of heavier species from Mars would help to constrain at least the value for the present epoch and verify the validity of the sputtering model itself. This is especially the case for C, because sputtering appears to be its dominant escape mechanism (for O, the other escape mechanisms will overwhelm the contribution from sputtering).

Bibliography

Acuña, M. H., J. E. P. Connerney, P. Wasilwski, R. P. Lin, K. A. Anderson, C. W. Carlson, J. McFadden, D. W. Curtis, D. Mitchell, H. Reme, C. Mazelle, J. A. Sauvaud, C. d'Uston, A. Cros, J. L. Medale, S. J. Bauer, P. Cloutier, M. Mayhew, D. Winterhalter, N. F. Ness, Magnetic Field and Plasma Observations at Mars: Initial Results of the Mars Global Surveyor Mission, *Science, 279,* 1676-1680, 1988.

Carr, M. H., *Water on Mars,* 229 pp., Oxford University, New York, 1996.

Fanale, F. P., J. R. Salvail, W. B. Banerdt, R. S. Saunders, Mars—the Regolith-Atmosphere-Cap System and Climate Change *Icarus, 50*, 381-407, 1982.

Forsythe, W. E., Smithsonian Physical Tables, *Smithson. Misc. Collect.,* Vol 120, *Publ. 4169*, Smithson. Inst., Washington D.C., 1954.

Goody, R. M., Y. L. Yung, *Atmospheric Radiation Theoretical Basis,* 2^{nd} ed., 519 pp., Oxford University, 1989.

Gurwell, M. A., Y. L. Yung, Fractionation of Hydrogen and Deuterium on Venus due to Collisional Ejection, *Planet. Space Sci., 41*, 91-104, 1993.

Haberle, R. M., Early Mars Climate Models, *J. Geophys. Res., 103*, 28467-28479, 1998.

Hodges Jr., R. R., Isotopic Fractionation of Hydrogen in Planetary Exospheres due to Ionosphere-Exosphere Coupling—Implications for Venus, *J. Geophys. Res., 98*, 10833-10838, 1993.

Hunten, D. M., Atmospheric Evolution of the Terrestrial Planets, *Science, 259*, 915-920, 1993.

Hutchins, K. S., B. M. Jakosky, J. G. Luhmann, Impact of a Paleomagnetic Field on Sputtering Loss of Martian Atmospheric Argon and Neon, *J. Geophys. Res., 102,* 9183-9189, 1997.

Ishimoto, M., M. R. Torr, P. G. Richards, D. G. Torr, The Role of Energetic O^+ Precipitation in a Midlatitude Aurora, *J. Geophys. Res., 91*, 5793-5802, 1986.

Jakosky, B. M., Mars Volatile Evolution—Evidence from Stable Isotopes, *Icarus, 94*, 14-31, 1991.

Johnson, R. E., D. Schellemberger, M. Wong, The Sputtering of an Oxygen Thermosphere by Energetic O^+, *Planet. Space Sci.,* submitted, 1999.

Johnson, R. E., M. Liu, Sputtering of the Atmosphere of Mars 1. Collisional Dissociation of CO_2, *J. Geophys. Res., 103*, 3639-3647, 1998.

Johnson, R. E., J. G. Luhmann, Sputtering Contribution to the Atmospheric Corona on Mars, *J. Geophys. Res., 103*, 2649-3653, 1998.

Johnson, R. E., *Energetic Charged-Particle Interactions with Atmospheres and Surfaces,* 232 pp., Springer–Verlag, New York, 1990.

Kass, D. M., Y. L. Yung, Loss of Atmosphere from Mars due to Solar Wind–Induced Sputtering, *Science, 268,* 697-699, 1995.

Kasting J. F., CO_2 Condensation and the Climate of Early Mars, *Icarus 94*, 1-13, 1991.

Kaye, G. W. C., T. H. Laby, *Tables of Physical and Chemical Constants*, Longman, New York, 1995.

Kieffer, H. H., B. M. Jakosky, C. W. Snyder, M. S. Matthews (Eds.), *Mars,* 1498 pp., The University of Arizona, Tucson, 1992.

Luhmann, J. G., R. E. Johnson, M. H. G. Zhang, Evolutionary Impact of Sputtering of the Martian Atmosphere by O^+ Pickup Ions, *Geophys. Res. Lett., 19*, 2151-2154, 1992.

Luhmann, J. G., J. U. Kozyra, Dayside Pickup Oxygen Ion Precipitation at Venus and Mars—Spatial Distributions, Energy Deposition, and Consequences, *J. Geophys. Res., 96*, 5457-5467, 1991.

McElroy, M. B., Mars: An Evolving Atmosphere, *Science, 175*, 443-445, 1972.

McElroy, M. B., T. Y. Kong, Y. T. Yung, Photochemistry and Evolution of Mars' Atmosphere: A Viking Perspective, *J. Geophys. Res., 82*, 4379-4388, 1977.

McKay C. P., W. L. Davis, Duration of Liquid Water Habitats on Early Mars, *Icarus, 90*, 214-221, 1991.

Nair, H., M. Allen, A. D. Anbar, Y. L. Yung, R. T. Clancy, A Photochemical Model of the Martian Atmosphere, *Icarus, 111*, 124-150, 1994.

Pollack, J. B., J. F. Kasting, S. M. Richardson, K. Poliakoff, The Case for a Wet, Warm Climate on Early Mars, *Icarus, 71*, 203-224, 1987.

Smith, P. H., N. K. Bewtra, Charge Exchange Lifetimes for Ring Current Ions, *Space Sci. Rev., 22*, 301-317, 1978.

Stebbings, R. F., A. C. H. Smith, H. Ehrhardt, Charge Transfer between Oxygen Atoms and O^+ and H^+ Ions, *J. Geophys. Res., 69*, 2349-2355, 1964.

Stevenson, D. J., T. Spohn, G. Schubert, Magnetism and Thermal Evolution of the Terrestrial Planets, *Icarus, 54,* 466-489, 1983.

Tombrello, T. A., Solar System Sputtering, *Radiation Effects, 65*, 149-158, 1982.

Watson, C. C., Topics in Classical Kinetic Transport Theory with Applications to the Sputtering and Sputter–Induced Mass Fractionation of Solid Surfaces and Planetary Atmospheres, Ph.D. Thesis, Yale University, 1980.

Yung, Y. L., J. S. Wen, J. P. Pinto, M. Allen, K. K. Pierce, S. Paulson, HDO in the Martian Atmosphere: Implications for the Abundance of Crustal Water, *Icarus, 76,* 146-159, 1988.

Zhang, M. H. G., J. G. Luhmann, S. W. Bougher, A. F. Nagy, The Ancient Oxygen Exosphere of Mars—Implications for Atmosphere Evolution, *J. Geophys. Res., 98*, 10915-10923, 1993.

Zhanle, K. J., J. C. G. Walker, The Evolution of Solar Ultraviolet Luminosity, *Rev. Geophys., 20*, 280, 1982.

Chapter 3 Modeling Atmospheric $\delta^{13}C$

3.1 Introduction

The current atmosphere of Mars is very thin (7 mbar) and it appears to be in vapor equilibrium with the permanent southern CO_2 polar cap [*Leighton and Murray* 1966, *Kieffer et al.,* 1992]. This implies that the present atmosphere is coupled to the solid planet. And yet, there is the evidence that for early Mars, the climate was significantly warmer [*Carr,* 1996, *Kieffer et al.,* 1992]. This early state was presumably partly due to a thick CO_2 greenhouse. In going from one state to the other, the atmosphere had to undergo significant climatic change. This raises the question of when it occurred, as well as the processes that caused the transition.

Since the current atmosphere is primarily CO_2, and it is thought that the past atmosphere was as well, understanding the history of the CO_2 provides a powerful understanding of the bulk atmosphere. In order to understand the CO_2, it is necessary to understand and model the Martian carbon "cycle."

The results from the previous chapter indicate that sputtering is a significant process in the evolution of the Martian atmosphere, but it is not the only process. Other processes play a significant role in determining how the atmosphere evolves over the history of the planet. In order to understand the history of the atmosphere, it is important to understand the various processes that can affect the atmosphere and determine the importance of each process. One of the few measurements that can indicate the importance of each process is the isotopic ratio of the various species in the atmosphere. They represent an integrated record of all the processes that have affected the atmosphere as it evolved. Due to the importance of the carbon "cycle" in understanding the evolution of the atmosphere, the carbon isotopic system is a particularly useful one to study.

3.2 Martian Carbon "Cycle"

Figure 3.1 is a simple conceptual model of the Martian carbon "cycle." There are five major carbon reservoirs on Mars and several major processes that affect carbon. The importance of the various processes and reservoirs has changed with time. While the conceptual model does not contain any temporal evolution, section 3.3 and the simple numerical model focus on the evolutionary aspects.

The starting place for the Martian carbon "cycle" is the mantle reservoir. This is carbon (mostly from carbonaceous condritic material) that was buried in the planet during accretion. There are very few constraints on the size of this reservoir. It was probably quite large initially and may still be very large, especially when compared to the other reservoirs.

The second reservoir is the Martian atmosphere. Most of the atmosphere is CO_2. This is the best characterized reservoir in the system and the only one where there are measurements determining its size (7 mbar). It is also one of two reservoirs (the other being the polar cap) that has actually been observed. The atmosphere is the main exchange center that all of the other reservoirs communicate with.

The carbon in the mantle is released into the atmosphere through volcanic emplacement of mantle material and the outgassing associated with the magma. This outgassing is the primary source of new carbon into the atmosphere. From estimates of the history of volcanic activity and the associated inferred intrusive emplacements, and the carbon content of the magmas, the flux can be crudely estimated. Outgassing is probably the source of the current atmosphere. Due to the need to fractionate the heavier noble gases, the atmosphere needed to exist before the end of late heavy bombardment and is more than just a late veneer of volatiles [*Pepin,* 1994, *Zahnle,* 1997].

The second main process affecting the atmospheric reservoir is non-thermal escape to space. For carbon, this is primarily sputtering, but solar wind pickup and dissociative recombination may contribute to the flux. This carbon is lost permanently from the planet and exits the "cycle." While there are no direct measurements of the carbon loss fluxes, models (such as the sputtering model in chapter 2) can give good

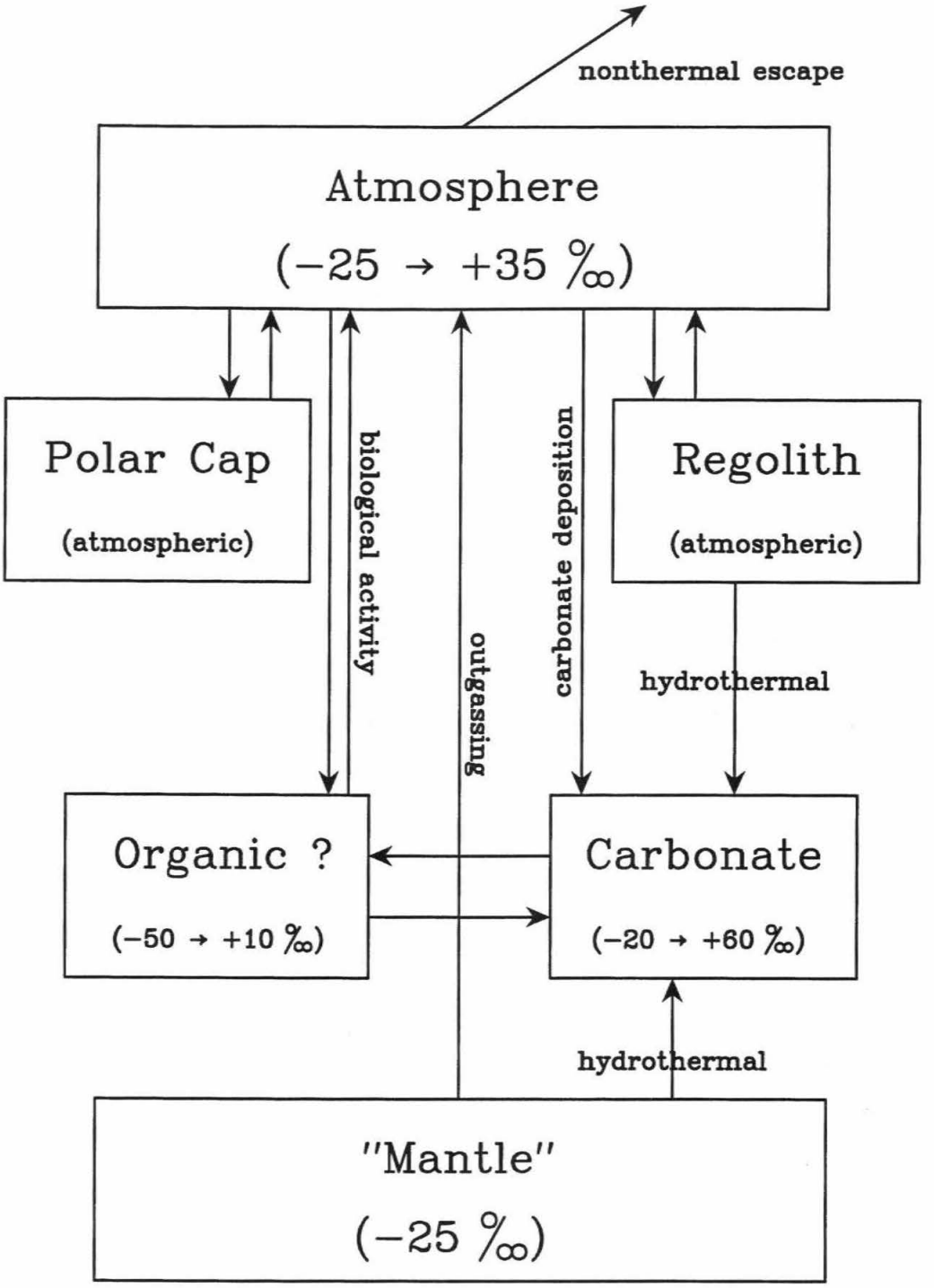

Figure 3.1: Martian Carbon "Cycle"
A simple model for the behavior of carbon at Mars. This is not actually a closed cycle and carbon usually only goes through the system once, either being lost to space or deposited as carbonate. The expected $\delta^{13}C$ ranges, based on the evolution of the system, are indicated for each reservoir.

estimates of the flux. There is hope that the upcoming Japanese Planet-B mission will directly measure the escaping flux.

The polar cap is the CO_2 frost (and ice) that has condensed out of the atmosphere at the poles. There are seasonal CO_2 caps at both poles and a permanent Southern cap. This carbon cycles through the atmosphere on a very short time scale. The Northern CO_2 cap is seen (from orbit) to sublimate away every year and there are hints that even the permanent part of the Southern CO_2 cap may occasionally disappear [*Jakosky and Barker*, 1984]. The obliquity changes the planet undergoes probably cause even the permanent Southern cap to exchange on a short time scale (compared to the age of the atmosphere) [*Ward,* 1992]. While surface pressure measurement give a good estimate of the size of the seasonal caps ($\sim$ 1 mbar), the size of the permanent Southern cap is not known.

The regolith reservoir is the CO_2 that is either adsorbed onto the regolith, or possibly, in places, deposited as a frost. This CO_2 comes from the atmosphere. Like the polar caps, it will probably also exchange with the atmosphere, but it is probably much less mobile and is buffered by the diffusion times through the regolith [*Fanale and Jakosky,* 1982]. This may result in a near surface component that exchanges rapidly (as fast as the permanent polar cap) while a deeper component responds much slower and may only be affected by its local vapor pressure equilibrium. Unlike the polar cap reservoir, there is little information on the actual volume. Laboratory measurements of adsorption amounts as well as model calculations are hampered by a lack of knowledge of the mineralogy of the Martian regolith [*Fanale et al.,* 1982]. There is also the question of whether water ice may exist and block any diffusion of the CO_2 into the deeper regolith.

Any life or fossil deposits that may exist at Mars represent a small element of the carbon cycle. The hostile surface environment would oxidize and remove any surface deposit and there is insufficient energy at depth to develop a significant biomass. There is no way of estimating the actual size of such a reservoir beyond assuming that it is small.

The final major reservoir is carbonate deposits. This may be deposited by several

different processes. There are traces of carbonate in several of the SNC meteorites and some of it appears to have been deposited in a hydrothermal system. The source of the carbon in the hydrothermal system may be the atmosphere itself, the magma that created the system or even preexisting CO_2 in the regolith. Also, carbonate may have been directly deposited from the atmosphere either in aquatic environments or in surface layers of water on mineral grains. Biological processes may have also deposited carbonate (much the way most of the Terrestrial carbonate is formed). There are no estimates of the volume of carbonates existing on Mars, but so far, no significant surface deposits have been detected [*Clark et al.*, 1998].

Since plate tectonics (or other crustal recycling) does not appear to occur on Mars, there is no way for the carbonate deposits to be recycled into the mantle (other than possible local recycling when new magma is emplaced adjacent to an old carbonate deposit). This has profound implications for the carbon "cycle" on Mars. In between the large flux lost to space and the fact that the carbonate is a permanent sink, the carbon "cycle" on Mars is not actually a cycle. Instead it is a one way path from the mantle into the atmosphere and then it is lost from the system (either to space or into carbonate). This is especially true if the atmosphere, polar caps, and regolith are considered one reservoir (since the latter two components exchange on very short time scales).

Since the bulk atmosphere is CO_2, the history of the atmosphere is effectively the history of the carbon. Thus, understanding the history of the carbon yields an understanding of the general history of the atmosphere.

3.3 $\delta^{13}C$ History

There are several interesting implications of this conceptual model of the Martian carbon "cycle." Since the system is open ended, there is no requirement for the fluxes to be balanced. Furthermore, it will evolve and change with time. One of the few probes into the past history of the planet, and especially the carbon "cycle," is the stable isotopic ratios. They contain a signature of past processes and reservoirs.

One of the simplest systems is the $\delta^{13}C$. For convenience, I will use $\delta\,{}^{0}/_{00}$ notation

$$\delta^{13}C = \left(\frac{R - R_{std}}{R_{std}}\right) \times 10^3 \tag{3.1}$$

where R is the ratio $^{13}C/^{12}C$ of the reservoir in question and R_{std} is the ratio of a standard (in this case the PDB standard). This is in units of per mille (or thousandths). It is also convenient to express fractionation factors (α—the ratio of the abundance of the isotopes in the product to that in the source) in per mille as well, where the source is assumed to be at the standard value. Since the atmospheric $\delta^{13}C$ is only affected by the processes that drive the evolution, it is an excellent probe of the evolution of the atmosphere and the carbon "cycle."

There are several measurements of the $\delta^{13}C$ values for Mars that are summarized in Figure 3.2. For each component/measurement, the measured value and uncertainty are graphed. The values for the equivalent Terrestrial reservoirs are included for comparison. The first thing to notice is that the range of values on Mars is much larger than for the Earth. This is partly due to the open nature of the Martian system that prevents homogenization as well as the fractionation caused by the various processes.

For the mineralogical carbon, the only measurements are from the SNC meteorites. Looking at the magmatic component (i.e., the primary rock), Wright *et al.* [1990] measured values between -32 ${}^{0}/_{00}$ and -17 ${}^{0}/_{00}$ (relative to the PDB standard–for consistency, all $\delta^{13}C$ values have been converted to the PDB standard), depending on the meteorite and the phase being measured. Based on the different values and laboratory measurements of fractionation effects during crystallization, they derived a bulk mantle value of -25 ${}^{0}/_{00}$. This can be compared to the terrestrial value of -5 ${}^{0}/_{00}$, but the terrestrial mantle is "contaminated" by recycled carbon that has been fractionated by other processes.

The small amounts of carbonate found in the SNC meteorites have PDB values between +6 and +10 ${}^{0}/_{00}$. This is significantly heavier than would be expected for carbonate deposition from Martian magmatic carbon and may be due to carbon from another source or an unusual deposition environment. In particular, it may reflect a

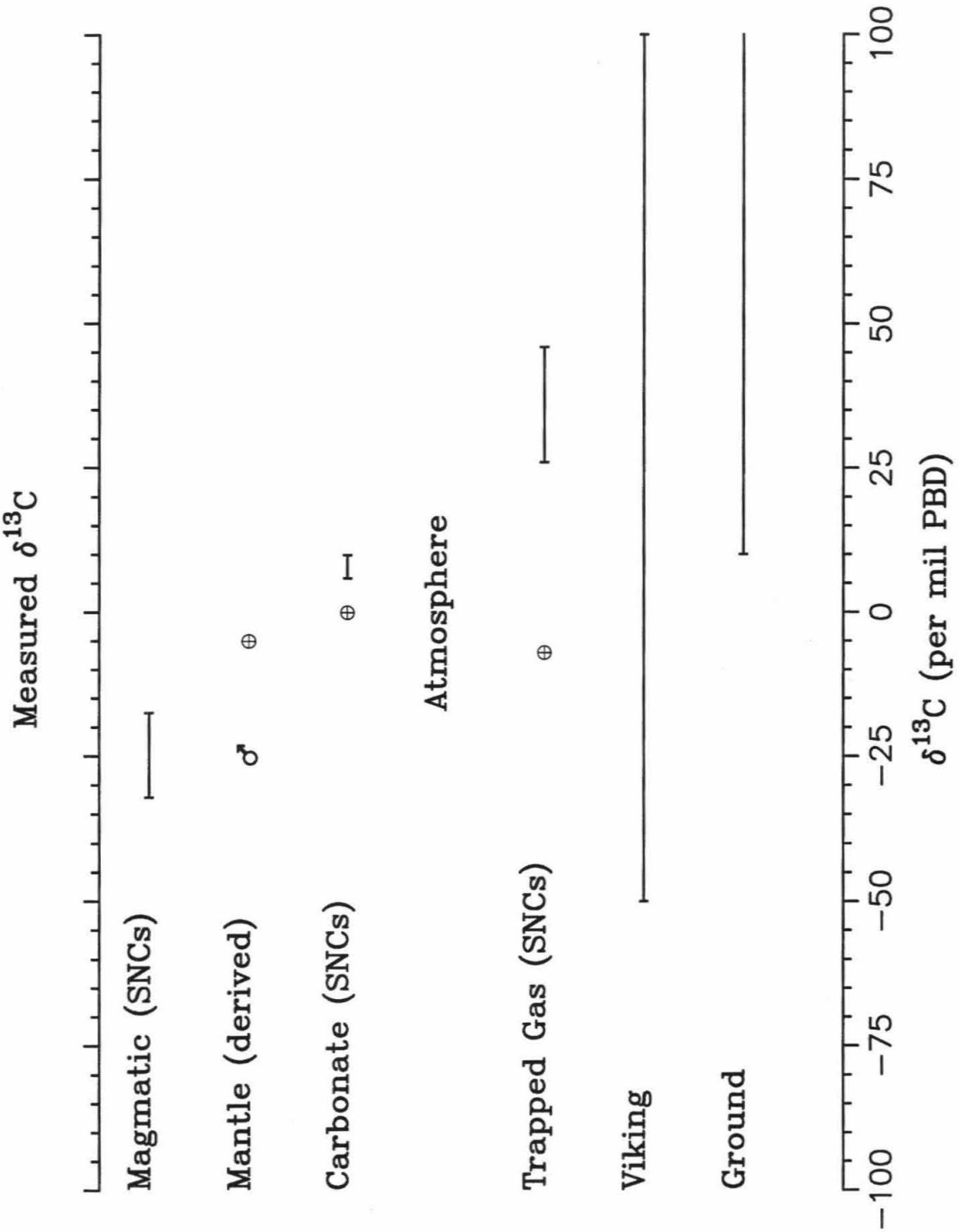

Figure 3.2: Summary of $\delta^{13}C$ Measurements

A table of $\delta^{13}C$ measurements for Mars. The first three values are values for solid reservoirs. The latter three values are measurements of the current atmospheric composition. Each value is indicated with its error bars. For comparison, the corresponding Terrestrial values are also included. [*Wright et al.*, 1990, *Jakosky et al.*, 1994]

depositional environment where the carbon was in equilibrium with an atmosphere that differed significantly from the mantle value.

There are three independent measurements of the current atmospheric $\delta^{13}C$. The first is from trapped gases in impact melts in the SNC meteorites. After correcting for the trapping process and composition of the glass, Wright *et al.* [1990] obtain $+36 \pm 10\,{}^0/_{00}$. The Viking landers were able to make a coarse measurement and obtained $+25 \pm 50\,{}^0/_{00}$ for the atmospheric CO_2. Ground based measurements of the Martian atmosphere obtained $+70 \pm 60\,{}^0/_{00}$ [*Jakosky et al.*, 1994].

In addition to measurements of the various reservoirs, it is possible to calculate the fractionation effects of the various processes that affect the carbon. These are shown in Figure 3.3 in schematic form. Starting with an atmosphere at $0\,{}^0/_{00}$ (and assumed to be infinite in size), the $\delta^{13}C$ vale of the carbon removed by the various processes is shown.

The carbon sputtered will be between $15\,{}^0/_{00}$ and $40\,{}^0/_{00}$ (depending on the epoch) lighter than the source atmosphere. The sputtering process itself does not fractionate the atmosphere (the carbon atoms that actually escape all have sufficient energy to escape regardless of the isotope). But the sputtering occurs at altitudes well above the homopause (see Table 2.3). Above the homopause, each species takes on its own mass dependent scale height. This also occurs for the different isotopes due to their different masses. Thus the atmosphere being sputtered is lighter than the total atmosphere and the net effect of the sputtering is to enrich the atmosphere.

When the Monte Carlo model calculates the sputtering fluxes, it also calculates the altitude that each sputtered atom comes from. This is used to calculate an effective exobase (see Figures 2.9, 2.10, and 2.11). The exobase altitude, along with other atmospheric parameters (including the homopause estimated from the density profiles), is used in the following formula [*Jakosky et al.*, 1994]:

$$\alpha = \exp\left(\frac{-g\,\Delta m\,\Delta z}{kT}\right) \tag{3.2}$$

where α is the fractionation factor (in this case, the ratio of the abundances of the two

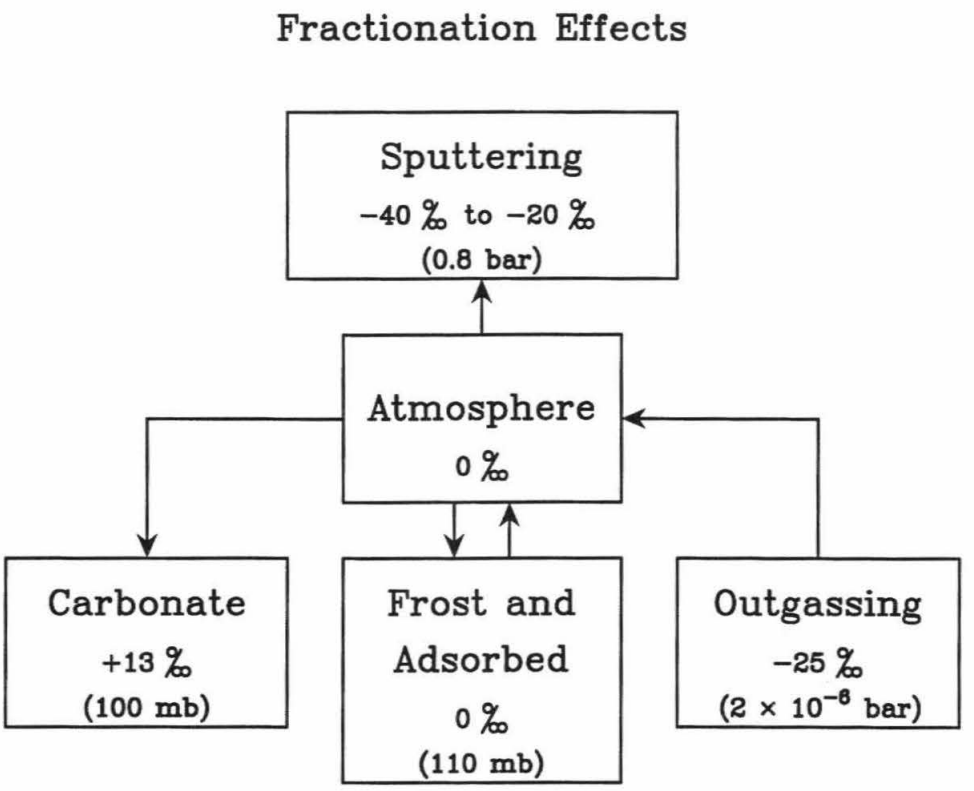

Figure 3.3: Fractionation of Loss Processes
For an atmosphere at 0 $^{0}/_{00}$, the fractionation that would occur for reservoirs in contact with it. In parentheses are the nominal sizes of the various reservoirs for the "Standard" numerical model case.

Sputtering Fractionation

	Δz	T	α
6 EUV	25	500	0.96
3 EUV	25	380	0.97
1 EUV	21	270	0.98

Table 3.1: Fractionation Factors Due to Sputtering
The parameters for equation 3.3 and the resulting fractionation factor. Δz is in km, and T is in Kelvins [*Zhang et al.,* 1993].

isotopes at the exobase to that at the homopause), g is the acceleration of gravity, Δm is the mass difference between the isotopes, Δz is the distance between the homopause and exobase, k is Boltzmann's constant, and T is the temperature in the upper atmosphere. For carbon at Mars, the equation reduces to:

$$\alpha = \exp\left(\frac{-0.446\,\Delta z}{T}\right) \tag{3.3}$$

with Δz in km and T in Kelvins. Using the results from Table 2.3, and upper atmospheric temperatures from Zhang *et al.* [1993], gives the values in Table 3.1. Sputtering is then assumed to be a Rayleigh distillation process:

$$R = R_0 f^{(\alpha-1)} \tag{3.4}$$

where R is the resulting isotopic ratio, R_0 is the initial ratio, and f is the fraction of the reservoir remaining. The less of the initial reservoir that remains, the more fractionated the remainder is. For the fractionation factors due to sputtering, the effect is not significant until the amount left is a small fraction of the initial reservoir.

The standard temperature dependent formulation of the carbon fractionation in carbonate formation [*Faure,* 1991] indicates that carbonate formed at 0 C is $\sim 13\,{}^0/_{00}$ heavier than the source atmosphere. This is for abiotic carbonate formation. 0 C was selected because most of the carbonate is expected to have formed in very cold water (since it is hard to warm the atmosphere even that much). Hydrothermal carbonate will probably have a different fractionation coefficient, depending on the

actual conditions under which it formed, but all of the carbonate is assumed to form with the same fractionation factor for simplicity.

Laboratory measurements [*Eiler and Kitchen,* 1999] indicate that the carbon in CO_2 does not fractionate during condensation. Thus the condensation of even large amounts of CO_2 does not affect the $\delta^{13}C$ of the atmosphere. While there is no laboratory data of fractionation of adsorbed CO_2, it was assumed that adsorption is not a fractionating effect, based on the lack of fractionation during condensation. Model runs with adsorption being a slightly fractionating process ($\sim 5\ ^0/_{00}$) exhibited minimal differences from those with no fractionation.

Terrestrial biological systems are up to 25 $^0/_{00}$ lighter than the source carbon. This depends on the type of biological processes being used but is a result of the binding energy difference between the two types of CO_2.

With the $\delta^{13}C$ measurements and estimates of the fractionation effect of the various processes involved in the Martian carbon "cycle," it is possible to estimate the expected $\delta^{13}C$ values for the various components (see Figure 3.1) as well as their changes with time. The mantle is at -25 $^0/_{00}$ and since there is no input and all of the removal is bulk removal, its value does not change with time. If the atmosphere is originally formed from outgassed carbon, its initial value will be -25 $^0/_{00}$, but current measurements indicate it has evolved to a present value of +35 $^0/_{00}$, primarily under the effects of sputtering (and other non-thermal escape) and carbonate deposition.

Depending on when any potential biota was created and the atmospheric (or other source of carbon) $\delta^{13}C$, it would be expected to have values in the range of -50 $^0/_{00}$ early and increasing to as high as +10 $^0/_{00}$. Most of it will probably have values at the lower end of the range, assuming conditions the most favorable for life early in Mars' history. The value of the carbonate will depend on the source of the carbon, how it is formed, and when it is formed. This creates a large range with early carbonate possibly being as light as -20 $^0/_{00}$, while later carbonate could be as heavy as +50 $^0/_{00}$.

3.4 Numerical Simulation

While the simple analysis of the $\delta^{13}C$ values of the various reservoirs presented in the previous section is useful, a time varying model is really needed to understand the behavior of the system. Furthermore, this allows some of the $\delta^{13}C$ values (especially the atmospheric value which contains a "memory" of the evolution) to be used as a constraint to help determine the values of various parameters. The model implementation is described in the appendix. The following section gives an overview of the model and addresses the major issues involved.

I created a simple box model to track the size and $\delta^{13}C$ of the Martian atmosphere. I started at the 6 EUV sputtering case (1 Gyr into the planet's history or 3.5 Gyr ago) and ran the model to the present. For all of the runs, I use 1 Myr time-steps. For convenience, the size of the atmosphere (and other reservoirs) is measured in bars of CO_2 (as a surface pressure).

The model includes four processes (see Figure 3.3). The two main processes are sputtering (representing all of the non-thermal escape processes) and carbonate deposition. It also includes outgassing and frost condensation and sublimation (this is also used to model the regolith absorption).

In addition to the atmosphere, the model has two other reservoirs it considers. Together they represent the polar cap and regolith reservoir from the carbon "cycle" model, but are divided along functional lines. The first is a short term reservoir, called the "Polar Cap" reservoir for convenience. This one exchanges with the atmosphere on short (compared to the 1 Myr time-steps) time scales and is always in isotopic equilibrium. Its size is constant (and is a parameter for the run). Its primary effect is to dilute the distillation effects of the loss processes by increasing the size of the remainder. The second reservoir is a long term reservoir, called the "Regolith" reservoir. It is formed when the atmosphere condenses (collapses), early in the history, and then loses mass to keep the atmosphere in vapor equilibrium. It may or may not be in isotopic equilibrium with the atmosphere, depending on the model run. The two reservoirs are parameterized by indicating their present size. The model deter-

mines their initial size to insure they finish with the desired size. The ability of the "Regolith" reservoir to exchange is controlled by a model switch.

The mantle and carbonate deposits are not reservoirs in the model, but are treated as sources and sinks, so their size and $\delta^{13}C$ do not need to be modeled explicitly. They are just represented by the appropriate fluxes. The mantle is represented by outgassing which is always at -25 $^{0}/_{00}$ and is controlled by the model volcanism history chosen. Two different models from Hutchins and Jakosky [1996] were used.

When doing a numerical model, it is necessary to consider several other important features. The first is the idea of a "catastrophic" atmospheric collapse. If the initial atmosphere is thick, warm and moist, at some point it needs to transition to the current thin, cold, dry one. While the transition might be continuous, atmospheric modeling by Pepin [1994] seems to indicate it would be rapid due to the following positive feedback mechanism. As the atmosphere gets thinner due to the various atmospheric loss processes, the greenhouse gets weaker and weaker. At some point, CO_2 starts to condense out. This thins the atmosphere further thus weakening the greenhouse more, causing the temperatures to fall even further. This rapidly removes most of the CO_2 from the atmosphere and it reaches the present vapor equilibrium state. The model has this occur in one time-step. Once the total atmospheric pressure reaches a specified level, the atmospheric pressure is set to 7 mbar and the extra is placed in the "Polar Cap" and "Regolith" reservoirs. From then on, the atmosphere is replenished each time-step from the "Regolith" reservoir so that it remains at the 7 mbar pressure equilibrium. This insures that the model pressure at the present is correct.

The sputtering loss flux and its isotopic fractionation is set by the results of the Monte-Carlo model. The total amount of carbonate created over the model history is an input parameter, as is the time variability of the process. I use three different models for the carbonate deposition. The first is a simple constant flux model where the carbonate deposition is independent of the model time and chosen to insure the desired amount of carbonate is created. The second carbonate deposition model deposits a constant fraction of the sputtering loss. The necessary fraction is calculated

at the beginning to insure the desired amount of carbonate is created. The final model is a step model. Most of the carbonate (95%) is deposited before the collapse occurs and the rest afterwards. Before and afterwards, the carbonate is deposited at a constant rate, but the two rates will be different. Again, the two rates are chosen to insure that the desired amount of carbonate is deposited. This last case is used to simulate the idea that in a warm, moist climate, carbonate is relatively easy to form, but in the current cold, dry climate (created when the collapse occurred), it will occur very slowly.

3.5 Results

Figure 3.4 shows the results of a typical run (this is run A in Table 3.2). The solid curve shows the evolution of the atmospheric $\delta^{13}C$ during the model run. The point with error bars indicates the current atmospheric $\delta^{13}C$ value. This run was chosen as the "standard" run because it reproduces the current measurements very well. The vertical dotted line at 1.15 Gyr is the age when the atmosphere collapses. The dashed lines are cases Y (the upper curve) and Z (the lower curve) from Table 3.2 and represent the history with only sputtering or only carbonate deposition, respectively.

This case (the solid line) has 100 mbar of carbonate deposition using the step function (so 95 mbar are deposited before the collapse at 1.15 Gyr and the final 5 mbar are deposited afterwards). There is 100 mbar of CO_2 that is assumed to be in the regolith at present. And finally, there are 10 mbar that are assumed to be in the polar caps (and shallow, fast exchanging regolith adsorbed CO_2). The atmosphere starts the run with 1.03 bars of CO_2 and the atmosphere collapses when the total pressure reached 0.75 bars.

In general, the atmospheric $\delta^{13}C$ increases with time due to the effects of sputtering. The various abrupt slope changes are due to the coarse resolution of the sputtering model with only 3 epochs. The exception is the change when the atmosphere collapses. Here, the rate of carbonate deposition changes abruptly (due to the step function) and this changes the balance between the fractionation due to carbon-

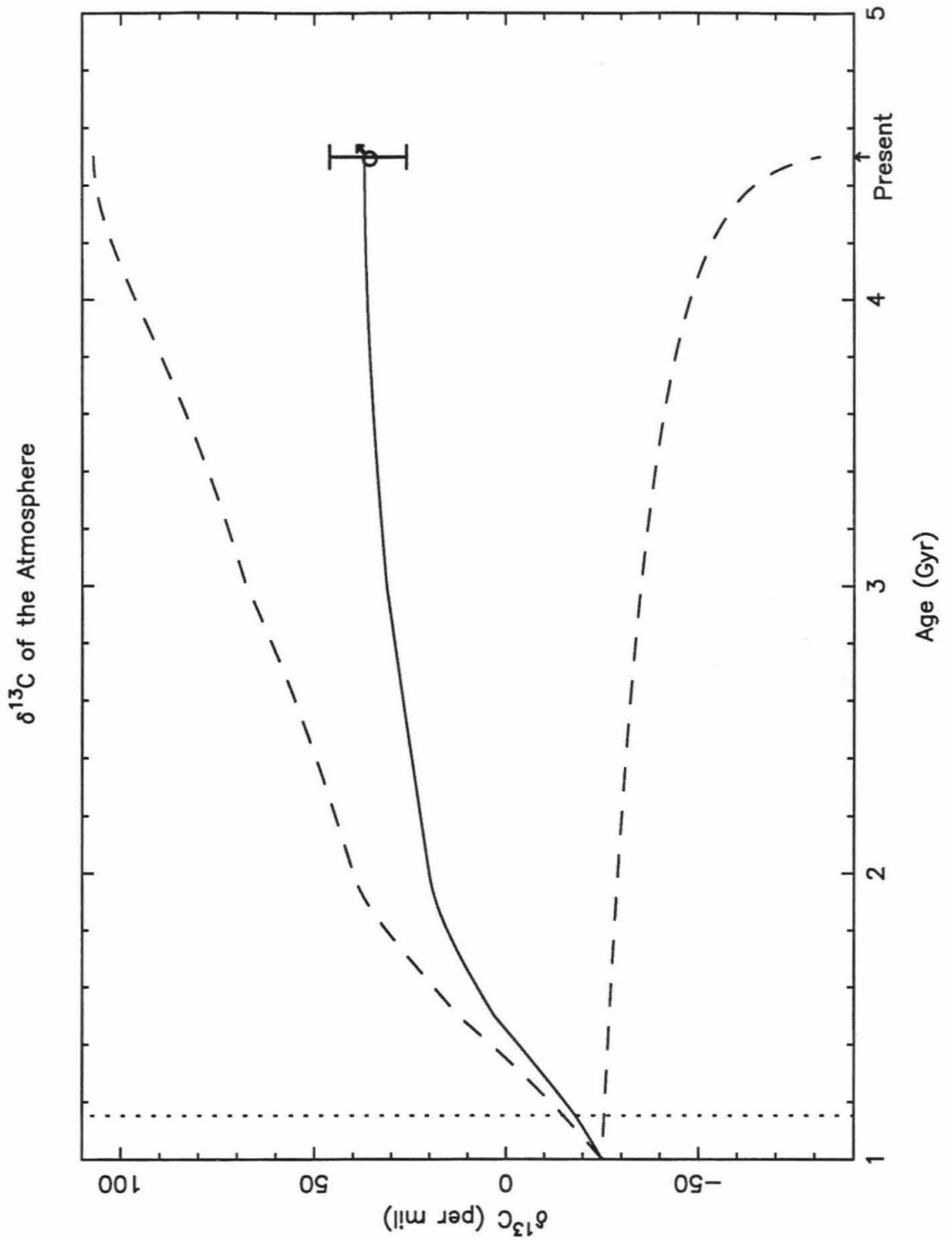

Figure 3.4: δ^{13}C History of the Atmosphere
The solid curve is the δ^{13}C value for the "Standard" model. The upper dashed curve is a case with only sputtering. The lower dashed curve is a case with only carbonate formation. The point at the present indicates the best estimate from the SNC meteorite measurements.

Model Results

	Polar Cap	Carbonate	Regolith	Collapse	$\delta^{13}C$	Note
A	0.01	0.1	0.1	0.75	37.0	
A[a]	0.01	0.1	0.1	0.75	23.0	a
A[b]	0.01	0.1	0.1	0.75	32.0	b
B	0.1	0.1	0.1	0.75	24.0	
C	0.02	0.1	0.1	0.75	35.3	
D	0.005	0.1	0.1	0.75	37.6	
E	0	0.1	0.1	0.75	37.8	
F	0.01	1.0	0.1	0.75	16.8	
G	0.01	0.5	0.1	0.75	27.0	
H	0.01	0.2	0.1	0.75	34.2	
I	0.01	0.05	0.1	0.75	38.4	
J	0.01	0	0.1	0.75	39.9	
K	0.01	0.1	0.5	0.75	4.0	
L	0.01	0.1	0.2	0.75	22.5	
M	0.01	0.1	0.05	0.75	50.2	
N	0.01	0.1	0	0.75	80.4	
O	0.01	0.1	0.1	2.0†	22.5	a
P	0.01	0.1	0.1	2.0†	31.5	b
Q	0.01	0.1	0.1	1.0‡	37.3	
R	0.01	0.1	0.1	0.5	36.0	
S	0.01	0.1	0.1	0	23.6	a
T	0.01	0.1	0.1	0	32.8	b
U	0.01	0.1	0.1	0.75	-1.2	c
V	0.1	0.1	0.5	0.75	11.4	c
W	0.01	1.0	0	0.75	42.6	
X	0.01	0.005	0.1	2.0†	39.2	b
Y	0	0	0	0	107.1	
Z	0	0.8	0	0	-81.3	a, d

† The model starts collapsed
‡ The model collapses on the first time-step
a: Constant carbonate deposition rate
b: Decreasing carbonate deposition rate
c: No isotopic exchange between regolith and atmosphere
d: No sputtering

Table 3.2: Model Results for Selected Cases
Results of changing the various model parameters. The "Polar Cap," "Carbonate," and "Regolith" columns refer to the current size of that reservoir, in bars. The "Collapse" column indicates the pressure at which the atmosphere collapses (in bars) and the "$\delta^{13}C$" column is the model value for the current atmosphere, expressed in per mille (the measured value is $\sim +35\,{}^{0}/_{00}$).

ate deposition and sputtering, allowing the atmosphere to fractionate faster. Towards the present, curve flattens significantly due to the low sputtering rate and its weak fractionation. This slow change is further accentuated by the carbonate deposition.

While model results are very useful, the model is significantly under-constrained. The only measurement the model tries to match is the current $\delta^{13}C$ of the atmosphere. While there are other constraints, they are either built into the model (such as the current atmospheric pressure) or very generous upper bounds on reservoir sizes that may not be full. For example, the maximum adsorbed regolith CO_2 is $\sim$ 1 bar, but that assumes an unreasonable regolith composition [*Haberle,* 1998]. Thus I ran several cases, summarized in Table 3.2, to explore the parameter space and determine how the various reservoir sizes and assumptions affect the model's $\delta^{13}C$ value for the present atmosphere value.

Unless indicated otherwise in the table (with either Note b or Note c), the model uses the step function for the carbonate deposition. A^b and A^c are the "Standard" case with the two alternative deposition regimes. The step deposition seems to be the most reasonable since it has almost all of the carbonate deposition occur very early—and thus it has the highest final atmospheric value of the three cases. Also, in all but the indicated cases (Note c), the atmosphere and adsorbed regolith CO_2 are assumed to isotopically equilibrate rapidly. In this standard case, the "Polar Cap" and "Regolith" reservoirs behave very similarly and are, for the most part, interchangeable (in the sense that the model is not sensitive to whether the mass is in the "Polar Cap" reservoir or the "Regolith" reservoir). I have generally put most of the mass in the "Regolith" with only a small bit in the "Polar Cap" reservoir. This is based on the thought that there is very little CO_2 in the southern cap and layered terrains.

The volcanism parameters are not noted in the table. The volcanism and outgassing is set to "default" values but has a minimal effect on the final results. Using either of the volcanic deposition histories and assuming a carbon composition of 4×10^{-8} g-C / g-planet changes the final $\delta^{13}C$ value by less than a hundredth of a per mille. Even increasing the outgassing rate by 3 orders of magnitude results in a

change of only a few tenths of a per mille—well below the error built into the model.

Table 3.2 has groups of model cases where each parameter was varied. B–E vary the size of the polar cap reservoir. F–J vary the amount of carbonates deposited. K–N vary the size of the regolith reservoir. O–T vary the pressure where the atmosphere collapses. U and V are cases with no isotopic exchange between the atmosphere and regolith.

In general, all of the parameters behave as expected. Increasing the size of any of the three reservoirs decreases the final $\delta^{13}C$ value (for example, B or K). As expected, the effect of changing the polar cap size and the regolith size are roughly equivalent– B and L both add $\sim$ 0.1 bars and both have similar $\delta^{13}C$ values. On the other hand, increasing the amount of carbonate deposited has a much smaller effect. It requires increasing the carbonate reservoir size by $\sim$ 0.4 bars to have the same effect as increasing the regolith reservoir size by 0.1 bars.

The effects of changing the pressure at which the atmosphere collapses is less obvious. For the cases (O and P) where the atmosphere starts out collapsed, the results are very similar to the identical cases where the atmosphere collapses as usual (A^b and A^c). The final $\delta^{13}C$ is strongly affected by the timing of the carbonate deposition (and thus the difference between O and P). For cases where the atmosphere collapses during the run, the lower the collapse pressure, the later the collapse. And the later the collapse, the later the step in the carbonate deposition, and the late carbonate generally reduces final $\delta^{13}C$. This is a relatively small effect. Even in the extreme case where the atmosphere does not collapse (equivalent to it collapsing at present) in case T, the effect is less than 10 $^0/_{00}$. For reasonable collapse pressures ($\gtrsim$ 0.5 bar), the atmosphere collapses within the first 500 Myr and the effect of the collapse time is relatively minor.

Case U is the lowest value and represents a case where the regolith does not isotopically exchange with the atmosphere. In this case, after the atmosphere collapses, the regolith reservoir keeps the same isotopic value it had when the collapse occurred ($-18.0\,^0/_{00}$ for this run) and replaces the CO_2 lost from the atmosphere with CO_2 with that isotopic composition. The resulting atmospheric isotopic composition is

then primarily controlled by the equilibrium between the resupplied CO_2 and the fractionation of the CO_2 lost by sputtering. For the early epochs where the fractionation is high, the equilibrium is high, but at present, with the relatively modest fractionation, the $\delta^{13}C$ value is low. This type of model is unlikely because the amount of adsorbed CO_2 is very temperature dependent and can diffuse fairly fast (at least compared to the 1 Myr time-steps the model uses). V is an attempt to push the limits of the other reservoir sizes and allow such a system to end with a value close to the measured values. The controlling factors are the sputtering fractionation (set by the sputtering model) and the isotopic composition of the regolith reservoir when it is created. The later it is created, the higher it is, thus increasing the size of the current regolith reservoir (thus causing the atmosphere to collapse later) is the most effective. It is difficult to move the collapse time much more than a billion years (and the geological history implies that the collapse was probably soon after the beginning of the model). This implies that the "Regolith" reservoir has to rapidly exchange with the atmosphere to prevent a value set early in the planet's history from overwhelming any fractionation since then.

Case W is an attempt to create a model that reaches the current atmospheric $\delta^{13}C$ without using a regolith reservoir. By depositing a bar of carbonate, the final $\delta^{13}C$ value is within the uncertainty. Most of the carbonate is deposited very fast and actually overwhelms the sputtering loss, causing the atmospheric $\delta^{13}C$ to become lighter. It reaches -30.4 $^0/_{00}$ when the atmosphere collapses. After that, the sputtering (offset a bit more than usual due to the still enhanced carbonate flux) enriches the atmosphere to the current value.

Case X is a look at a case where the atmosphere collapses before the model starts and, due to the low expected temperatures, little carbonate is subsequently deposited. As expected from J, and the fact that the major effect of the timing of the collapse is on the timing of the carbonate deposition, the resulting $\delta^{13}C$ is quite reasonable. This is because carbonate, unless it is in large quantities or deposited very late in the the planet's history, has little effect on the $\delta^{13}C$ value of the current atmosphere.

Cases Y and Z are extreme cases to show the effects of sputtering and carbonate

deposition operating alone. In Y, the only process to affect the atmosphere is sputtering and the final result is an extremely enriched atmosphere. Case Z has only the deposition of 0.8 bars of carbonate (the same amount that was lost from Mars in case Y). The rapid decrease in the $\delta^{13}C$ at the end (in Figure 3.4) is due to the use of constant carbonate deposition, but the final value is independent of the depositional pattern (since it is just a distillation result).

3.6 Discussion

The first important result of the $\delta^{13}C$ modeling is that neither carbonate deposition, nor sputtering, can remove significant amounts of CO_2 without fractionating the atmosphere to extreme values. The sputtering can easily balance the fractionation due to carbonate formation but the opposite is less true. Since carbonate appears to require aqueous environments for deposition [*Stephens,* 1995], most of the deposition probably occurs before the collapse of the atmosphere. A small amount might be deposited after that (either in water films on individual grains or in hydrothermal systems). Thus the step model is the most reasonable of the three carbonate models used.

The model atmosphere generally collapses early in the model–this is reasonable based on the geological evidence that most of the fluvial activity occurs in the oldest terrains [*Carr,* 1996]. The early collapse means that most of the carbonate is deposited during the initial part of the model when the atmosphere is relatively unfractionated, and thus it is difficult to balance the fractionation due to sputtering with carbonate formation (see case W).

On the other hand, an adsorbed regolith, or polar reservoir can significantly dilute the effects of the sputtering. By assuming that the current size of the CO_2 reservoir being fractionated by sputtering distillation is larger than just the atmosphere, the fraction of the total reservoir lost is decreased and the f term in the Rayleigh distillation (equation 3.4) is much larger. With 100 mbar of CO_2 adsorbed in the regolith at present, $f \sim 0.1$ instead of $f \sim 0.01$ for the atmosphere by itself. This is why the

effects of changing the size of the current polar cap reservoir or the current adsorbed regolith reservoir are so large. Laboratory experiments [*Fanale and Jakosky,* 1982] indicate that most reasonable regolith compositions will adsorb some CO_2 at Martian temperatures. Thus, given the normal Martian environment, it is quite easy to envision 100 mbar of adsorbed CO_2. This also means that it is not necessary to invoke chance that the sputtering process has left exactly the right amount of CO_2 on the planet at present for the polar caps to be in vapor equilibrium due to the insolation. The polar cap–regolith–atmosphere system acts, on short time scales, to keep the atmospheric pressure in vapor equilibrium [*Fanale et al.,* 1982].

While the model assumes that sputtering occurs unhindered, there is a possibility that, especially during the early part of the model, that Mars had an intrinsic magnetic field that reduced the effectiveness of sputtering [*Kieffer et al.,* 1992, *Stevenson et al.,* 1983, *Hutchins et al.,* 1997]. Within the context of the $\delta^{13}C$ evolution model, the primary effect would be to require a smaller current "Regolith" reservoir. Furthermore, the sputtering fractionation factors are quite sensitive to the distance between the homopause and exobase, as well as the atmospheric temperatures, so it is quite possible for sputtering to be more efficient at fractionating the atmosphere. This would allow even a reduced sputtering rate to significantly fractionate the atmosphere.

The model includes the continuous loss processes, but it ignores the possibility of impact erosion. If comets or asteroids impact Mars with sufficient velocity, they can eject a plume of atmosphere with velocities above the escape energy. Large impactors are potentially capable of removing large amounts of atmosphere [*Walker* 1986]. The majority of such impacts probably occurred during heavy bombardment and the process should only be a minor contribution during the period of interest. Furthermore, impact erosion of the atmosphere is a bulk process and does not fractionate the gases. Thus, even if it did occur during the period modeled, it would not affect the $\delta^{13}C$ value of the model. It would affect the total atmospheric pressure, or if the atmosphere had collapsed, the size of the various reservoirs. The model's initial pressure is chosen to insure the correct ending pressure and is not constrained beyond what is

necessary to match the selected final state. Thus it is quite possible to increase the initial pressure of the atmosphere to account for impact erosion without substantially affecting the model results.

In the model, I assumed that the initial atmospheric $\delta^{13}C$ is the same as the expected Martian mantle value. This is based on modeling that indicates that any early atmosphere would be essentially removed by hydrodynamic escape and impact erosion during heavy bombardment [*Pepin,* 1994]. Thus, in the early stages of the model, Mars will have outgassed a secondary atmosphere with an isotopic composition close to that of the mantle. Furthermore, while I have not done extensive experiments, it is possible to find a set of parameters that allow for $-50\,{}^{0}/_{00} \lesssim \delta^{13}C_{\text{initial}} \lesssim +10\,{}^{0}/_{00}$ to work. This primarily involves modifying the size of the "Regolith" reservoir. If the initial value is very light (possibly due to massive carbonate formation), then the reservoirs are small and the sputtering can drive the process further. On the other hand, if the initial atmosphere is heavy, larger reservoirs can be used to keep the sputtering from fractionating the atmosphere as much.

3.7 Conclusion

Overall, the model of the behavior of $\delta^{13}C$ derived from the simple model of the Martian carbon "cycle" is very reasonable. The model matches the current $\delta^{13}C$ value with reasonable parameter choices. While it cannot tightly constrain the sizes of the various reservoirs, the model does set some limits. Unfortunately, the $\delta^{13}C$ of the atmosphere is not sensitive to the amount of carbonate deposited but primarily reflects the extent to which sputtering has Rayleigh distilled the carbon. And thus it primarily measures the size of carbon reservoirs that are in equilibrium with the atmosphere. Within the context of Mars, these are probably an adsorbed regolith reservoir and polar deposits—both the southern polar cap and possibly ground CO_2 ices.

The model indicates that the sputtering loss and associated fractionation calculated by the Monte-Carlo model in chapter 2 can be accommodated without overly

fractionating the atmosphere. And furthermore, a certain amount of sputtering appears to be necessary to enrich the carbon to the degree seen. A modest ($\sim$ 100 mbar) amount of CO_2 needs to be stored in a reservoir capable of isotopically exchanging with the atmosphere on moderate time scales. The model allows for the formation of moderate amounts of carbonate ($\lesssim$ 0.5 bar) over the last 3.5 Gyr, but does not require such formation.

Putting this modeling work together with the geological evidence and atmospheric thermal budget models for early Mars [*Carr,* 1996, *Haberle,* 1998, *Kieffer et al.,* 1992] allows for a general evolutionary model for the Martian atmosphere. At the end of late heavy bombardment, about 3.5 Gyr ago, the atmosphere was quite thick (at least one bar of CO_2 and possibly 5 or 10 bars). This was sufficient to allow for liquid water (at least on parts of the planet during some seasons) near the surface. The water created the geomorphic features attributed to flowing water seen in images (which mostly seem to date from this time period [*Carr,* 1996]). It also caused significant amounts of carbonate to deposit [*Haberle,* 1998] which depleted the atmosphere to $\sim$ 1 bar. At this point, the atmospheric greenhouse collapsed catastrophically and rapidly reached a state very similar to the current one, but with large volatile inventories in the regolith and polar caps. Under the new conditions, carbonate formation is suppressed (no liquid water) and sputtering becomes the dominant mechanism for removing CO_2 from the planet. As the sputtering occurs, it fractionates the atmosphere, preferentially enriching it in ^{13}C and raising the $\delta^{13}C$ value to the that of the present atmosphere. Over the last 3.5 Gyr or so, the sputtering has removed much of the atmosphere left from the collapse, but not all of it. This leads to the current state where there is a small regolith and polar reservoir remaining from the atmosphere that collapsed. And the regolith, coupled with the polar caps, buffers the current atmosphere at its vapor equilibrium of $\sim$ 7 mbar.

Appendix 3.A Computational Model

The computational model used to study the evolution of the atmosphere is fairly simple. This is primarily a description of the actual model used. The issues and choices involved are discusses in section 3.4.

The model is basically a box model. The model only has two boxes and tracks the mass and $\delta^{13}C$ of each box. The mass terms evolve actively due to the changing conditions while the $\delta^{13}C$ changes in response to the mass changes. The first box is the atmosphere and the second box is the "Regolith" reservoir. In addition to exchange between these two reservoirs, it has three other fluxes that all affect the atmospheric box. The three fluxes are sputtering loss to space, carbonate deposition and outgassing from the interior.

3.A.1 Mass Calculation

The model uses an explicit Euler method for the timestep. The governing equation for the atmospheric mass is:

$$M_a(t) = M_a(t-1) + \Delta t \left[F_s(t) + F_c(t) + F_m(t)\right] + E(t) + E_p(t) \tag{3.5}$$

where $M_a(t)$ is the atmospheric mass at time t. $F_s(t)$, $F_c(t)$, and $F_m(t)$ are the fluxes due to sputtering, carbonate formation, and mantle outgassing respectively. Note that F_s and F_c are always negative since they represent atmospheric loss, while F_m is always positive. $E(t)$ is the exchange with the "Regolith" reservoir. $E_p(t)$ is the exchange with the "Polar Cap" reservoir. Δt is the size of the timestep and was 1 Myr for all cases studied. The time step was chosen to insure that the model remained numerically stable despite the potential instability of the explicit method. Throughout the rest of the appendix, the fluxes will be normalized to the timestep (as in the numerical model). The result of equation 3.5 is the solid line in Figure 3.5.

The equation for the mass of the "Regolith" reservoir is much simpler:

$$M_r(t) = M_r(t-1) - E(t) \tag{3.6}$$

where $M_r(t)$ is the mass of the "Regolith" reservoir. It is the dash dotted curve in Figure 3.5.

$F_s(t)$ and $F_m(t)$ are specified explicitly in the model. The values for $F_s(t)$ are the dashed line in Figure 3.6 and are from the results of chapter 2. This is the same result as curve (e) in Figure 2.12 with a different scaling. $F_m(t)$ is negligible, but was taken from Hutchins and Jakosky [1996].

Instead of specifying initial conditions, the model is designed to allow the user to specify final (or present—t_p) conditions for most parameters. This includes the final size of the "Regolith" reservoir $M_r(t_p)$, the final size of the carbonate reservoir ($C_{\text{tot}} = -\sum_t F_c$), the size of the "Polar Cap" reservoir M_p, and the size of the current atmosphere ($M_a(t_p) = 7$ mbar). These values are then used to calculate the correct initial conditions.

Initially the "Regolith" reservoir is empty and $M_r(t_0) = 0$ since the planet is assumed to be too warm for there to be adsorbed or condensed CO_2. All of the CO_2 is assumed to start in the atmosphere:

$$M_a(t_0) = M_a(t_p) + M_r(t_p) + C_{\text{tot}} + M_p - \sum_{t_0}^{t_p} (F_s + F_m) \tag{3.7}$$

Note that the term for the sputtering flux adds mass to the initial atmosphere while the flux from the mantle decreases the initial mass.

In addition to calculating the initial values, the model also calculates the time when the atmosphere will catastrophically collapse (t_c). This is based on the selected collapse pressure (M_c). The carbonate flux F_c is selected so that 95% of the carbonate is deposited before the collapse. Thus the atmosphere collapses when

$$M_a(t_c - 1) > M_c \geq M_a(t_c) = M_a(t_0) + \sum_{t_0}^{t_c} (F_s + F_m) - .95\, C_{\text{tot}} \tag{3.8}$$

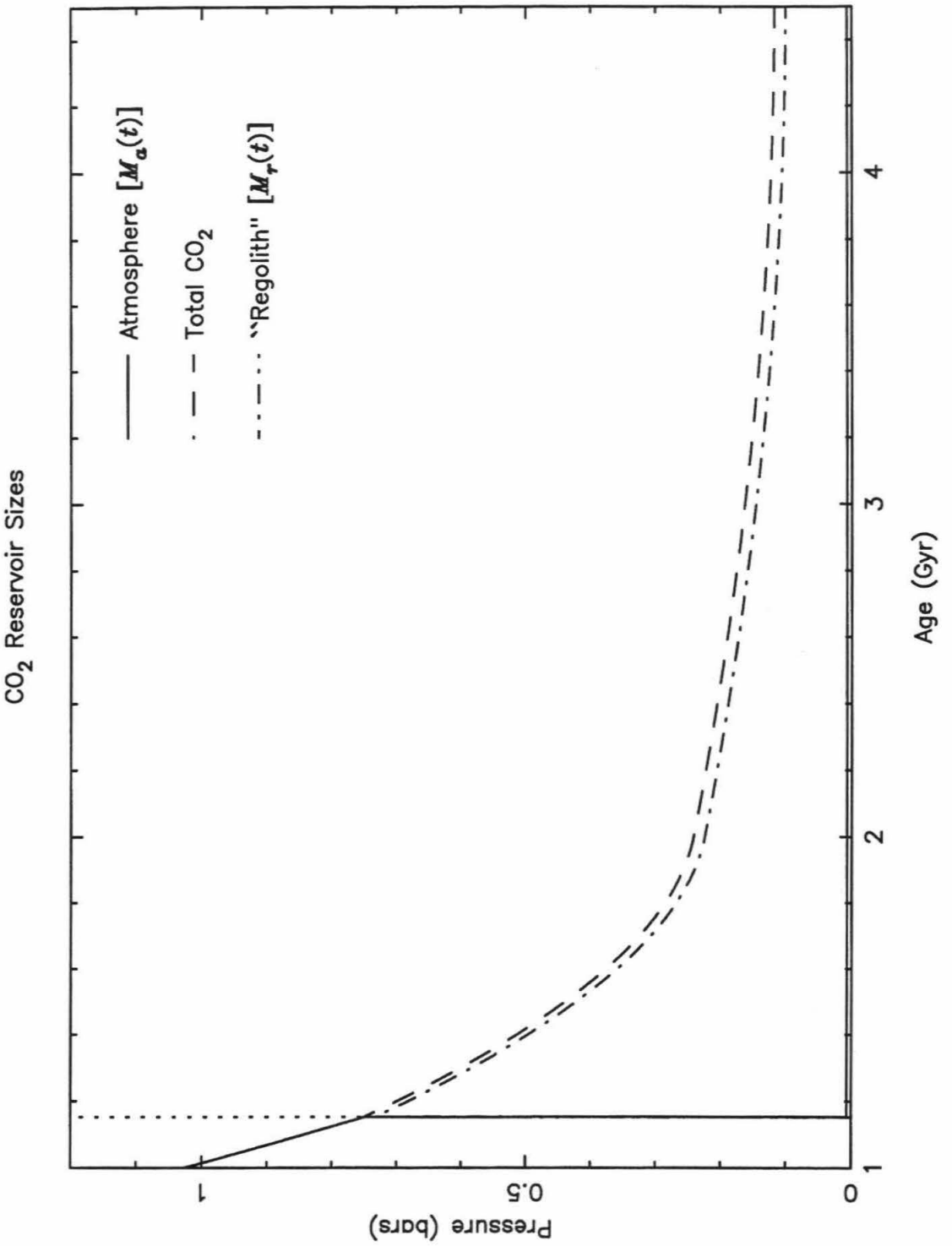

Figure 3.5: Size of the Model Reservoirs
This is for the "Standard" model. The solid curve is the atmospheric pressure (in bars) over the course of the model. The dashed curve is the total CO_2 in the atmosphere-cap-regolith system over the course of the model. The dash-dotted line is the amount of CO_2 in the "Regolith" reservoir.

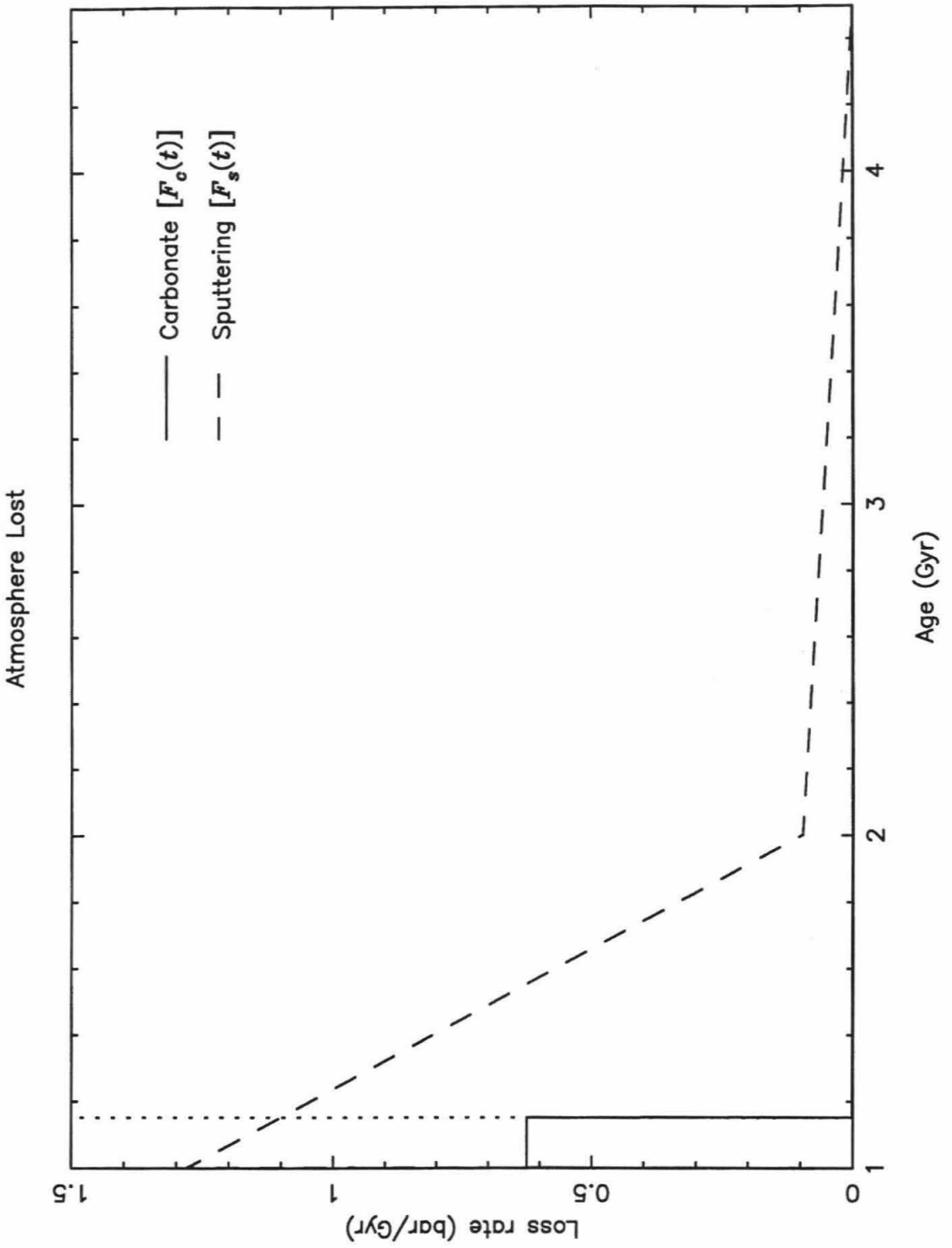

Figure 3.6: Loss Rates from the Atmosphere
The rate at which various model processes remove atmosphere in the "Standard" model. The dashed curve is the loss rate due to sputtering. The solid curve is the loss rate for carbonate. The step in it is defined to occur at the time of collapse. The carbonate loss rate will change depending on the model parameters to insure this.

t_c is explicitly found by searching, starting with t_0.

F_c and $E(t)$ are calculated dynamically by the model based on the parameters selected. As mentioned in section 3.4, F_c is a step function with a constant value before and after collapse. Thus

$$\begin{aligned} F_c(t \leq t_c) &= .95\, C_{\text{tot}}/(t_c - t_0) \\ F_c(t > t_c) &= .05\, C_{\text{tot}}/(t_p - t_c) \end{aligned} \tag{3.9}$$

In some cases, the simpler carbonate fluxes were used. One case (note a) was a constant carbonate rate $F_c = C_{\text{tot}}/(t_p - t_0)$, the other (note b) was proportional to F_s. In that case,

$$F_c(t) = F_s(t)\frac{C_{\text{tot}}}{\sum_{t_0}^{t_p} F_s} \tag{3.10}$$

$E(t)$ is an implicitly defined discontinuous function that creates the "Regolith" reservoir when the atmosphere collapses and then keeps the atmosphere at the condensation equilibrium pressure for the rest of the history of Mars. The equation thus has the following form:

$$\begin{aligned} E(t < t_c) &= 0 \\ E(t_c) &= M_a(t_c - 1) + F_s(t_c) + F_c(t_c) + F_m(t_c) - M_a(t_p) \\ E(t > t_c) &= -F_s(t) + F_c(t) + F_m(t) \end{aligned} \tag{3.11}$$

Since E_p is zero except at t_c (when the "Polar Cap" reservoir is formed), M_a is constant for $t > t_c$ as shown in Figure 3.5. While M_a does not change after t_c, the actual CO_2 in the atmosphere (as well as the total CO_2—the dashed line in Figure 3.5) do change and drive isotopic evolution.

3.A.2 δ^{13}C Calculation

In some ways, the δ^{13}C calculations are simpler than those for mass since the δ^{13}C just reacts to the changes in mass. Unfortunately, the isotopic system is non-linear so the governing equations are more complex. The isotopic effects can be broken

into two categories. The first are loss processes which are assumed to be Rayleigh distillation processes. The second group are ones that add mass and are treated as mixing processes.

In the model, the "Polar Cap" reservoir (M_p) is always assumed to be in isotopic equilibrium with the atmosphere. Thus the effective atmosphere size for isotopic equilibrium is $M_a(t) + M_p = M'_a(t)$ for convenience. In the period before the atmosphere collapses ($t \leq t_c$), the "Polar Cap" reservoir is part of the atmosphere and $M_p = 0$.

The two loss processes are the sputtering and the carbonate formation. They use the following form of equation 3.4:

$$\delta_a(t) = \delta_a(t-1)\left[1 - \frac{F_s(t)}{M'_a(t-1)}\right]^{(\alpha_s(t)-1)}\left[1 - \frac{F_c(t)}{M'_a(t-1) - F_s(t)}\right]^{(\alpha_c(t)-1)} \tag{3.12}$$

Where $\delta_a(t)$ is the δ^{13}C value of the atmosphere. $\alpha_s(t)$ is the sputtering fractionation coefficient and $\alpha_c(t)$ is the carbonate formation coefficient (see section 3.3 for further discussion). For computational simplicity, the two fractionation processes are actually performed serially (thus the extra term in f for the carbonate fractionation). Because of the small timestep, the amounts lost are small and this is a sufficiently good approximation. When the collapse occurs, both E and E_p are nominally loss processes as well. Since the model assumes that condensation is not fractionating [*Eiler and Kitchen,* 1999], they do not affect the isotopic value.

Like the loss processes, the mixing processes are performed serially (and after the loss distillation). For simplicity, $\delta' = \delta + 1000$ [*Faure,* 1991]. t' is the value at this timestep after the previous processes have taken effect. The general form of the mixing equation is:

$$\delta_a(t) = \frac{(R_{\text{std}}\,\delta'_n(t')+1)\,\delta'_a(t')\,M'_a(t') + (R_{\text{std}}\,\delta'_a(t')+1)\,\delta'_n(t')\,M_n}{R_{\text{std}}\left[(\delta'_a(t')+1)\,M_n + (\delta'_n(t')+1)\,M'_a(t')\right]} - 1000 \tag{3.13}$$

where M_n is the amount mixed into the atmosphere and $\delta'_n(t') - 1000 = \delta_n(t')$ is its δ^{13}C value. This is basically an average of the δ^{13}C values weighted by the size of each reservoir, but it accounts for the way δ is defined (equation 3.1).

The first mixing process is the outgassing. It uses equation 3.13 with $M_n = F_m(t)$ and $\delta_n(t') = \delta_m = -25\,{}^0/_{00}$, the δ^{13}C value of the mantle. t' is after performing the loss processes so $M'_a(t') = M_a(t-1)+F_s+F_c+M_p$ and $\delta_a(t')$ is the result of equation 3.12. Due to the extremely small value of F_m, the effect on δ_a is unnoticeable.

When the atmosphere collapses, and the "Regolith" reservoir is created, its δ^{13}C value is set to the atmospheric value ($\delta_r(t_c) = \delta_a(t_c)$). In the "Standard" model, it is assumed to isotopically equilibrate with the atmosphere at each timestep after the collapse. This is calculated after all the other fluxes have affected the atmosphere. It uses equation 3.13 with $\delta'_n(t') = \delta'_r(t-1)$ and $M_n = M_r(t-1)$ so that the entire reservoir equilibrates with the atmosphere. Here, t' is the value after including outgassing and $M'_a(t') = M_a(t-1)+F_s+F_c+F_m+M_p$. Since it is assumed that there is no isotopic fractionation during condensation and adsorption, $\delta_r(t) = \delta_a(t)$. In cases where the atmosphere and "Regolith" reservoir do not mix, M_r is replaced by E (and δ_r does not change over time).

It is not useful to express the δ^{13}C calculation as a single step since it is performed serially—the complexity would just obscure the structure of the operation. The mass calculation is also done serially, but in that case the combined equation is simple. Furthermore, it is equivalent to the simultaneous calculation. It is possible to do the two distillation processes simultaneously (the model almost does so in its current form). It is also possible to simultaneously do both mixing processes (although much more complicated mathematically). It is not computationally feasible to perform the mixing and loss processes simultaneously. Since the timestep is sufficiently small, the integrated flux in any timestep is a small fraction of the effective atmosphere and serial calculations are an adequate approximation.

Bibliography

Carr, M. H. *Water on Mars,* 229 pp., Oxford University, New York, 1996.

Clark, R. N., T. M. Hoefen, R. Pearson, N. Gorelick, P. R. Christensen, TES Science Team, Low Spatial Resolution Mars Global Surveyor Thermal Emission Spectrometer Imaging Spectroscopy Maps of Mars: A Global Perspective, *Bul. Amer. Astro. Soc., 30,* 1049, 1998.

Eiler, J. M., Kitchen N., Experimental Study of the Stable-Isotope Systematics of CO_2 Ice/Vapor Systems and Relevance to the Study of Mars, *LPSC Abstract,* 1999.

Fanale, F. P., B. M. Jakosky, Regolith–Atmosphere Exchange of Water and Carbon Dioxide on Mars: Effects on Atmospheric History and Climate Change, *Planet. Space Sci., 30,* 819-831, 1982.

Fanale, F. P., J. R. Salvail, W. B. Banerdt, R. S. Saunders, Mars—the Regolith-Atmosphere-Cap System and Climate Change *Icarus, 50,* 381-407, 1982.

Faure, G., *Principles and Applications of Inorganic Geochemistry,* 626 pp., Macmillian Publishing, New York, 1991.

Haberle, R. M., Early Mars Climate Models, *J. Geophys. Res., 103,* 28467-28479, 1998.

Hunten, D. M., Atmospheric Evolution of the Terrestrial Planets, *Science, 259,* 915-920, 1993.

Hutchins K. S., B. M. Jakosky, Evolution of Martian Atmospheric Argon: Implications for Sources of Volatiles, *J Geophys. Res., 101,* 14933-14949, 1996.

Hutchins, K. S., B. M. Jakosky, J. G. Luhmann, Impact of a Paleomagnetic Field on Sputtering Loss of Martian Atmospheric Argon and Neon, *J. Geophys. Res., 102,* 9183-9189, 1997.

Jakosky, B. M., Mars Volatile Evolution—Evidence from Stable Isotopes, *Icarus, 94,* 14-31, 1991.

Jakosky, B. M., E. S. Barker, Comparison of Groundbased and Viking Orbiter measurements of the Martian Water Vapor: Variability of the Seasonal Cycle, *Icarus, 57,* 322-334, 1984.

Jakosky, B. M., R. O. Pepin, R. E. Johnson, J. L. Fox, Mars Atmospheric Loss and Isotopic Fractionation by Solar-Wind-Induced Sputtering and Photochemical Escape, *Icarus, 111,* 271-288, 1994.

Kieffer, H. H., B. M. Jakosky, C. W. Snyder, M. S. Matthews (Eds.), *Mars,* 1498 pp., The University of Arizona, Tucson, 1992.

Leighton, R. B., Murray, B. C., Behavior of Carbon Dioxide and other Volatiles on Mars, *Science, 153,* 136-144, 1966.

Pepin, R. O, Evolution of the Martian Atmosphere, *Icarus, 111,* 289-304, 1994.

Stephens, S. K., Carbonate Formation on Mars: Experiments and Models, Ph.D. Thesis, California Institute of Technology, 1995.

Stevenson, D. J., T. Spohn, G. Schubert, Magnetism and Thermal Evolution of the Terrestrial Planets, *Icarus, 54,* 466-489, 1983.

Walker, J. C. G., Impact Erosion of Planetary Atmospheres, *Icarus, 68,* 87-98, 1986.

Ward, W. R., Long–Term Orbital and Spin Dynamics of Mars, in *Mars,* edited by H. H. Kieffer, B. M. Jakosky, C. W. Snyder, and M. S. Matthews, 298-320, The University of Arizona, Tucson, 1992.

Wright, I. P., M. M. Grady, C. T. Pillinger, The Evolution of Atmospheric CO_2 on Mars—The Perspective from Carbon Isotope Measurements, *J. Geophys. Res., 95,* 14789-14794, 1990.

Zhanle, K. J., The Case Against Cometary Origin of the Oceans, *Eos Trans. AGU, 78,* F542, 1997.

Zhang, M. H. G., J. G. Luhmann, S. W. Bougher, A. F. Nagy, The Ancient Oxygen Exosphere of Mars—Implications for Atmosphere Evolution, *J. Geophys. Res., 98*, 10915-10923, 1993.

Chapter 4 Data Assimilation Methodology

4.1 Introduction

On Earth, there is a large number of meteorological stations as well as several weather satellites in various orbits, giving good global coverage of the planet. On Mars, this is not the case. At best there have been two meteorological stations and two spacecraft. And during the current Mars Global Surveyor (MGS) mission, there is only one spacecraft. This makes it much harder to examine the atmospheric dynamics because the spacecraft only samples one location at a time. In order to help solve this problem we developed a data assimilation technique that allows us to extrapolate the limited available data to a global view of the planet. This will allow us to study global issues involving Martian dynamics.

The global dynamical fields can be used to obtain climatological mean values for the season of the data. This can then be compared to models of the Martian atmosphere and improve our understanding of the climatology of Mars. An understanding of the climatology then allows the weather itself to be studied. By studying the weather, we not only get the mean state of the atmosphere but range of variability and change around that mean state.

Furthermore, since the Martian atmosphere is, in many ways, simpler than the terrestrial atmosphere, it is a useful dynamical system to study and use to test our understanding of the dynamics of terrestrial planet atmospheres. While there are similarities between the Earth's atmosphere and that of Mars, there are also significant differences, and a global view of the meteorology can help indicate processes that are relevant for Mars, but which do not play a role on Earth.

This chapter discusses the method implemented to successfully assimilate the

MGS data. The next section discusses the background, then section 4.4 describes the method implemented and finally the latter part of the chapter covers several tests used to show the accuracy of the method. Chapter 5 is devoted to the results of the actual assimilation.

4.2 Theoretical Background

The basic idea behind meteorological data assimilation is to merge observations with a model. The goal is to have the resulting state be better than either the model or the data. The data improves the ability of the model to reflect the real atmosphere and the model allows the data to be extrapolated to cover the entire planet at all times.

This is generally done by using the observations as a guide for modifying the state of the model. This can equivalently be considered dynamically interpolating between the observations. There are two different states of the model at each time. The first state is the prediction state and is the result of running the model by itself. The length of the "free running" is variable and can result in different prediction states for any given time—for example a three–day prediction (i.e., a model run for three days without assimilation) will be different from a three–hour one. The extent to which predictions, made for the same time, will differ depends on how accurate the assimilation is and the ability of the model to track the weather. The second state is the analysis state and is the result of the full assimilation procedure. This is the system's best guess at what the state of the atmosphere was at that point in time. The observations can be called a third state but are often not on the grid of the model, nor do they usually cover the entire model domain.

There are several different approaches and techniques that can be used for data assimilation. The ones used for a given assimilation project generally reflect the model being used and the type of data available. In our assimilation, we used the following

basic assimilation equation [*Banfield et al.,* 1995]:

$$\Psi_k^a = \Psi_k^p + \sum_{j=1}^{N} \alpha_{kj} \left(\Psi_j^o - \Psi_j^p \right) \tag{4.1}$$

where Ψ_k is an element of the state vector (at point k). The superscript denotes the state vector in question, a for the analysis state, p for the prediction state, and o is for the observations. The α_{kj} are the weighting coefficients (or gain functions) specifying how the discrepancy between the observations and the model predictions at point j affect the state variable at point k. N is the total number of observations that will be assimilated during this update of the model. In our case, this is usually one time-step—about 6 minutes.

There are a broad class of assimilation techniques that use equation 4.1. They differ in the gain α_{kj} used. We based our technique on Optimal Interpolation (OI) and steady state Kalman Filtering [*Banfield et al.,* 1995]. In OI, the gains are generated to minimize the analysis errors resulting from the merging of the data and model [*Bengtsson and Gustavsson,* 1971, *Ruthford,* 1972, *Bergman,* 1979]. Two assumptions are made in deriving the gains for Optimal Interpolation. First that the physical system (and observing system making the measurements) are statistically steady and secondly that the observational error are uncorrelated with the prediction errors. For a full derivation of OI, see Rutherford [1972]. In order to minimize the error in the analysis state vector, the assimilation gains for point k are the solution to the following system of equations:

$$\sum_{j=1}^{N} \left(\overline{\epsilon_i^p \epsilon_j^p} + \overline{\epsilon_i^o \epsilon_j^o} \right) \alpha_{kj} = \overline{\epsilon_k^p \epsilon_i^p} \tag{4.2}$$

where there is one such equation for each observation i. ϵ_i^p is the prediction error at point i, and ϵ_i^o is the observational error at i. The over-bars indicate averaging over many different states. This basically describes the necessary relationships between the various covariance terms of the errors at the relevant points.

The equations for the gains and the assimilation are general. They apply to the

assimilation of any observation into any point of the state vector. Thus both j and k can range over the entire state vector. Efficient, but sub-optimal implementations of OI are obtained by assuming that α_{jk} is compact. Under that assumption, an observation only affects nearby points and often only affects the field actually measured (thus temperature measurements only correct the model temperature near the actual measurement). The errors are also often assumed to have a simple Gaussian distribution around the observation point. Because we have so little data (relative to the size of our state vector), we use an implementation where k is global and thus we correct the temperature everywhere based on one measurement, and furthermore, we also correct the pressure and wind velocities using the same measurement. In order to do this, we calculate the actual covariances to determine the gains.

4.3 Assimilation Resources

The data available for assimilation are the atmospheric retrievals from the Thermal Emission Spectrometer (TES) on the Mars Global Surveyor (MGS) spacecraft orbiting Mars [*Christensen et al.*, 1992, *Conrath et al.*, 1998]. The data of interest are temperature profiles (temperature at a set of fixed pressures) that are obtained by inverting the structure of the 15 μm CO_2 line. We did not do the inversion and thank John Pearl, Barney Conrath and the rest of the TES Team for the processed data.

Unfortunately, due to structural problems with the spacecraft, MGS was unable to rapidly enter into the sun synchronous orbit it was designed for. Instead, it spent the first year and a half in a highly elliptical orbit. The net result was an unusual TES observing pattern (Figure 4.1). Each orbit can be broken into three components. The first component (in red in the figure) is the close pass to the planet during the periapse. In this phase, the spacecraft is close to the planet and takes a line of data with a high spatial density. The second component (in blue) is the roll out and roll in before and after the near pass. In this case the data is a small area with a fairly high sampling. The third section of the data is in green and is the data from the rest of the orbit while the spacecraft is at a significant distance from Mars. During this period,

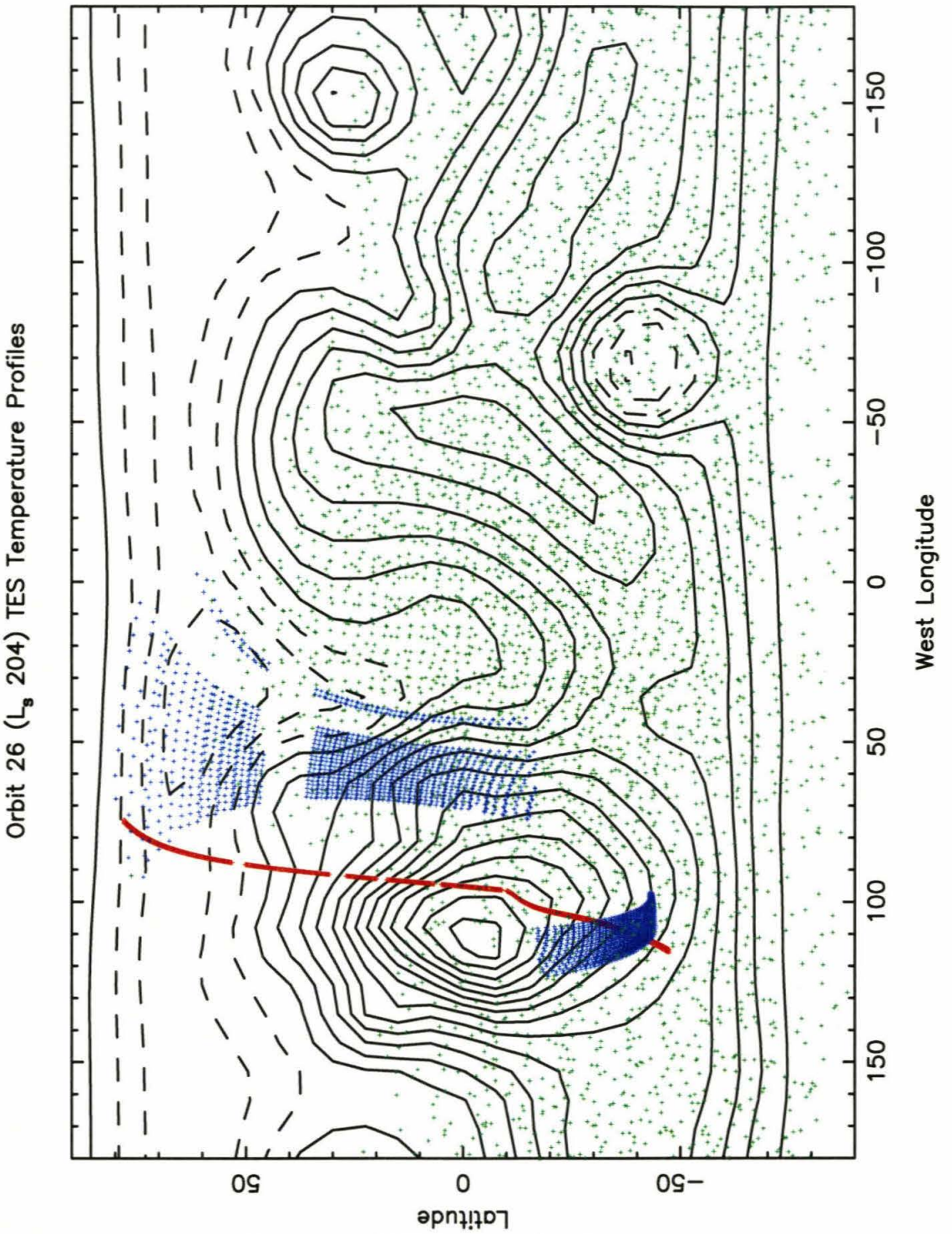

Figure 4.1: TES Observation Pattern

The usable observations made by TES during one orbit (orbit 26). The observations have been grouped by color. The red ones are the close pass of the planet. The green ones are long range observations. The blue observations are from just before and after the close pass. The background contours, at 3 km intervals, are the topography of Mars at the resolution of the GCM. Low contours are dashed.

the spacecraft rolls and Mars is in the instrument's field of view during a few minutes every hour and a half. During that time, the combined roll and instrument pointing allow it to perform a crude raster scan across the disk. The collection of points are the result of several of these scans. While the spacecraft spends the majority of its time in the green phase, most of the data is collected during the close pass (red+blue). This observation pattern greatly increased the difficulty because the method has to be able to handle periods when there are long continuous streams of very tightly spaced data as well as periods of no data interspersed with short bursts of data covering significant portions of Mars.

The second component necessary for the assimilation process is a global model of the atmosphere. Because of its robustness and availability, we used the Ames Mars GCM (General Circulation Model) [*Pollack et al.*, 1990, *Haberle et al.*, 1993]. We thank Ames team for giving us access to the MGCM code and assisting us with understanding how it operates.

The Ames Mars GCM (MGCM) is a low resolution GCM based on the primitive equations and a hydrostatic atmosphere. It uses a finite difference technique with, in our experiments, 25 latitude grid points and 40 longitude grid points, for a horizontal resolution of 7.5×9 degrees (see the topography in Figure 4.1). It has 13 layers in a σ-coordinate system with the "top" of the dynamical atmosphere being ~ 0.067 mbar. We were primarily interested in the four dynamical fields (surface pressure, temperature and the zonal (U) and meridional (V) components of the wind fields) that the model calculates at each time-step. It also tracks many other variables (such as radiative heating and surface frost) that allow it to model the progression of the seasons on Mars. The model calculations are entirely deterministic based on the starting time for the model run. The model starts a run from a cold start but can be restarted from intermediate conditions. The GCM uses a leapfrog method to advance in time with 6 minute time-steps and it smoothes every fifth step (to keep the even and odd solutions from diverging). The GCM output includes the full dynamical fields every hour and a half (note that the GCM runs on Martian time and this is actually every $1/16^{\text{th}}$ of a sol). The GCM is designed to run in 10 sol increments and

string them together using the final conditions from one as the initial conditions of the next one to create longer sets of results.

4.4 Assimilation Method

In implementing an assimilation scheme, there are three main constraints that need to be considered. The first is that the method needs to be successful on the data it is designed to assimilate. It needs to correct the model and reduce the analysis error as much as possible for the available data. Secondly, it needs to be computationally feasible. This is both in terms or running in a reasonable amount of time as well as in using a reasonable amount of memory. In many ways, these are complementary in that reducing the amount of memory reduces the computation by requiring operations on fewer data points. The final, but still critical, requirement is that it needs to match the theory.

In Optimal Interpolation and its variants, the data assimilation process can be divided into two steps. The first is the calculation of the gains (or weights) used for the assimilation process—the α_{kj} in equation 4.1. This includes gathering the necessary statistics and the calculation of the actual gains. The second step is the actual assimilation. In this step, the differences between the model and the observations need to be calculated and multiplied by the gains to modify the model state vector. The two steps can be done interleaved (required if the gains are calculated during the assimilation) or they can be done separately. Due to the number of gains, we do the two steps independently and calculate the gains off line before doing the actual assimilation (see Figure 4.2).

Due to the continuous stream of data (at least during parts of the spacecraft orbit), we use a scheme where data is assimilated every time-step. The process (illustrated in Figure 4.2) starts by reading in the data for the current time-step. It then calculates the differences between the model and observations. Next it uses the necessary gains to update the GCM state vector. Finally, it has the GCM step the model forward one time-step to create the prediction state for the next time-step. This continues

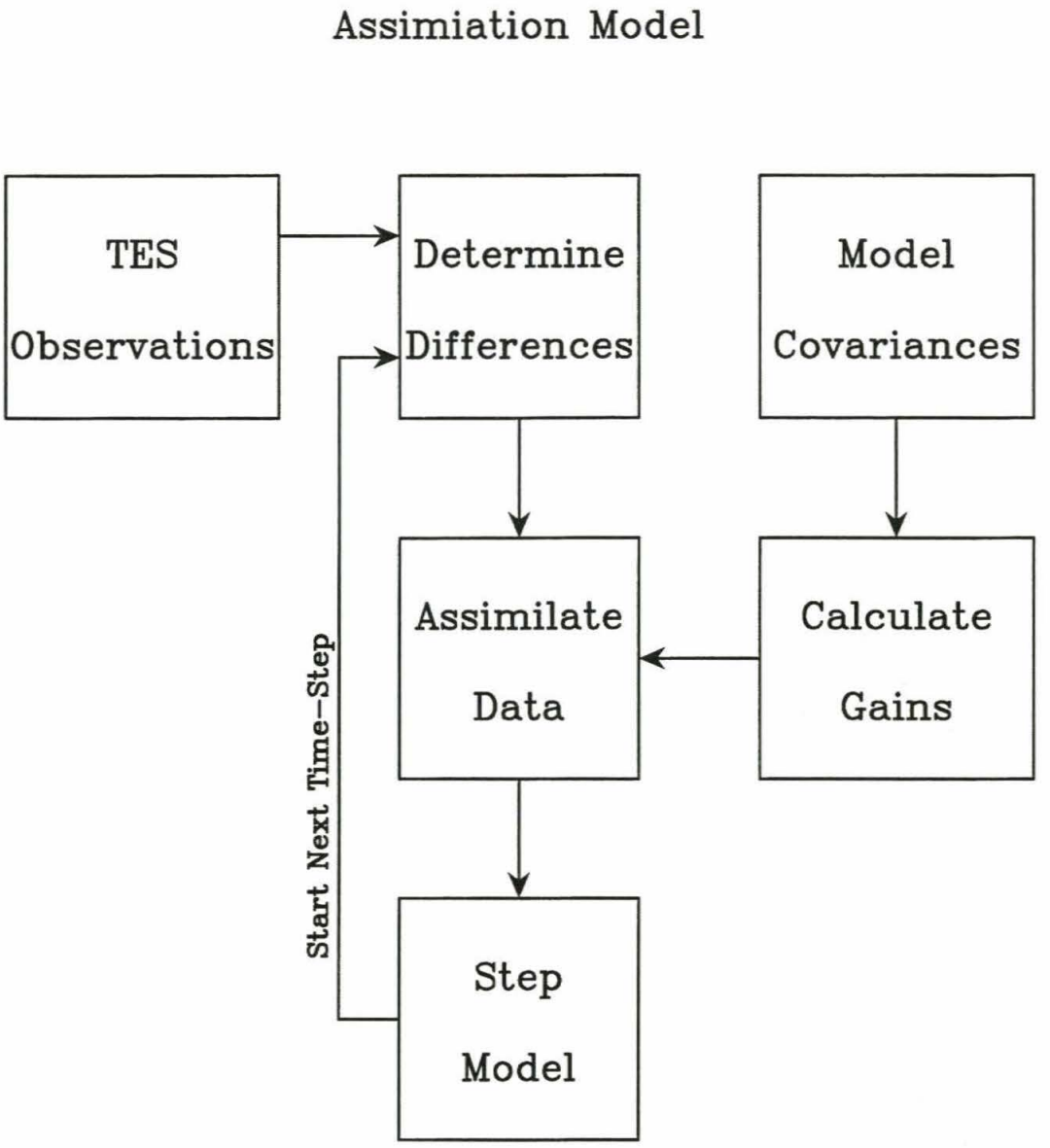

Figure 4.2: Assimilation Flow Chart
This is a continuous process and represents one time-step in the model. The gains and data for the time-step are read in from external sources and used during this time-step.

until the model is stopped or it runs out of data.

4.4.1 Gain Calculation

There are several problems with directly implementing the full Optimal Interpolation equation (4.2) for calculating the gains. First, the equation implies that the gains for assimilating an observation depends not only on that observation but on all other observations being assimilated with it (through the system of equations and error covariance terms). Secondly, it requires knowing the errors (difference between the actual atmosphere and the model or observations) which are not known (otherwise there would be no point to the assimilation). This system, as described, is extremely large, since for each observation, there is one gain per point in model state vector. In order to gather the statistics, this requires calculating the error covariances for each pair of points in the model.

In calculating the gains, we made several assumptions to simplify the process. The primary goal of the assumptions was to allow the gains to be calculated off-line. Like that they only need to be calculated once and the expensive matrix inversions do not need to be done each time-step. Furthermore, if precalculated, only the gains need to be used by the model; it does not need to store the covariance values while actually assimilating.

The first assumption is that each vertical profile is assimilated alone. Thus the assimilation of a given profile does not affect how other profiles are assimilated. This is equivalent to assuming that equation 4.2 is block diagonal—the only non-zero, non-diagonal terms relate to other points in the same profile. This converts equation 4.2 into M smaller independent matrices:

$$\sum_{j=1}^{P_l} \left(\overline{\epsilon_i^p \epsilon_j^p} + \overline{\epsilon_i^o \epsilon_j^o} \right) \alpha_{kj} = \overline{\epsilon_k^p \epsilon_i^p} \tag{4.3}$$

where M is the number of vertical profiles in the time-step and P_l is the number of points in the l^{th} profile.

This transformation encompasses two implied points. It assumes that the obser-

vation error between different profiles is uncorrelated. Like this, $\overline{\epsilon_i^o \epsilon_j^o} = 0$ if i and j are in different profiles. This is a reasonable assumption for the TES instrument. Also, previous work shows that even moderate inter-observational correlations do not degrade OI assimilation schemes much [*Banfield et al.*, 1995]. Secondly, it ignores the prediction error covariance between the locations of the two profiles ($\overline{\epsilon_i^p \epsilon_j^p} = 0$ when i and j are in different profiles). Mathematically, these terms tell the assimilation process how many observations are being assimilated and where the observations are in relation to each other. The former effect is the dominant portion of the process. In a simple system where there is only one grid location, but multiple observations at that location, the effect of the cross terms is to weight each coefficient by the total number of observations being assimilated. We recover this portion of the term later in the process by dividing all the differences by the total number of profiles in the time-step and can thus ignore the term here.

With the first assumption, the only variable portion of the gains is the number and location of the observations within each profile. Because all the observations in a single profile are actually from the inversion of a single spectral feature, with overlapping weighting functions, they are not uncorrelated and $\overline{\epsilon_i^o \epsilon_j^o} \neq 0$. While we cannot diagonalize the set of equations within each profile, it is possible to use further assumptions to make them time independent. Since the assimilation process moves the observation to the nearest vertical grid point (after calculating the difference using an interpolated field value), the model only actually needs gains at the 13 model layers. In order to make the gains time independent, we assume, while calculating them, that each profile has exactly 13 points and that each is at one of the model layers. This is equivalent to assuming $P_l = 13$ in equation 4.3 for both i and j. While not perfect it is a decent representation of the observations. From the actual data, the profiles average 15 points (and none exceed 20). Since each observation is on a discreet pressure surface, they spread out well within the GCM altitude range (many of the profiles extend beyond the top of the GCM upper boundary, but such points are discarded). We still assimilate each point in the profile (and thus N in equation 4.1 is the number of actual observations), but the gain (α_{kj}) is calculated with $P_l = 13$

and our assumed profile, ignoring the real pattern of observations.

Combined, the assumptions that each profile is independent and contains exactly 13 points, one per layer, allow us to calculate the gains off-line while preserving many of the features of the full Optimal Interpolation gains. Thus, for each point k in the GCM state vector and each possible profile location (each distinct latitude and longitude in the GCM) we solve the matrix equation:

$$\sum_{j=1}^{13} \left(\overline{\epsilon_i^p \epsilon_j^p} + \overline{\epsilon_i^o \epsilon_j^o} \right) \alpha_{kj} = \overline{\epsilon_k^p \epsilon_i^p} \tag{4.4}$$

where each value of i and j corresponds to one of the GCM layers. Since we don't know, in advance where the observations will be, we calculate the gains for observations over the entire GCM domain. Unfortunately, the actual dependence of the observational noise ($\overline{\epsilon_i^o \epsilon_j^o}$) within the profiles is not available, so we generated a simple synthetic noise pattern. We assumed that each point affects two neighbors to either side (as well as itself) with a Gaussian distribution (using the probability for the ranges [-0.5,0.5], [0.5,1.5], and [1.5,∞), and the corresponding negative intervals).

In order to reduce the number of gains that were needed, we assumed, based on work by Banfield *et al.* [1995], that the gains are longitudinally independent. We only consider the relative longitudinal separation between the observation and the point being corrected. Thus assimilating an observation with a longitude of 120 degrees into a point at 90 degrees is equivalent to assimilating an observation at 60 degrees into a point at 30 degrees (assuming the altitudes and latitudes all match). Like this we don't need gains to assimilate observations at all longitudes but can use one set regardless of the longitude of the observation (and then calculate the relative distance for each location). This reduces the number of gains by 40 (the number of longitudinal GCM grid points) and allows us to actually store sufficient gains in memory to run the assimilation model.

While we do not worry about the longitude, we do consider the local time of day. The relation between a measurement at night and the model is different than the relationship between a daytime measurement and the model. While it would be

ideal to have gains for every local time of day, this was not possible due to memory constraints. We calculate two sets of statistics, and subsequently two sets of gains—one at 2 am and the other at 2 pm. These times were chosen primarily because the eventual circular orbit will be inertially fixed at these local times. They also capture the significant differences between day and night. Observations are assimilated using whichever of the two gains is closer (in local time).

The gains require the covariances to be measured over the entire globe. Because the actual error values are not available, it is necessary to estimate what they are. As described above, we use a simple Gaussian to estimate the observational covariances. The prediction covariances (especially the $\overline{\epsilon_k^p \epsilon_i^p}$ term) are what drive the assimilation process. They encode the weather variability that tells the process how to correct the model based on the observations. We use a twinned model approach (Figure 4.3). In a twinned model approach, one model is called the "Truth" and the other the "Analysis." The "Truth" is used to generate synthetic observations. For each synthetic observation i and each point in the GCM state vector, the model calculates $\epsilon_k^p \epsilon_i^p$ (the other prediction errors are just subsets). These terms are then used to calculate the statistics for the prediction errors. The synthetic observations can also be assimilated into the "Analysis" model to test the ability of the technique to do assimilation (since the truth is known, the quality of the assimilation can easily be measured). The two models can either be run simultaneously (as shown in Figure 4.3) or the synthetic data can be used as input to a scheme like Figure 4.2.

In order to generate sufficient statistical data, it is necessary to collect statistics for 90 sols. Furthermore, the system needs to be sampling a statistically constant system while gathering the covariances. This is done by taking three pairs of initial conditions at the same point in the Martian year and gathering statistics from each pair over a 30 sol period and combining them. The six initial conditions are generated by perturbing the GCM fields and then allowing it to run for 30 sols. This insures that the initial states are dynamically uncorrelated but seasonally identical.

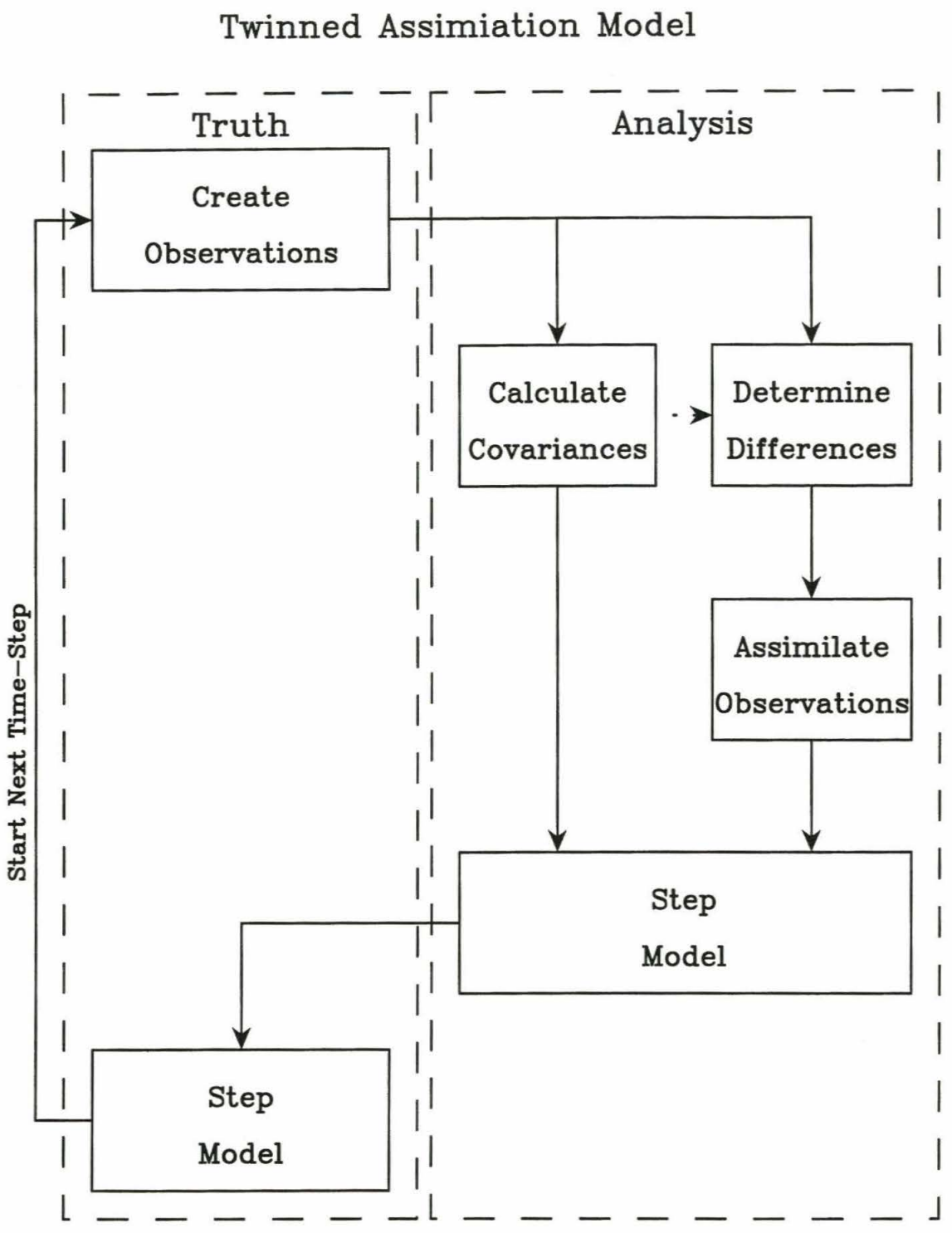

Figure 4.3: Twinned Model Flow Chart

This is the process used for generating the statistics (covariance terms) for the gains. Two copies of the GCM are run in parallel with different initial conditions but the same season. It is also possible to use the explicit twinned models for assimilating synthetic data. The two processes can be combined (the dashed arrow), but that is for calculating higher order Kalman gains.

4.4.2 Assimilating

The second step of the assimilation process is to use the gains to actually assimilate data. While computationally intensive (since each observation point has to be assimilated into the entire state space), this step is much simpler conceptually as it just implements equation 4.1. One of our goals in implementing the actual assimilation process was to avoid modifying the MGCM itself as much as possible. Apart from modifying the state vector, the modifications to the GCM are limited to three subroutine calls (one to initialize the data assimilation, one to perform the assimilation each time-step, and one to end the assimilation).

While the model assimilates each profile independently, it calculates all of the differences $\left(\Psi_j^o - \Psi_j^p\right)$ before doing any of the assimilation. Like this, the processing is effectively in parallel even though the model actually assimilates in series since it adds the corrections to the state vector as they are calculated. The parallel process matches the assumptions built into the assimilation equations that the method was derived from. It also insures that the assimilation is independent of the order of the profiles within the time-step (which actually allows the program to more efficiently do the assimilation calculations).

As mentioned when discussing the gains, the construction of the gains (equation 4.4) is independent of M and N and thus ignores any term referring to the number of gains in the time-step. This is done to allow the gains to be constant. While ignored, these terms are important since the gains for OI includes terms that reduce the weight of each gain depending on the total number being assimilated. Effectively, the combined weight of all the gains is constant (so that the weighting of the model and data remains independent of the number of observations). In order to recover this behavior, all the gains are divided by M, the total number of observation profiles in that time-step (for computational reasons, it is actually the differences that are divided, but this is mathematically equivalent). The precalculated gains approximately account for the other points within a profile, so the additional term only needs to account for the number of profiles in the time-step.

When the system was fully implemented, it was found to be unstable under many conditions. The combined MGCM and assimilation system would generate instabilities that grow and eventually become numerically unstable at which point the MGCM would attempt an illegal operation and crash. This is known to happen when performing data assimilation on large, complex systems [*Bierman,* 1977]. In order to resolve the problem, we introduced an f factor into the assimilation equation (equation 4.1):

$$\Psi_k^a = \Psi_k^p + \sum_{j=1}^{N} \frac{f}{M} \alpha_{kj} \left(\Psi_j^o - \Psi_j^p \right) \tag{4.5}$$

where M is then number of profiles in the time-step and $f \leq 1$ is the control factor. The f values required for stability vary depending on the data being assimilated. While the "goodness" of the assimilation is somewhat noisy as f changes, there is usually a broad minimum around some f value. Significantly below that value, the assimilation is much worse because the data is not weighted sufficiently and the model goes off on its own. If f is too large, the fit gets worse and worse as the instability type behavior sets in. At some f value, the system goes sufficiently unstable for numerical instabilities to develop.

While it is possible to run the method with $f > 1$, this was not done. In such cases, the data is being emphasized at the expense of the model and there does not seem to be a reason for doing so. In such cases, the error of the observations is probably being overestimated, and correcting that should make the optimum value for f be 1. With $f = 1$ (which is actually stable for some synthetic data regimes), the method is operating "optimally." This is generally due to the data being very similar to the model, resulting in a simpler assimilation problem. These cases emphasize some of the dangers of using a twinned model approach.

4.5 Results

In order to test out and fine tune the assimilation method, it was necessary to run experiments in controlled cases. In order to do this, we created synthetic data by

running the MGCM and flying a simulated spacecraft over the planet. As the spacecraft traveled, it would make measurements of the MGCM atmosphere below it. We use two different cases. In the first case, the spacecraft is in a circular orbit like the expected final MGS orbit. In this mode, synthetic profiles are created about once a minute. This is done by selecting a set of pressures and then interpolating the model coordinate grid to get the temperature at those positions. The temperatures are modified by a random noise term. The second synthetic data set is developed using the actual TES observation pattern. In this case, the model uses the location (latitude, longitude and altitude) of each observation in the TES data set and makes a synthetic observation in the same location. This allowed us to insure that the assimilation technique works not only for the expected data pattern, but also for the data actually available.

While the synthetic data has several significant advantages, it is effectively a twinned model system (although the two models are partly decoupled) and thus has several potential problems. With the synthetic data, the "Truth" is known, so it is fairly easy to measure the effectiveness of the assimilation technique. The major problem is that since the same model generates the data as well as assimilating it and is also the source of the gains used for the process, it is possible for non-physical aspects of the model to drive the assimilation process. This is alleviated somewhat by the use of an observational error in calculating the synthetic data. In several cases, multiple different sets of synthetic data were developed for the same cases from different initial conditions. Another effect of a twinned system is that the assimilation of the synthetic data will be more effective than assimilating the actual data. This is primarily because the error terms used to build the gains are exact instead of approximations, but also because the synthetic data is not as different from the model as the actual data. The model considers itself to be climatologically correct and is only assimilating the weather.

Figure 4.4 shows the success of the method at assimilating synthetic data from a circular orbit. This figure compares the "Truth" (i.e., source of the synthetic data) and the results of the assimilation run. It shows the square root of the mean of the

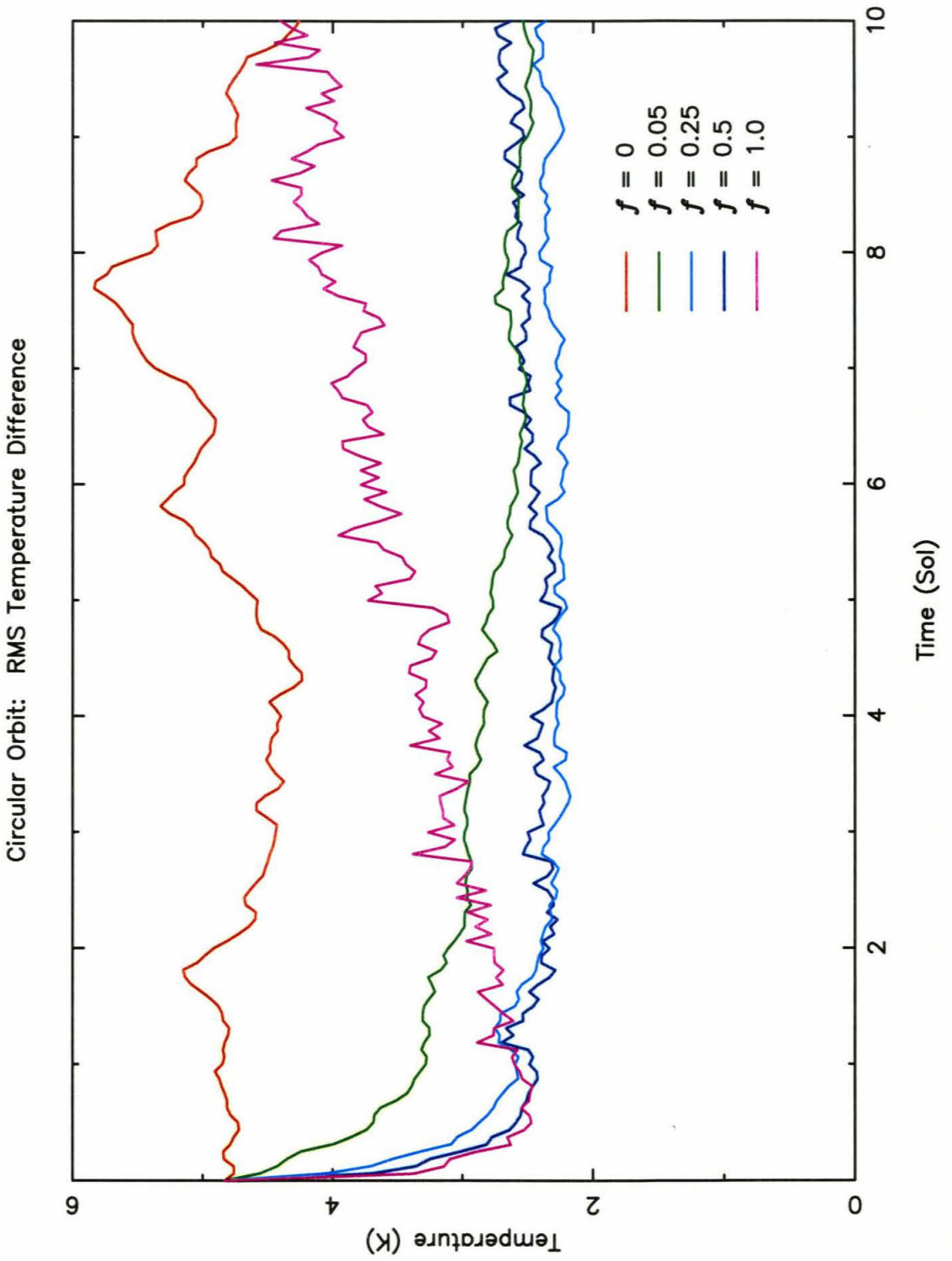

Figure 4.4: Assimilation Results for Circular Synthetic Data
The RMS temperature difference (D_T) between the "Truth" and the "Analysis" model for several different f values. The case with $f = 0$, in red, is the case with no assimilation and represents the background weather variability.

squared temperature differences between the two cases (D_T). At each time-step in the output files (every 1.5 hours), the difference between each point of the temperature field of the "Truth" and the "Analysis" are calculated, squared and averaged.

$$D_T(t) = \sqrt{\frac{\sum_k^T \left(\epsilon_k^a(t)\right)^2}{T}} \tag{4.6}$$

where $\epsilon_k^a(t)$ is the error of the analysis model at time t at point k and T is the temperature domain of the state vector. Thus $D_T(t)$ is the global temperature difference between the two MGCM runs at time t. Each of the five curves is an assimilation run with a different value for f. The red line ($f = 0$) is effectively the difference between two runs of the GCM with the same season but different initial conditions, since with $f = 0$ no assimilation is performed. The other four cases are for four different f values, showing the range of possible fits. The first feature to notice is that all of the curves improve rapidly at the very beginning. This represents an initial correction as the data assimilation gathers enough data to reasonably sample the planet. Once it has mostly corrected the initial problems, generally after the first sol, the assimilation keeps the models as close as possible for that f value. In the best case, $f = 0.25$, the temperature error after the first few sols is a bit above 2K. The mean observational noise used in creating the synthetic data was 2K so the assimilation is doing quite well.

Note that at $f = 1$, magenta line, while there is a large and very fast initial improvement, the difference then starts to get larger and larger with time. While the errors successfully correct the model, the further differences are amplified by the assimilation process and the model becomes unstable. If run for a sufficiently long time ($\sim$ 30 sols) the instability grows and reaches the point where it creates a numerical instability causing the model to crash.

In the case of $f = 0.05$, green line, the assimilation improves the model, but it does not do so as much as the others initially. Starting around sol 3 or so, the value begins to improve slightly and (coupled with a slight worsening of the $f = 0.25$ and $f = 0.5$ cases) ends up being about as good as the other two cases at sol 10. This is

actually deceptive, and over a longer period the two diverge with the $f = 0.05$ case remaining $\sim$ 1K worse than the other two (see also Table 4.1). There is a long period undulation of $\sim$ .5K in all three fits that is not understood at this time.

Figure 4.5 shows the RMS difference for the zonal component of the winds (D_U). The colors are identical to Figure 4.4, as are the general results. While not shown, the fields for the meridional winds and pressure are also significantly improved. This shows that the assimilation technique is capable of correcting the dynamical fields not measured. This is especially useful since it is much easier to measure temperatures than it is to measure winds.

While the ability to successfully assimilate data in a circular orbit validates the assimilation technique, the data that TES has collected is not in a circular orbit. Using the available TES data as a template, the model was tested using non-circular data. The results of the assimilation are shown in Figure 4.6. The figure is comparable to Figure 4.4—the red line is again the inherent variability of the weather. Figure 4.7 shows the number of observation profiles being assimilated each time-step. Note how the abrupt transitions in the assimilation (Figure 4.6) correspond to the spikes with the peak numbers of observations per time-step. These periods are the close passes at the beginning of each orbit. Unlike the circular orbit, the synthetic TES orbital data is actually stable at $f = 1.0$. This appears to be due primarily to the sampling pattern that generally allows the model time to stabilize after each block of data input. Thus instead of continually being modified, there are periods where the model can drift and damp out some of the instabilities. Being stable at $f = 1.0$ means that the assimilation is extracting all of the available information from the data. Unfortunately, while the process can better use the available information, there is less information than in the circular orbit and the assimilation is less effective at recovering the truth.

Looking at a plot of the RMS temperature difference (or the RMS difference for any of the fields) is useful for qualitatively determining how good a fit is doing, but there is also a need for a quantitative measure of the effectiveness of a fit. This can be done by calculating an explicit goodness value—this is especially convenient for

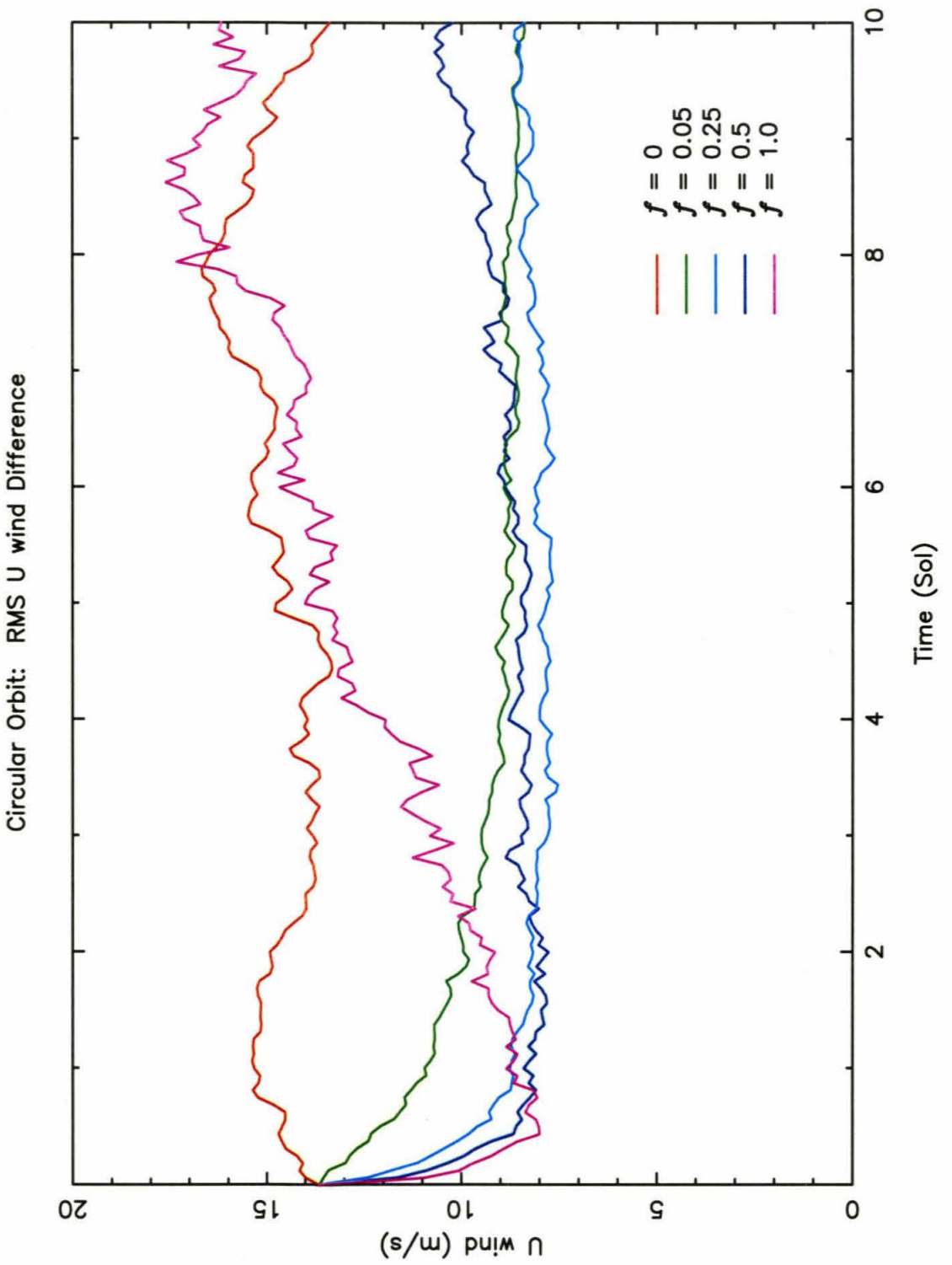

Figure 4.5: Zonal Wind Assimilation for Circular Synthetic Data
The RMS Zonal Wind difference (D_U) between the "Truth" and the "Analysis" model for several different f values.

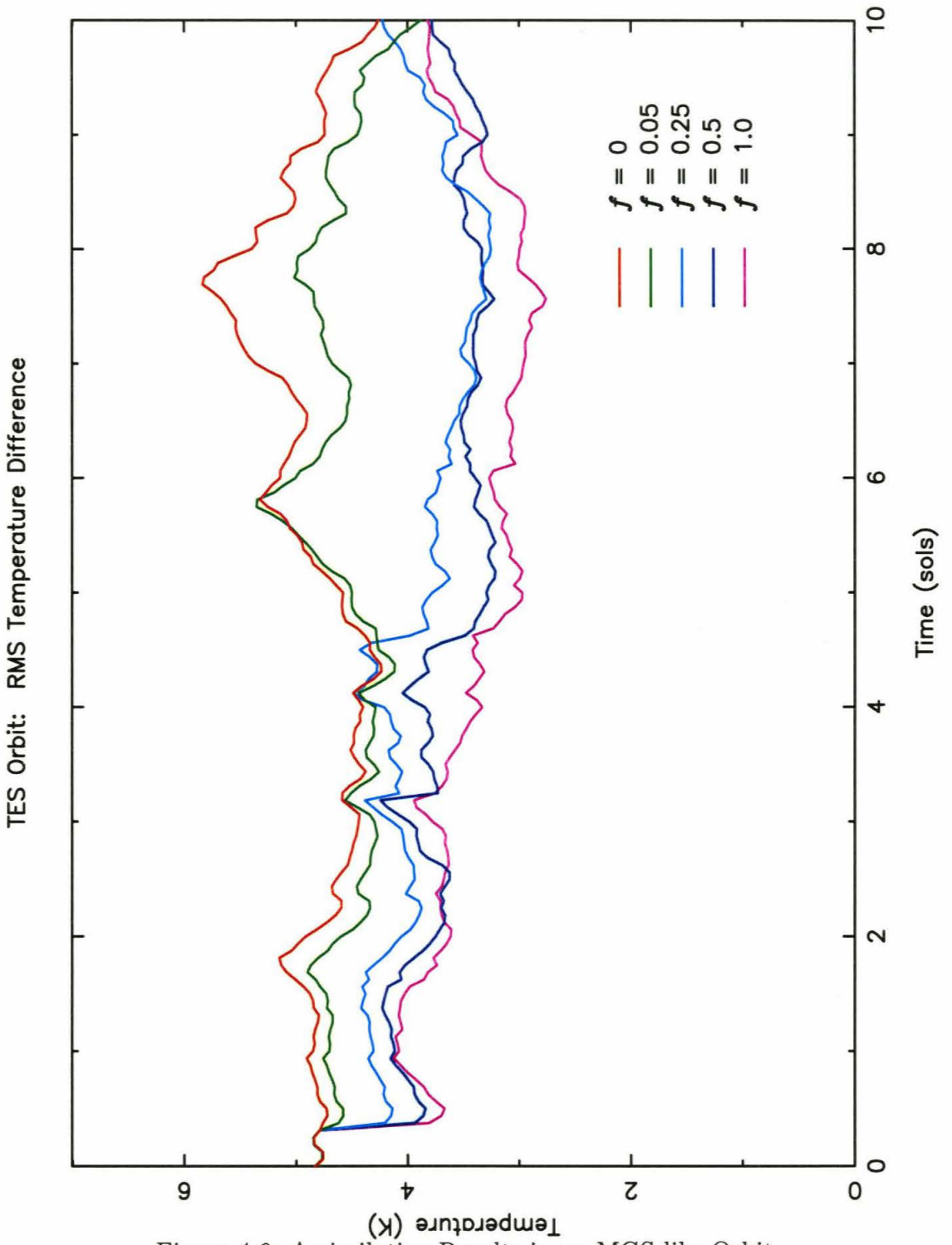

Figure 4.6: Assimilation Results in an MGS-like Orbit

The RMS temperature difference between the "Truth" and the "Analysis" model for several different f values. The case with $f = 0$, in red, is the case with no assimilation and represents the background weather variability. The other three fields (P, U and V) are similar.

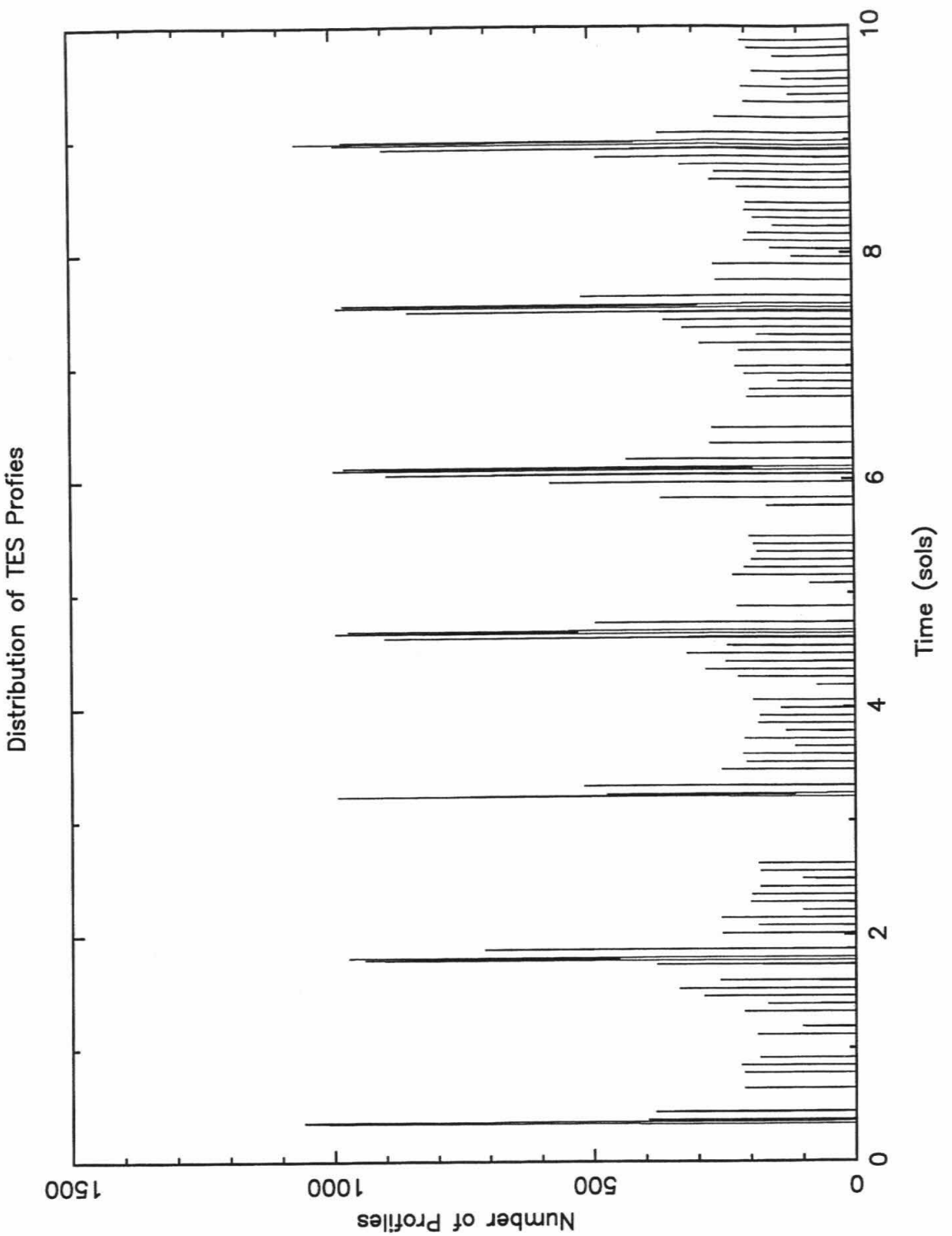

Figure 4.7: Temporal Density of Observation Profiles
The number of profiles in each time-step (6 minutes) of the assimilation. The dense, closely spaced clusters usually involve 5 to 10 time-steps and represent the close pass of the planet during the orbit (the red crosses in Figure 4.1). The individual spikes between represent scans across the planet during the rest of the orbit.

comparing the different fields since they are all in different units and/or have different errors associated with them. A useful definition for "goodness" (G) is the following:

$$G_T = \frac{1}{t}\sum_t \left(\frac{D_T^a(t)}{D_T^p(t)}\right)^2 \tag{4.7}$$

G_T is the goodness value for temperature. Basically it is the mean of the squared fractional improvement of the analysis model (when compared to pure weather variability). The square term is included to give it the variance like behavior expected for such a measure. All G values are scaled identically, with 0 being perfect assimilation and 1 being no assimilation (values greater than 1 imply over-forcing). This allows the G value for the different fields to be compared directly. For both the TES case as well as the circular data, Table 4.1 shows the change in G with f for the four dynamical fields and then the mean value. A curve can then be fit to the goodness values (as in Figure 4.8) and the optimum f value determined. As can be seen, the optimum value is close to $f = 1.0$ and the assimilation at that value will be used for the rest of the discussion.

As can be seen from Table 4.1, the irregularity of the TES orbit significantly degrades the ability of the algorithm to actually assimilate data. G_{Mean} for the optimum circular case is 0.26, but only 0.51 for the TES data. Despite this, the technique is still capable of modifying the "Analysis" model. The error in the model has been reduced and it is significantly closer to the "Truth." The primary problem with the irregular observations is that there is poor coverage of the northern hemisphere, especially at the midlatitudes. While each orbit has one line of data during the close pass, this only occurs every $\sim$ 1.5 sols instead of the 12 times per sol in the circular orbit. This problem is aggravated by the season. At this season ($L_s \sim 315$–mid Winter in the Northern hemisphere), most of the active meteorology is in the Northern storm belt. This is a region of low and high pressure systems that encircle the planet as they travel around.

The storm belt is best seen in the pressure field (Figure 4.9). This is a snapshot of the surface pressure from the MGCM that has had the mean for a 10 sol period

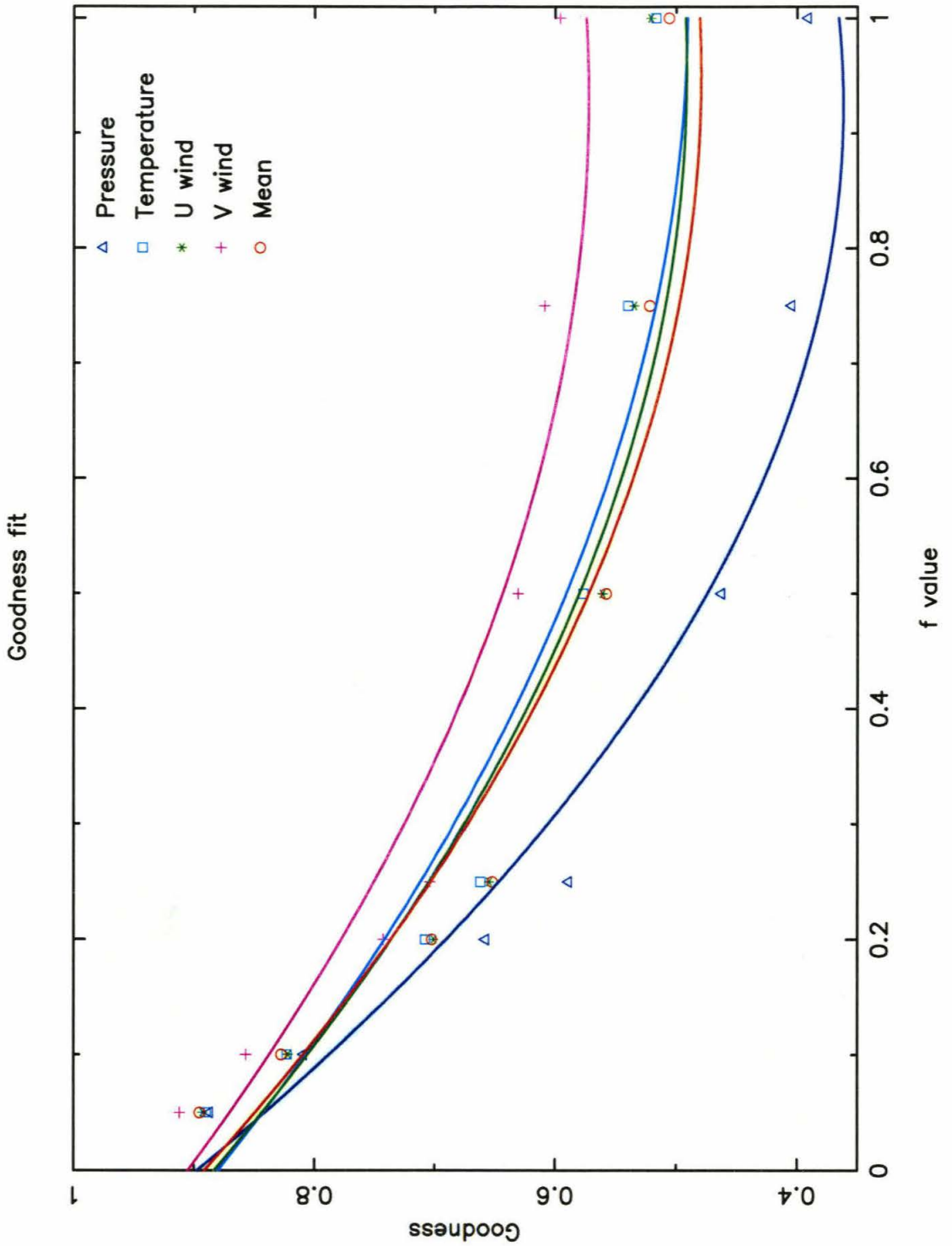

Figure 4.8: Goodness Value Fits

A parabolic fit to the goodness values for the synthetic TES data. Each field is fit independently and the mean value is fit as well. The minimum in the parabolas is $f \sim 1.0$.

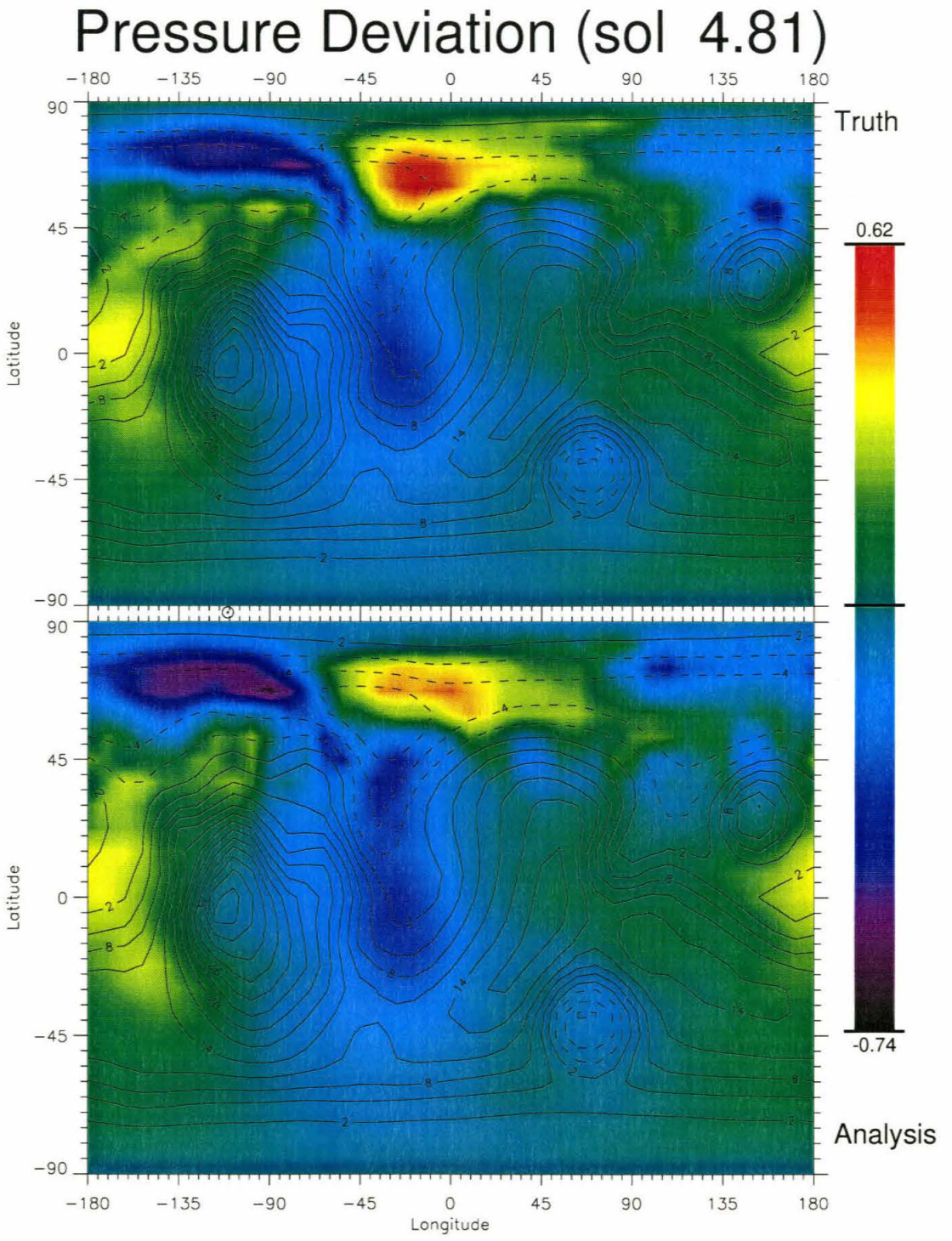

Figure 4.9: Truth and Analysis Pressure Map

Map of the surface pressure. The upper panel is the "Truth" model at 4.81 sols. The lower panel is the "Analysis" model at the same time. The mean pressure field (combined for the two models as well as the "Prediction" model) over a 10 sol period was subtracted from each point to remove the topographic effect. The GCM topography is contoured for reference.

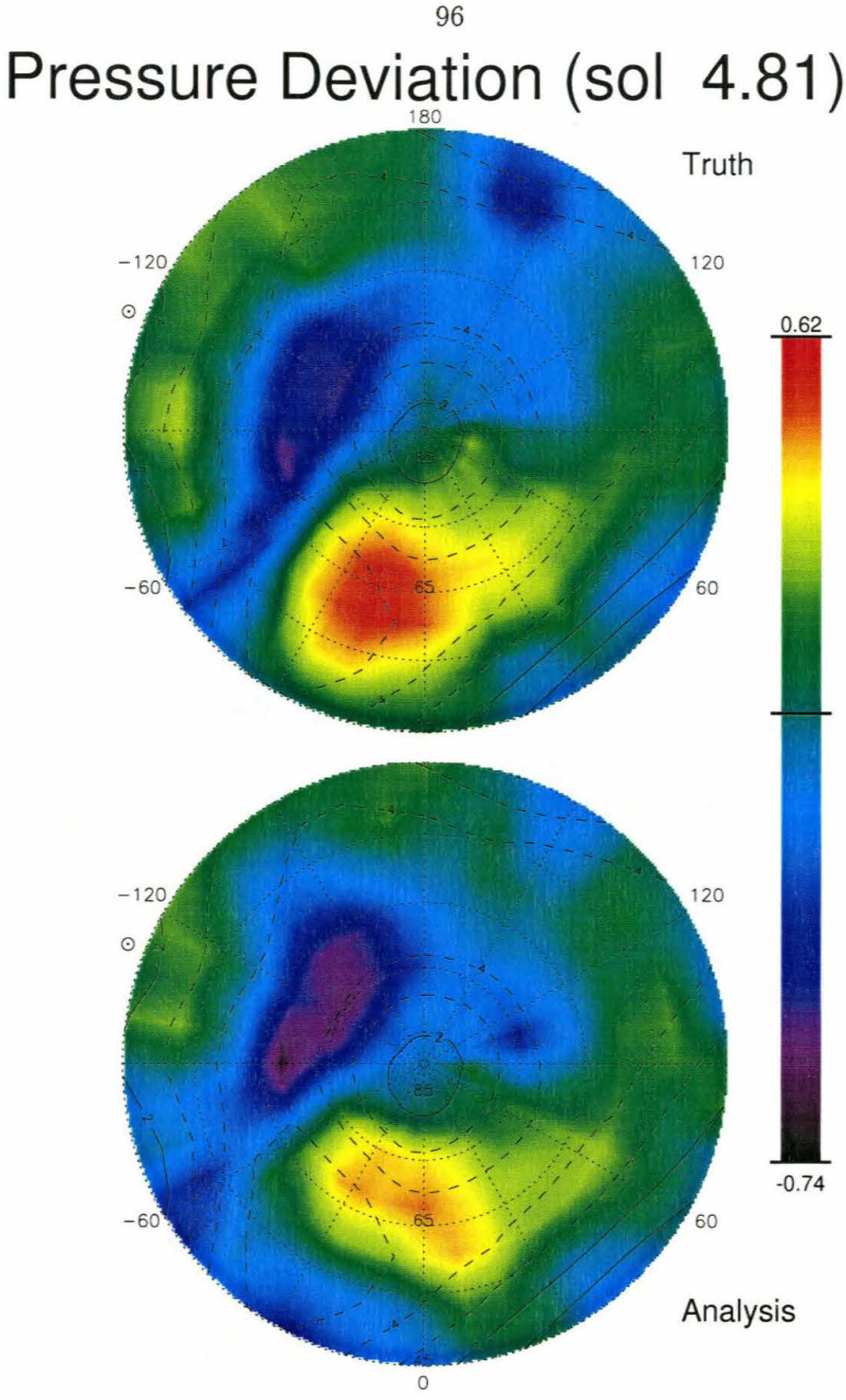

Figure 4.10: Truth and Analysis Polar Map
Polar map of the surface pressure. This is the same data as in Figure 4.9, but a polar plot of the northern hemisphere. The outer contour is 45N.

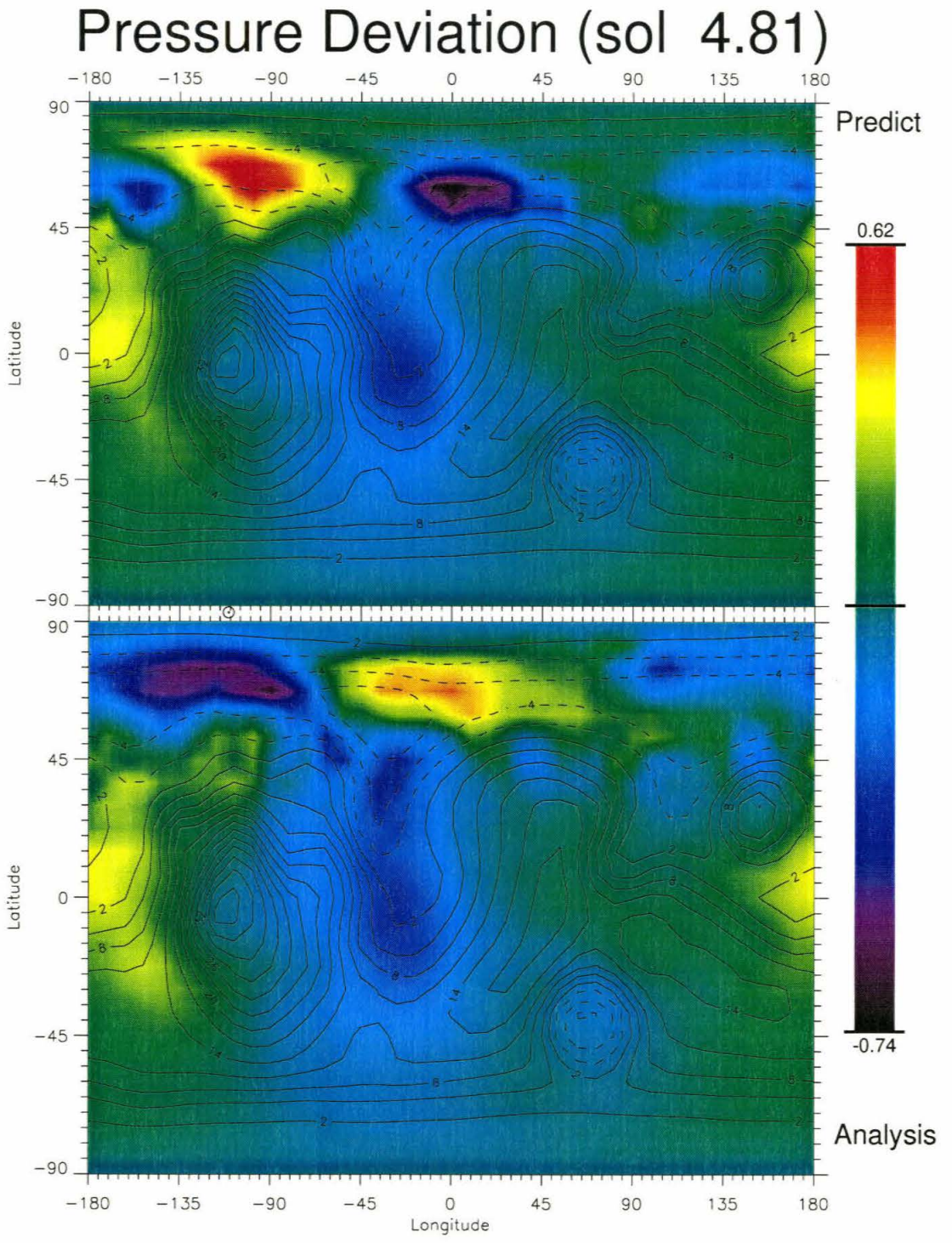

Figure 4.11: Prediction and Analysis Pressure Map
Map in latitude and longitude of the surface pressure. The upper panel is the value for the "Prediction" model (a model run with the same initial condition as the "Analysis," but no assimilation). The lower one is the "Analysis" model at the same time. This figure uses the same scaling (and is directly comparable) as Figure 4.9.

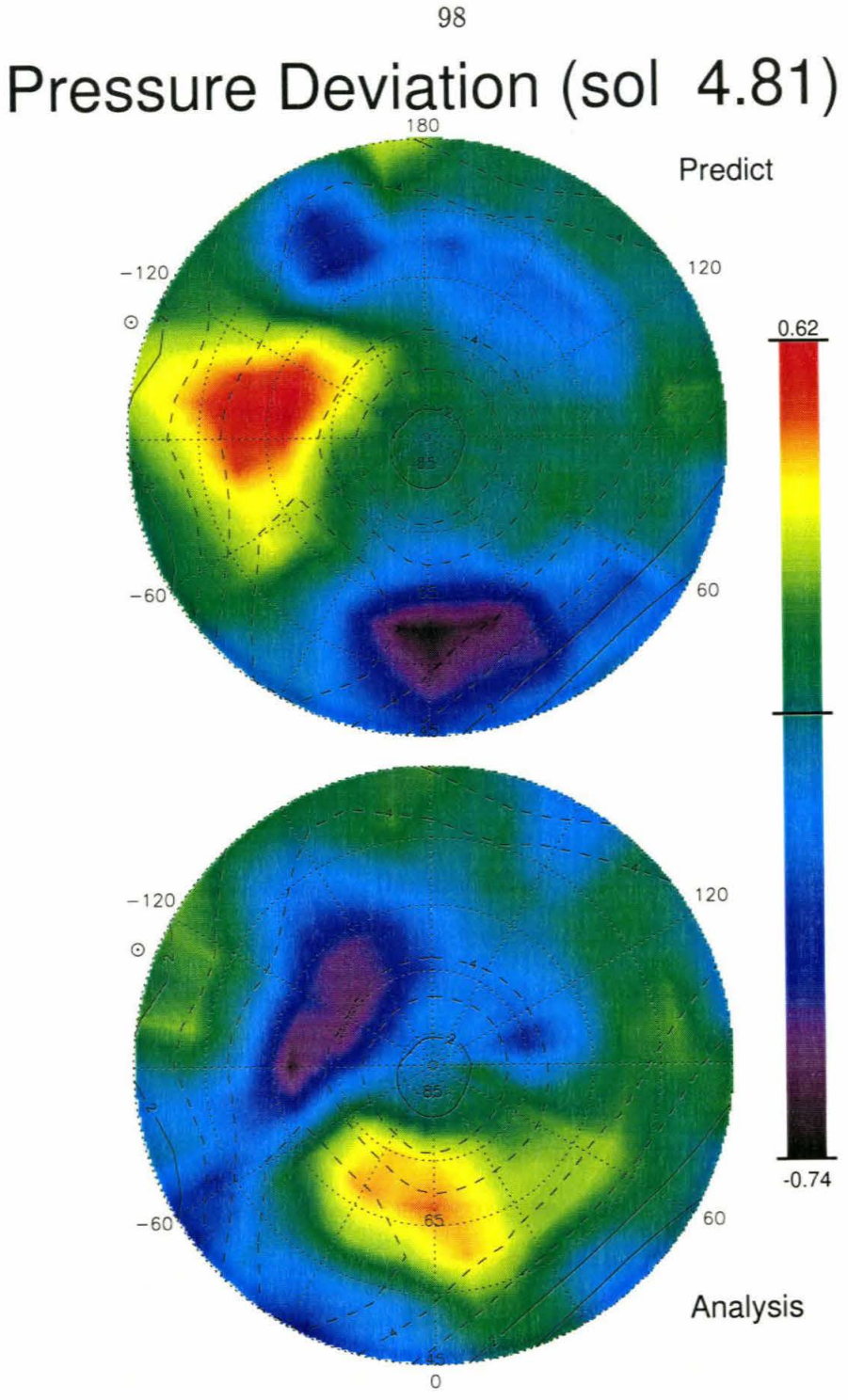

Figure 4.12: Prediction and Analysis Polar Map

Polar map of the surface pressure. Map in latitude and longitude of the surface pressure. This is the same data as Figure 4.11, but a polar plot of the northern hemisphere. It is directly comparable to Figure 4.10.

Goodness Values for Assimilating Synthetic Data

	Goodness				
	G_P	G_T	G_U	G_V	G_{Mean}
Synthetic TES Data					
$f = 0.05$	0.89	0.89	0.89	0.91	0.90
$f = 0.1$	0.81	0.82	0.82	0.86	0.83
$f = 0.2$	0.66	0.71	0.70	0.74	0.70
$f = 0.25$	0.59	0.66	0.66	0.70	0.65
$f = 0.5$	0.46	0.58	0.56	0.63	0.56
$f = 0.75$	0.41	0.54	0.54	0.61	0.52
$f = 1.0$	0.39	0.52	0.52	0.60	0.51
Synthetic Circular Data					
$f = 0.05$	0.21	0.35	0.40	0.46	0.35
$f = 0.25$	0.11	0.24	0.31	0.36	0.26
$f = 0.5$	0.19	0.26	0.36	0.39	0.30

Table 4.1: Goodness Values for Various f Values
The G value (see equation 4.7) for each dynamical variable. The mean is the average of the four fields to get a total fit. Both the synthetic TES data and synthetic circular data are shown (the circular case with $f = 1.0$ is omitted since it is not stable).

removed from each point, showing the deviations. This is done to remove the huge effects of topography across the planet. For reference, the topography of the GCM is shown with the contours. These are the same contours as Figure 4.1. The upper panel is the truth model at this time; the lower one is the analysis model during the assimilation. The two models begin with initial conditions whose weather patterns are uncorrelated. Both panels have the same mean pressure removed from each point and are directly comparable. For the analysis case, $f = 1.0$ and the subsolar longitude is -110. The high pressure ridge west of there, and the minimum at ~ -45, are the solar thermal tides (which are quite strong at the surface on Mars). One important feature is the series of "waves" (or storms) encircling the planet between 45 N and 80 N. This can also be seen in Figure 4.10 which is a polar view of the same data, but only the the northern hemisphere north of 45N. The system of waves is quite variable as individual storms break up and string out and reform as they go around the planet. Despite the poor data sampling in the northern hemisphere, the "Truth" and the "Analysis" models show significant similarities. While the finer structure is

not the same, the longitude, strength and some of the structure of the high and low pressure regions correspond fairly well. Figure 4.11 shows the same information, but this time the upper panel is the "Prediction" model—the same initial condition as the "Analysis" model, but without assimilation (and Figure 4.12 is the corresponding polar plot). The difference between the two is noticeable, especially when compared to the similarities between the "Truth" and "Analysis." Thus, even though it is poorly sampled, the assimilation scheme is capable of correcting, at least to some extent, differences between the truth and prediction models.

4.6 Variations in the Method

While the results presented in the previous section were obtained with the assimilation method used for the actual data (the "Standard" method), we experimented with several alternative possibilities. The three main issues studied were the equation to use for the gains, the method of actually applying equation 4.5, and the possibility of averaging data.

The alternative for calculating the gains is to assume that each observation, including within each profile, is completely independent and assimilated by itself. In this case, the equation for calculating the gain simplifies to [*Banfield et al.*, 1995]:

$$\alpha_{ki} = \frac{\overline{\epsilon_k^p \epsilon_i^p}}{\overline{\epsilon_i^p \epsilon_i^p} + \overline{\epsilon_i^o \epsilon_i^o}} \tag{4.8}$$

In this case, M in equation 4.5 becomes N–the total number of observations. This form eliminates the need to assume one observation per GCM layer, but requires the points within the profile to be uncorrelated. A look at Table 4.2 shows that this method, labeled "Simple Gains," is significantly worse than the standard technique. Assuming that the points are uncorrelated (when the assumed error function has the nearest neighbors significantly correlated) end up being much worse than the inaccurate vertical distribution used in the "Standard" gains. This is due to the high point to point error correlation in the observations. Furthermore, since the

Comparison of Alternative Methods

	Goodness				
	P	T	U	V	Mean
Standard	0.37	0.50	0.50	0.58	0.49
Simple gains	0.84	0.82	0.81	0.83	0.83
Serial processing	0.46	0.53	0.52	0.60	0.52
Averaging	0.49	0.55	0.56	0.61	0.55

Table 4.2: Goodness of Alternative Methods
Each row represents an alternative method of performing the assimilation. The values represent the goodness (G is a measure of the remaining error–0 is perfect). All of the experiments were done with synthetic data based on the the TES observations. Note that these values are not comparable with the ones in Table 4.1 due to different initial conditions.

computational savings is in calculating the gains, it is less important.

Optimal Interpolation, if fully implemented, can be performed in series instead of in parallel as implied by equation 4.1. Since the actual assimilation has to be done in series (computers can really only do one thing at a time), we considered the possibility of using serial assimilation [*Banfield et al.,* 1995], in which the Ψ^p are updated after each observation is assimilated instead of waiting until all observations for that time-step have been assimilated. This has the advantage that the gains are actually computationally less expensive (since matrix inversion is no longer required). Unfortunately, the cost is still too high to calculate the gains during assimilation. Attempting to use the precomputed gains has two unfortunate effects. First, the order in which the observations are assimilated matters. While with full OI, this is not the case, the precomputed gains cannot compensate for the changes to Ψ^p as the observations are assimilated. Secondly, the observations take on differing importance in such a scheme and the precomputed gains cannot compensate for this. Combined, these two effects cause serial methods to be both less rigorous as well as less performant (see Table 4.2).

The third major variation was to use averaging. Since, during the close pass as well as during parts of the roll-out, the data density is much higher than the resolution of the MGCM, averaging could reduce the computational burden while

not actually loosing information. Averaging was done by dividing each grid box into four (2×2) subgrids. For each time-step, all the profiles within each subgrid box were averaged on each pressure surface of the data. The process reduces the total number of profiles by about one-third—as suspected, the parts of the orbit at long range offer few opportunities for averaging. This is not sufficient to make any of the more computationally intense methods viable. Furthermore, the averaged data actually performs worse during assimilation (see Table 4.2). This is because the assimilation method implicitly averages the data (via the $1/M$ term), assuming all profiles are equal and have identical error characteristics. Thus it does not assign sufficient weight to the averaged profiles and loses the actual quality of those profiles.

4.7 Conclusion

In this chapter, we describe a data assimilation technique based on Optimal Interpolation that can successfully assimilate data taken in the observation pattern of the MGS-TES data during the elliptical aerobraking orbits. The method also works quite well for simpler circular orbits, but can probably be improved to use the regularity in the observation pattern (possibly by using a steady–state Kalman Filters).

In order to implement the method efficiently and effectively, it was necessary to significantly modify the theoretical method. This is especially true of the gains used for the assimilation. We are able to, at least in the context of Mars, keep the gains global and use each observation to update the entire model atmosphere.

The method, as implemented, is capable of performing meaningful assimilation and appears to drive the "Analysis" model towards the truth. This is especially apparent in the phasing of the northern storm belt, but can also be seen in the global dynamical fields. While the improvements are not as large as we hoped for, they are significant and will allow meaningful assimilation of actual TES data.

Bibliography

Banfield, D., A. P. Ingersoll, C. L. Keppenne, A Steady–State Kalman Filter for Assimilating Data from a Single Polar Orbiting Satellite, *J. Atmos. Sci., 52,* 737-753, 1995.

Bengtsson, L., N. Gustavsson, An Experiment in the Assimilation of Data in Dynamical Analysis, *Tellus, 23,* 328-336, 1971.

Bergman, K. H., Multivariate Analysis of Temperatures and Winds using Optimum Interpolation, *Mon. Wea. Rev., 107,* 1423-1444, 1979.

Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation,* Academic Press, New York, 1977.

Christensen, P. R., D. L. Anderson, S. C. Chase, R. N. Clark, H. H. Kieffer, M. C. Malin, J. C. Pearl, J. Carpenter, N. Bandiera, F. G. Brown, S. Silverman, Thermal Emission Spectrometer Experiment: Mars Observer Mission, *J. Geophys. Res., 97,* 7719-7734, 1992.

Conrath, B. J., J. C. Pearl, M. D. Smith, P. R. Christensen, MGS TES Results: Characterization of the Martian Atmospheric Thermal Structure, *Bull. Amer. Astro. Soc., 30,* 1031, 1998.

Haberle, R. M., J. B. Pollack, J. R. Barnes, R. W. Zurek, C. B. Leovy, J. R. Murphy, H. Lee, J. Schaeffer, Mars Atmospheric Dynamics as Simulated by the NASA Ames General–Circulation Model; 1. The Zonal–Mean Circulation, *J. Geophys. Res., 98,* 3093-3123, 1993.

Pollack, J. B., R. M. Haberle, J. Schaeffer, H. Lee, Simulations of the General–Circulation of the Martian Atmosphere. 1. Polar Processes, *J. Geophys. Res., 95,* 1447-1473, 1990.

Ruthford, I. D., Data Assimilation by Statistical Interpolation of Forecast Error Fields, *J. Atmos. Sci., 29,* 809-815, 1972.

Chapter 5 Assimilation of TES Data

5.1 Introduction

From a dynamical point of view, the atmosphere of Mars is simpler than that of the Earth. Since the total mass of the Martian atmosphere is much lower, it responds to radiative forcing much faster than the Terrestrial atmosphere. Also, precipitation appears to be confined to high latitudes during the winter and possibly locally during certain times of the day, thus atmospheric water is not a critical source of heat transport like on the Earth. While the Martian atmosphere is potentially an easier system to understand, there are several fundamental issues that make it very different from the Earth. First, the bulk constituent, CO_2, condenses out of the Martian atmosphere and is probably in vapor equilibrium with CO_2 ice. Also, the atmosphere contains a large amount of dust that modifies the behavior of the atmosphere in several ways.

A good understanding of the dynamics of Mars is important for addressing several problems. These include questions about dust, like where does it come from, how do dust storms start, and how do they grow? The dynamics affect the latitudinal heat transport that determines the heat budget for the planet as a whole. They also control the movement of water on the planet.

Much like the Earth, the Martian atmosphere has weather as well as change on longer timescales. Thus, in order to understand the atmospheric dynamics, it is important to understand the mean large scale behavior—the climatology—and also to understand the change and variability that occurs within the pattern of the seasonal climate. This variability, or weather, is as important as the climatology itself. It shows what is actually happening and is what the atmosphere (and surface) actually experience. The differences and locally changing weather may be the trigger or controlling factor for much of the larger scale more "seasonal" behavior that is

seen.

The previous chapter developed a method for assimilating TES (Thermal Emission Spectrometer) atmospheric data into the Ames Mars GCM (General Circulation Model). In this chapter we use the method to assimilate actual TES observations and then analyze the results of the assimilation. By assimilating the TES data, we can obtain a better climatology and we can look at the actual Martian weather patterns.

5.2 TES Data

Currently the available TES atmospheric data is limited to 25 sols of data around $L_s = 200$ (early fall). This is 17 orbits (orbits 20 through 36) during the first hiatus period. The TES data is a series of temperature measurements at fixed pressures obtained by inverting the shape of the 15 μm CO_2 line. Each spectrum is inverted to produce 10 to 20 temperature measurements in a vertical column. One column is termed a "profile." During these observations, the spacecraft was in a highly elliptical 35 hour orbit. Near the periapse of each orbit, the instrument records a very concentrated swath of data (up to 1100 profiles per 6 minute time-step). Figure 5.1 shows the number of profiles per time-step for the full set of data available. Sol 0 is the first sol that contains observations. The timing of the observations is selected so that this is the same L_s as the actual observations and so that the local time of day is correct for the observations. While the footprint of the instrument varies considerably over the orbit, it is significantly smaller than the MGCM grid and therefore the actual footprint is ignored.

During the actual assimilation (as opposed to assimilation of synthetic data), it is necessary to filter the data. This is done by removing any profile where $\Psi_j^o - \Psi_j^p > 50\text{K}$ for at least one point in the profile. Since there appears to be significant observational error correlation within a profile, if any point in the profile is problematic, the rest are also problematic. Points with large differences cause problems because they result in large forcing effects which tend to destabilize the system. A look at the $\sim$ 170 profiles that were not assimilated (less than 0.1% of the all the profiles) shows that

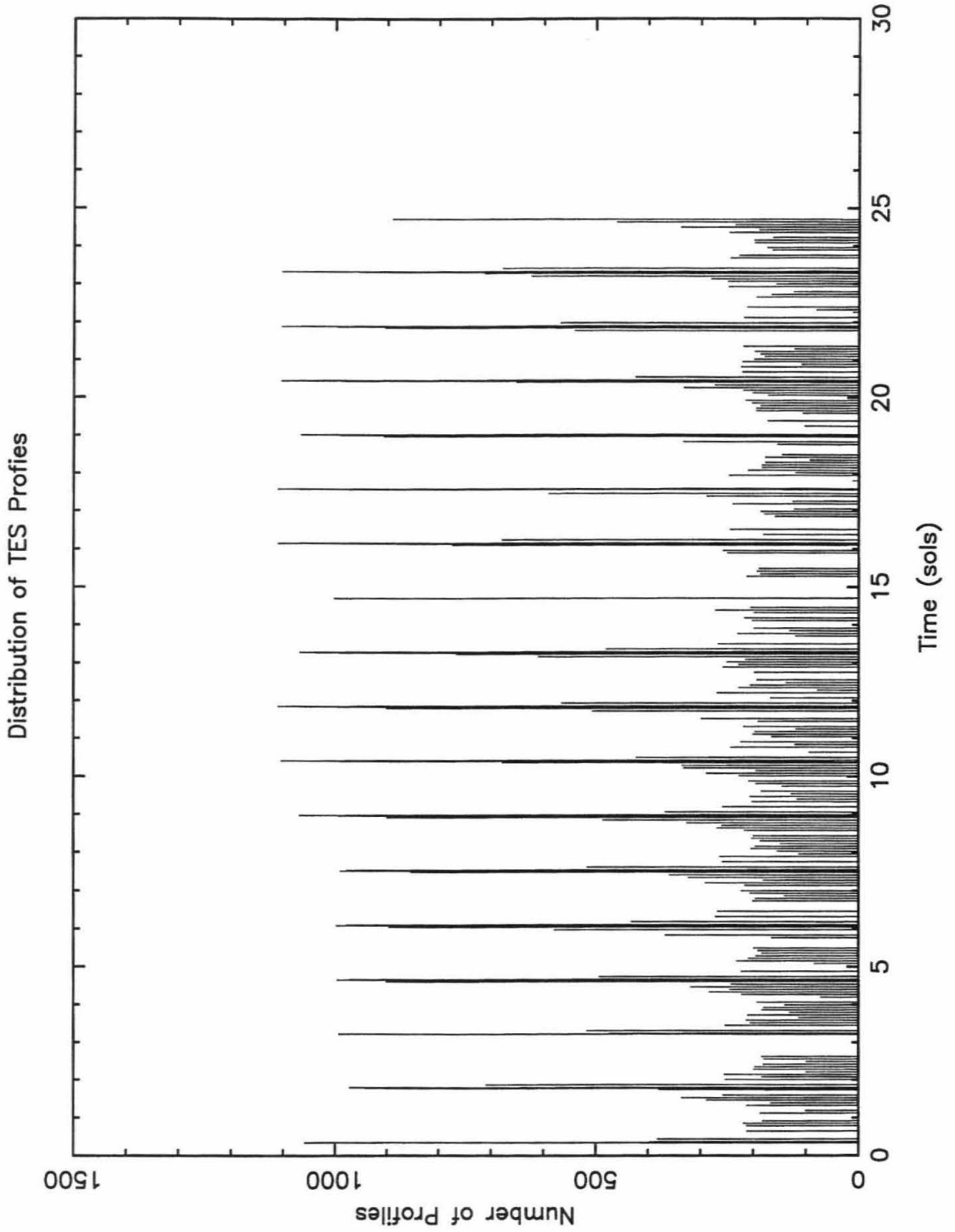

Figure 5.1: Data Density for Available TES Data
This indicates the total number of TES profiles available for each 6 minute time-step of the MGCM. The close passes are the locations with the highest density and mark the beginning of each orbit.

they are either cases where the inversion algorithm used to generate the profiles failed or the instrumental noise was extreme (we cannot actually distinguish between the two). Even though very few profiles are removed, the net effect is large. Without removing those profiles, the assimilation only produces marginal improvements in the analysis model (the process would change the model, but apparently not in a useful manner). Roughly 1/3 of the filtered profiles are in a tight cluster at very high altitudes (surface pressures of $\sim$ 1 mbar) on the flank of Olympus Mons. The high altitude profiles probably reflect the fact that the inversion process has problems with very low surface pressures. The other 2/3 of the profiles are scattered around the planet, somewhat grouped near the equator, with no apparent pattern.

The choice of 50K is somewhat arbitrary. It was selected as being small enough to remove the worst of the problematic profiles (as measured by the improvements in the analysis model), but not so small that valid profiles (i.e., ones with significant contributions) are removed. Most of the removed profiles either show the signs of low surface pressure inversion problems or unphysical behavior probably due to instabilities in the inversion algorithm (for example, many have very cold temperatures, as much as 30K below the CO_2 condensation temperature). A few profiles appear fairly reasonable but are either very cold or very warm. Since there are very few of them and they are not particularly clustered, it is not clear if these are intense local phenomena, instrument noise or inversion problems. While the 50K filter catches many of the worst profiles, there are several other profiles that were very likely also problems (often having differences of 40K or larger). Unfortunately, especially during the first few sols of the assimilation, there are some valid large differences, making it difficult to use a smaller filtering value.

Attempts to pre-filter the data, using some simple rejection criteria, were not successful. At best, they performed about as well as the simple 50K cutoff, and were generally a bit worse. The two criteria used were profiles with temperatures more than 5K below the condensation temperature (which removed very few profiles) and those with surface pressures less than 2.5 mbar. The latter category was probably overly broad and removed a significant number of valid profiles, and since they were

concentrated in regions, it was sufficient to hamper the assimilation. We currently hope that work on the inversion side of the process (where error information about the inversion is available) can do a better job of removing profiles and improve the assimilation. At present the only solution would be to manually check all 177,000 profiles.

In order to insure that the MGCM matches the actual observations, the MGCM was run with conditions as near to that of the data as possible. This was done in several ways. First, the season of the initial model states was identical to that of the data. This insures, to the extent that the MGCM seasons are right, that the data and model are similar. The same season was also used to build gains covering the period of the observations. Secondly, the MGCM time—it primarily uses the hour since the cold start—of the observations was selected to insure that the local time of day was correct for each observation. This allows the observations to modify the correct state of the atmosphere. Finally, the opacity (or dustiness) used in the model was twice the 9 μm to 10 μm infrared opacities estimated from the TES observations [*Kieffer et al.,* 1992]. This results in $\tau = 0.32$. While there may have been some spatial and temporal variability of the dust seen by TES [*Smith et al.,* 1998], the MGCM is designed to have a constant mass density of dust.

5.3 Validating the Assimilation

While the assimilation technique works with the synthetic data, it is necessary to also verify that it works with the actual data. This is especially important since the synthetic testing was done with a twinned model system. For the assimilation to be working, it needs to reduce the analysis error (when being compared to the actual state of the atmosphere) compared to what the MGCM running alone would do (the prediction error). Unlike the synthetic case, the truth is not actually available for comparison (if it were, there would be no need for data assimilation). While it is possible, using several different Mars GCM's to avoid the twinned problem, we currently do not have access to another one. And even then, there are still issues of

the actual observational error and fundamental atmospheric assumption that require some sort of validation of the assimilation process with actual observations.

We use several methods for validating the results of the actual assimilation process. The first requirement for the assimilation to be successful is that it needs to modify the atmospheric state. While the climatology could be correct and not require modification, it is extremely unlikely that the model weather would correspond to the actual weather. Thus, in order to be successful, the assimilation process needs to modify the model. Figure 5.2 shows that this is the case, despite the very low f (0.075) needed for the process to be stable. Here, the solid line is the RMS temperature difference (D_T) between two runs of the MGCM at the season of the data with different initial conditions. There is no assimilation and the difference gives a feel for the meteorological variability on Mars at this season. The dashed line is the RMS temperature difference (D_T^a) between the assimilation results and the same initial conditions without assimilation. The vertical dashed lines represent the beginning and end of the available data. As expected, the difference is zero until data is assimilated. Once the assimilation begins, the difference jumps rapidly each time a close pass is made (see Figure 5.1). The analysis and prediction soon differ by more than the inherent variability; which reflects the modification of the model's climate.

Tests with synthetic data showed that it is possible for the assimilation process to modify the analysis model without actually improving it (i.e., without decreasing the difference between the analysis and truth). So, while changing the model is necessary for successful data assimilation, it is not sufficient. In order to determine if the assimilation is actually driving the model towards the state of the Martian atmosphere, we use a second test. We assimilate the TES observations using two different initial conditions–both at the correct season. The results of the two runs are then compared to determine if they are closer to each other than random weather. As shown by Figure 5.3, the assimilation process is successful. The curves are the RMS temperature differences (D_T) between the two analysis models for several different f values. $f = 0$ is the case where no assimilation occurred and is effectively the expected weather difference for this season. Like in Figure 5.2, the vertical dashed

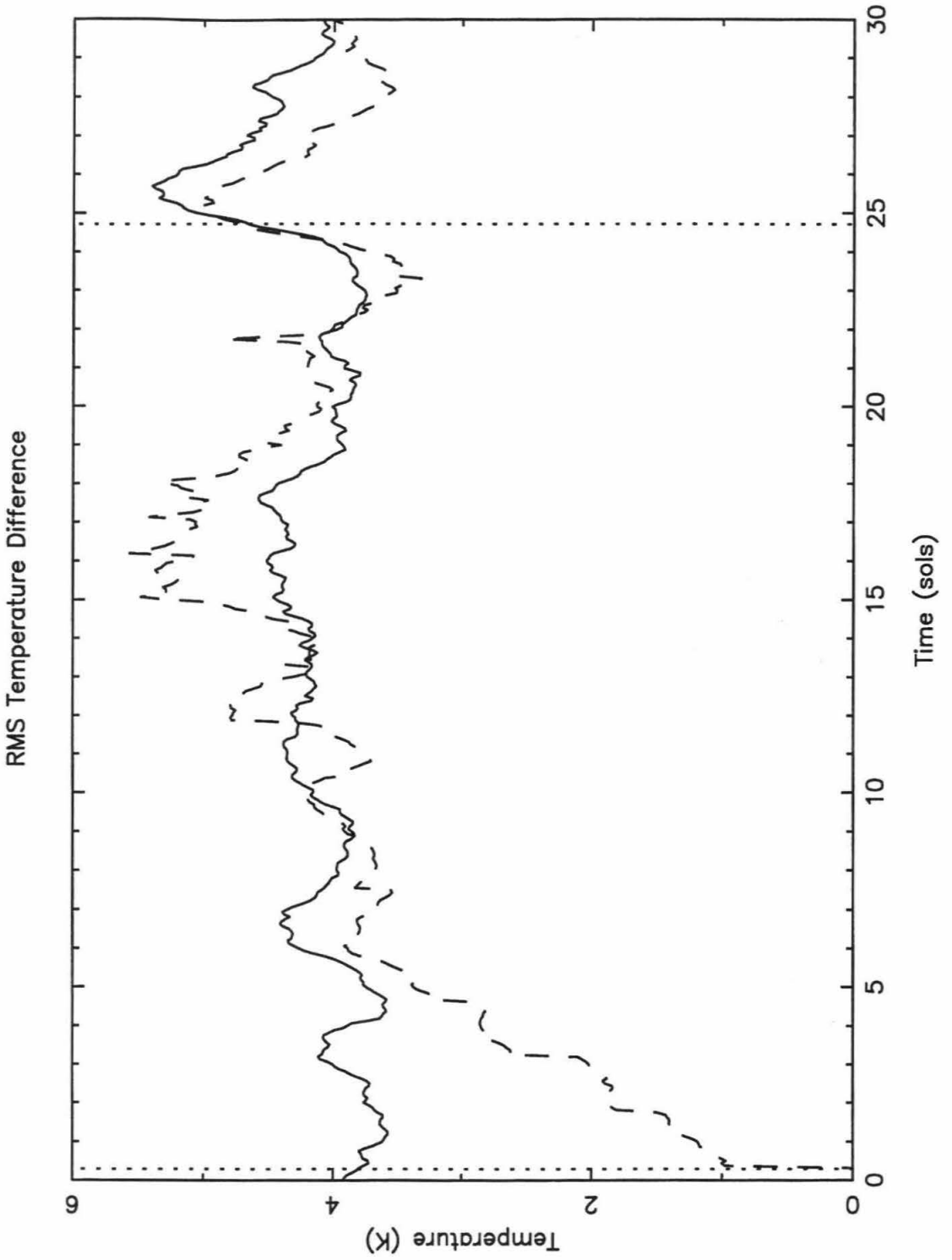

Figure 5.2: RMS Temperature Difference Due to Assimilation
The solid line is the variability due to weather at this season (D_T). the dashed line is the difference between the assimilation results and the same initial conditions without assimilation. The dotted lines mark the beginning and end of the available data.

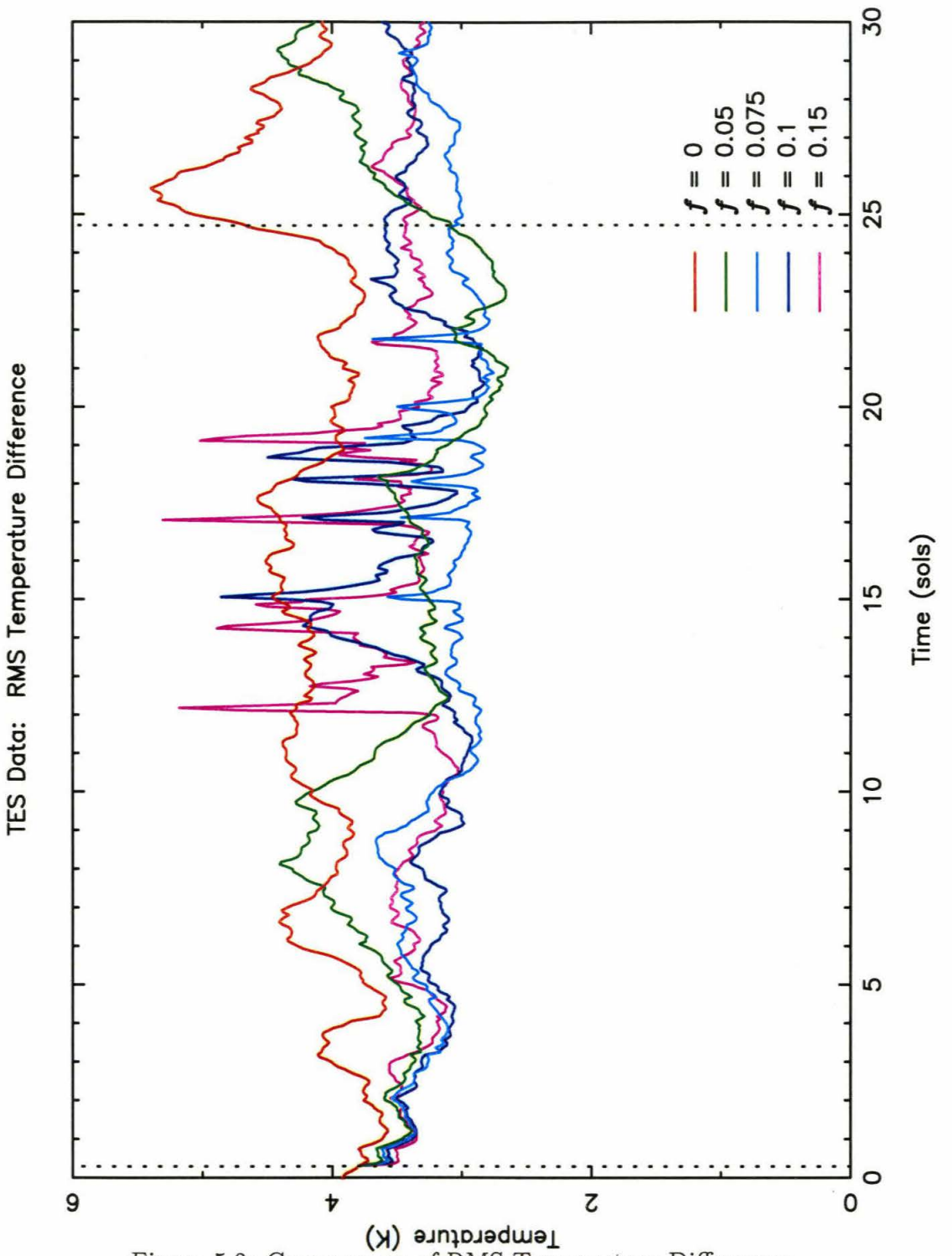

Figure 5.3: Convergence of RMS Temperature Differences

The RMS temperature difference D_T between two analysis models with the indicated f value, but different initial conditions. The red case, $f = 0$, shows the basic variability of the weather. The other cases indicate the convergence of the two initial conditions due to the data assimilation.

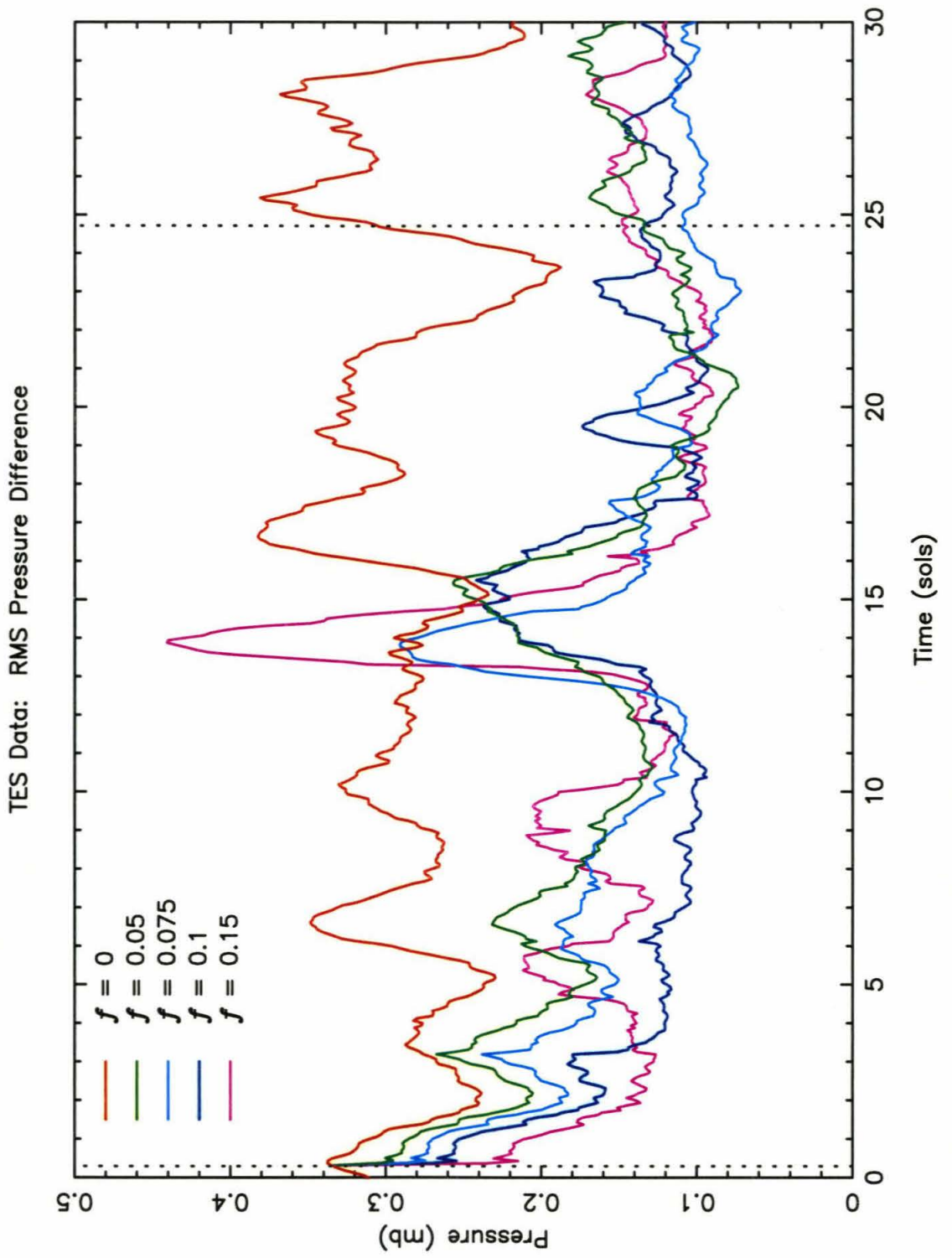

Figure 5.4: Convergence of RMS Pressure Differences
The RMS pressure difference D_P between two analysis models with the indicated f value, but different initial conditions.

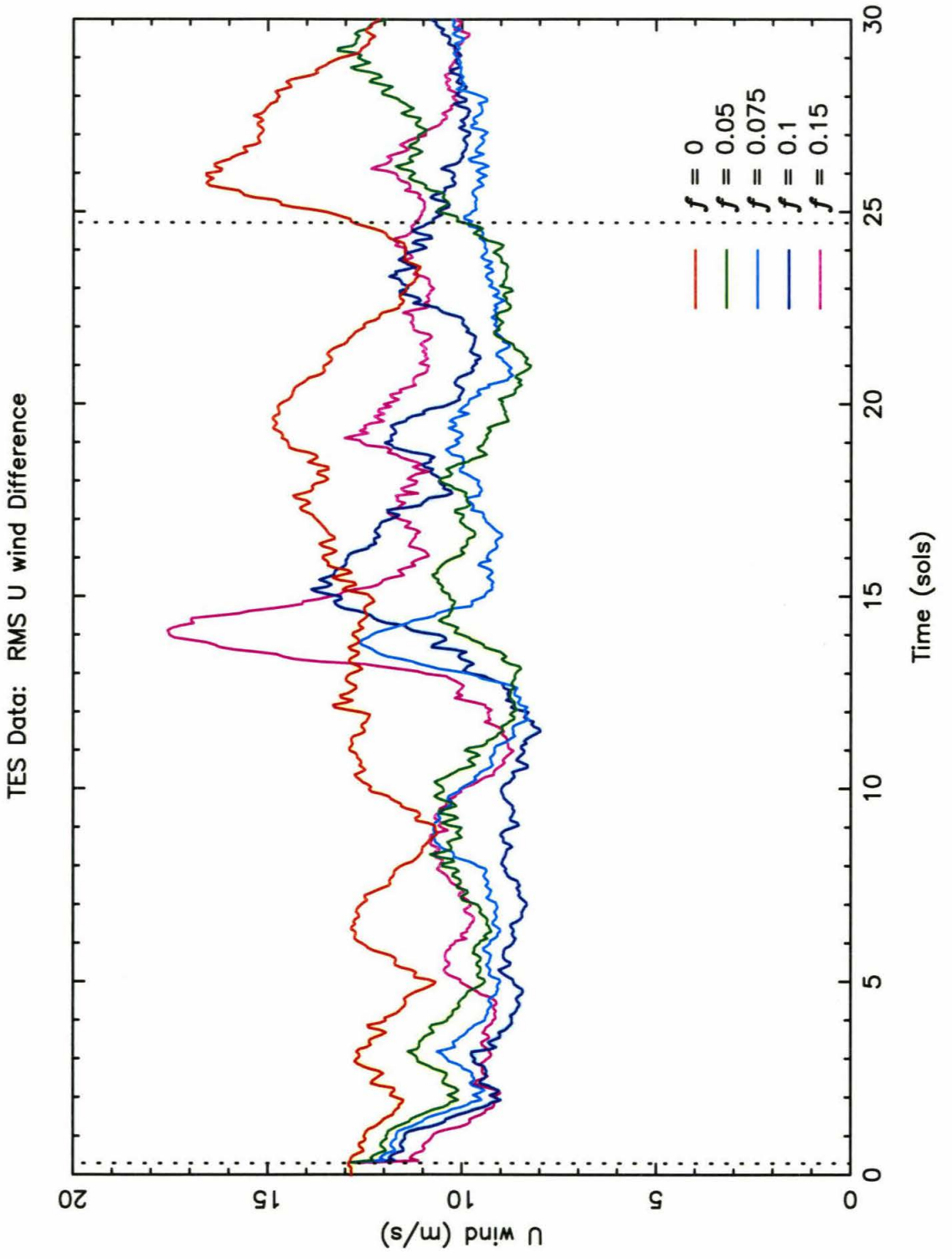

Figure 5.5: Convergence of RMS U Wind Differences
The RMS U wind difference D_U between two analysis models with the indicated f value, but different initial conditions.

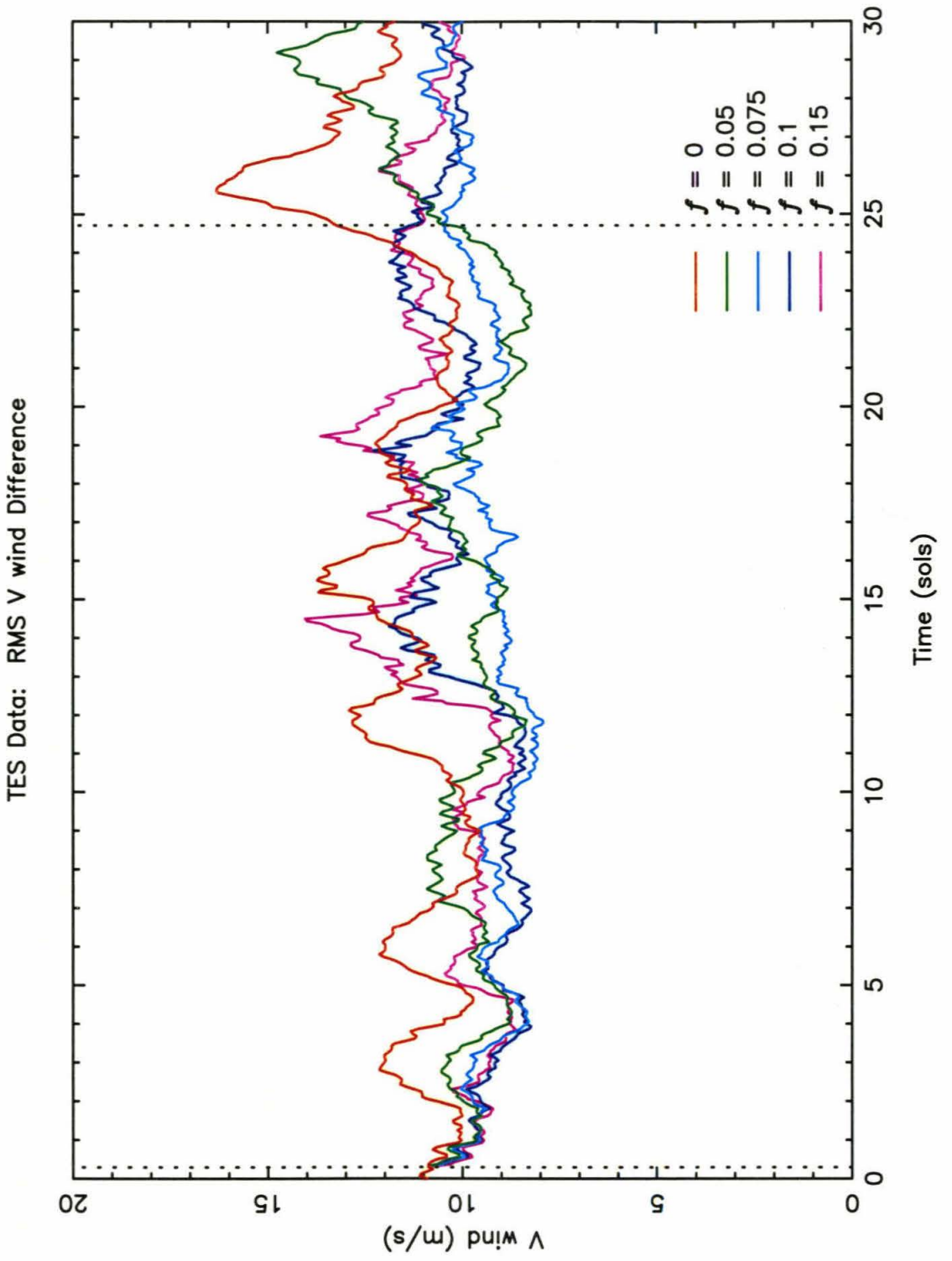

Figure 5.6: Convergence of RMS V Wind Differences
The RMS V wind difference D_V between two analysis models with the indicated f value, but different initial conditions.

Goodness Values for Assimilating TES Data

	Goodness				
	G_P	G_T	G_U	G_V	G_{Mean}
$f = 0.05$	0.40	0.71	0.63	0.77	0.63
$f = 0.075$	0.34	0.62	0.62	0.71	0.57
$f = 0.1$	0.29	0.68	0.66	0.81	0.61
$f = 0.15$	0.38	0.74	0.78	0.92	0.70

Table 5.1: Goodness Values for Various f
The G value (see equation 4.7) for each dynamical variable from the assimilation of the TES data. The mean is the average of the four fields to get a total fit. The model is unstable for $f \gtrsim 0.2$.

lines indicate the beginning and end of the available data. As f increases, it reaches a "best" assimilation (the case where the difference between the two assimilated curves is the smallest) and then worsens again. While these curves involve different components than in chapter 4, it is possible to calculate a goodness, G, in this case as well (Table 5.1). For $f \gtrsim 0.2$, the assimilation is unstable (much like for large f values in the synthetic case). The results indicate that the optimum $f = 0.075$. Looking at the actual curve, this is somewhat deceiving. For the first 12 sols, slightly higher values are better (and are more decoupled from the initial conditions), but then there is a change in the quality of the assimilation and the $f = 0.075$ case ends up being better over the entire period. This is especially true in the other dynamical fields (Figures 5.4 to 5.6). For the remainder of the chapter, the case with $f = 0.075$ will be used.

5.4 Results of Assimilating TES Data

While looking at the TES results, we will compare them to the MGCM at the same season. The dynamics of the MGCM are already well described [*Pollack et al.,* 1990, *Haberle et al.,* 1993, *Hollingsworth et al.,* 1996], thus we can build upon the results of the MGCM and only discuss the differences and their implications. Furthermore, the data assimilation, using the available TES data, cannot fully correct the fundamental

climatological assumptions of the MGCM.

There are several interesting components to the atmospheric dynamics of Mars at this season. The first is variable or changing part of the dynamics–the weather. At this season, the main weather features are traveling baroclinic waves in the Northern hemisphere. The second component of interest is the zonal mean climatology. Due to the restricted temporal coverage, we really only sample one season and thus cannot study the changing of the Martian climatology over the seasons.

A third interesting component are the thermal tides. Looking at the maps of surface pressure (where the tides are the most visible, Figure 5.15) we see very few differences between the tides with assimilation and without (upper and lower panel respectively). The pressure field expresses the tides as "waves" along the equator, synchronized west of the sub-solar longitude. The first possibility is that since the tides are controlled almost entirely by the solar forcing, the MGCM is correctly modeling them. The other possibility is that the radiative forcing of the model tides is too strong for the data to modify.

5.4.1 Baroclinic Waves

During the season of the TES observations, one of the major features of the Martian atmosphere is a series of baroclinic waves (or storms) that encircle the planet in the northern hemisphere [*Zurek et al.,* 1992, *Barnes et al.,* 1993, *Barnes,* 1980]. These waves were observationally detected by the Viking Lander meteorology packages [*Tillman et al.,* 1979, *Barnes,* 1980]. They are also well expressed in the MGCM [*Barnes et al.,* 1993].

The waves (or storms) can clearly be seen in the global maps of the pressure fields (Figures 5.7 to 5.18). These maps are similar to the ones in chapter 4 (Figures 4.9, 4.10, 4.11 and 4.12). The mean surface pressure at each grid point (over all three MGCM runs) for the 30 sols was subtracted from each map. Then all the deviations were normalized using the same color scale. For each cylindrical projection, there is the corresponding northern polar projection (down to 45N) using the same data

and color scale. The background contours are the GCM topography with dashed contours below the mean pressure and solid ones above it. Note that there is a bit of saturation in Figures 5.7 and 5.13 (and in the corresponding polar plots 5.8 and 5.14), primarily in the maps of the assimilation results. The sub-solar longitude is indicated by the sun symbol between the two maps (and the symbol at the outer edge in the polar plots). There are two figures (containing three different pressure maps) at each of three times during the model run. Each figure is followed by the corresponding polar plot of the same data. The first figure in each pair contains the pressure map from two different assimilation attempts. The second figure contains one of the assimilation cases and the corresponding MGCM initial condition with no assimilation.

At sol 3.5, while the data assimilation has significantly affected the state of the model, it has not yet fully decoupled the model from the initial state (see Figure 5.4). This can be seen in the differences between the two assimilation maps (Figure 5.7) and in the similarities between the TES and MGCM fields (Figure 5.13). By sol 17, the assimilation is in control of the pressure field and there is little resemblance of the storm belt between the MGCM and TES results (Figures 5.15 and 5.16). Furthermore, the two TES results (Figures 5.9 and 5.10) are very similar, especially in the phasing of the baroclinic waves. This continues throughout the rest of the assimilation as shown by the last maps (Figures 5.11, 5.12, 5.17 and 5.18) at sol 24, almost at the end of the data.

When looking at the raw pressure data, the largest variability between independent model cases is in the northern hemisphere, extending poleward of 30N. This is also one of the regions where the two assimilation cases are the most similar. This is clearly visible in the polar projections of the data. Thus the data assimilation is significantly modifying the pressure fields (despite the fact that TES is measuring temperatures). The figures with two assimilation maps (Figures 5.7 through 5.12) show that the data is actually driving the MGCM to a common state. The midlatitude and polar regions of both cases with assimilation are similar. This is especially true when comparing the assimilation cases with the corresponding case without as-

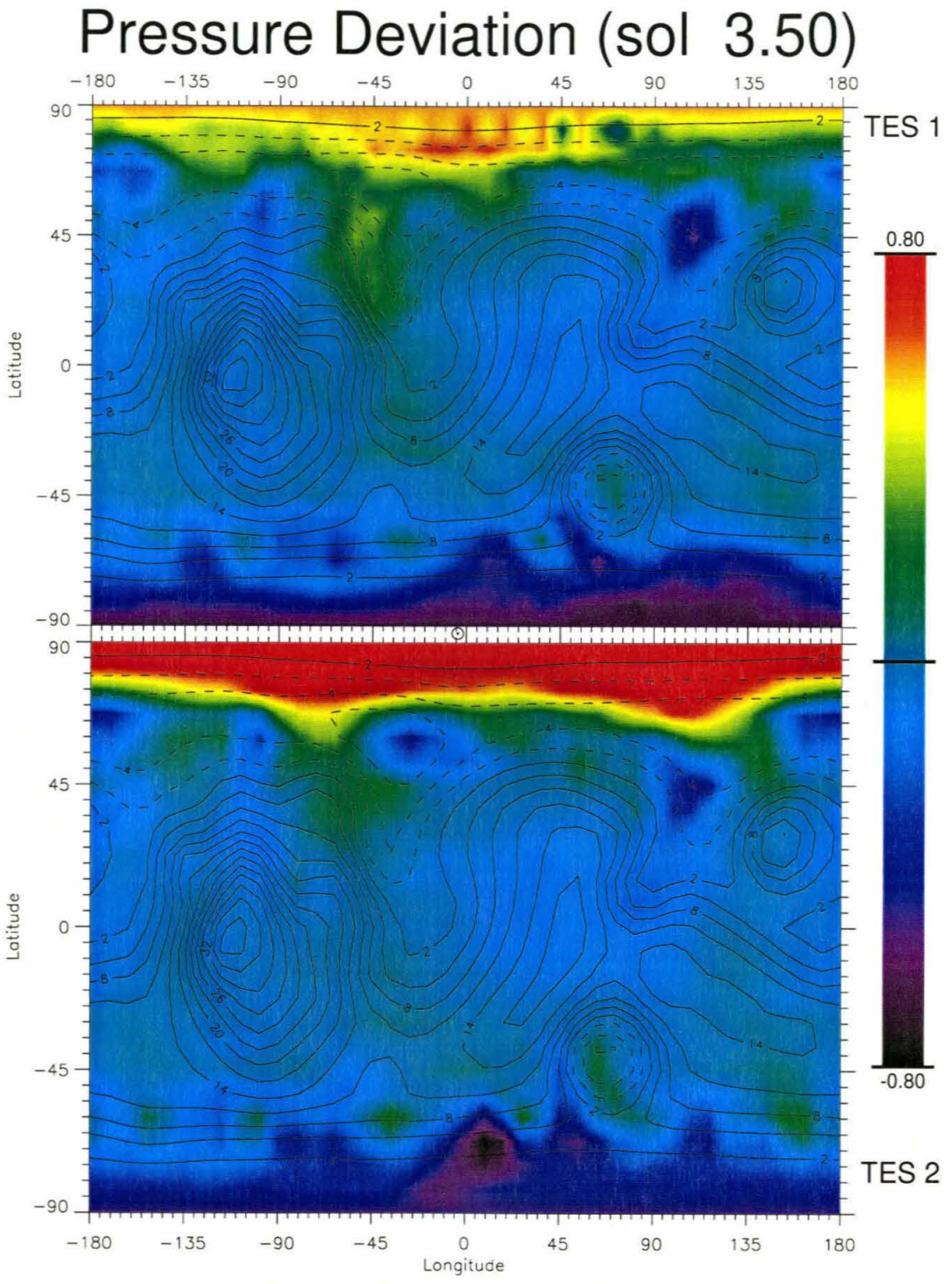

Figure 5.7: Assimilation Pressure Maps at Sol 3.5

Surface pressure deviation for each point in the assimilation model at sol 3.5 for two different initial conditions. The same mean pressure was subtracted from each map. Note that the pressure scale is saturated at the north pole of the TES 2 map.

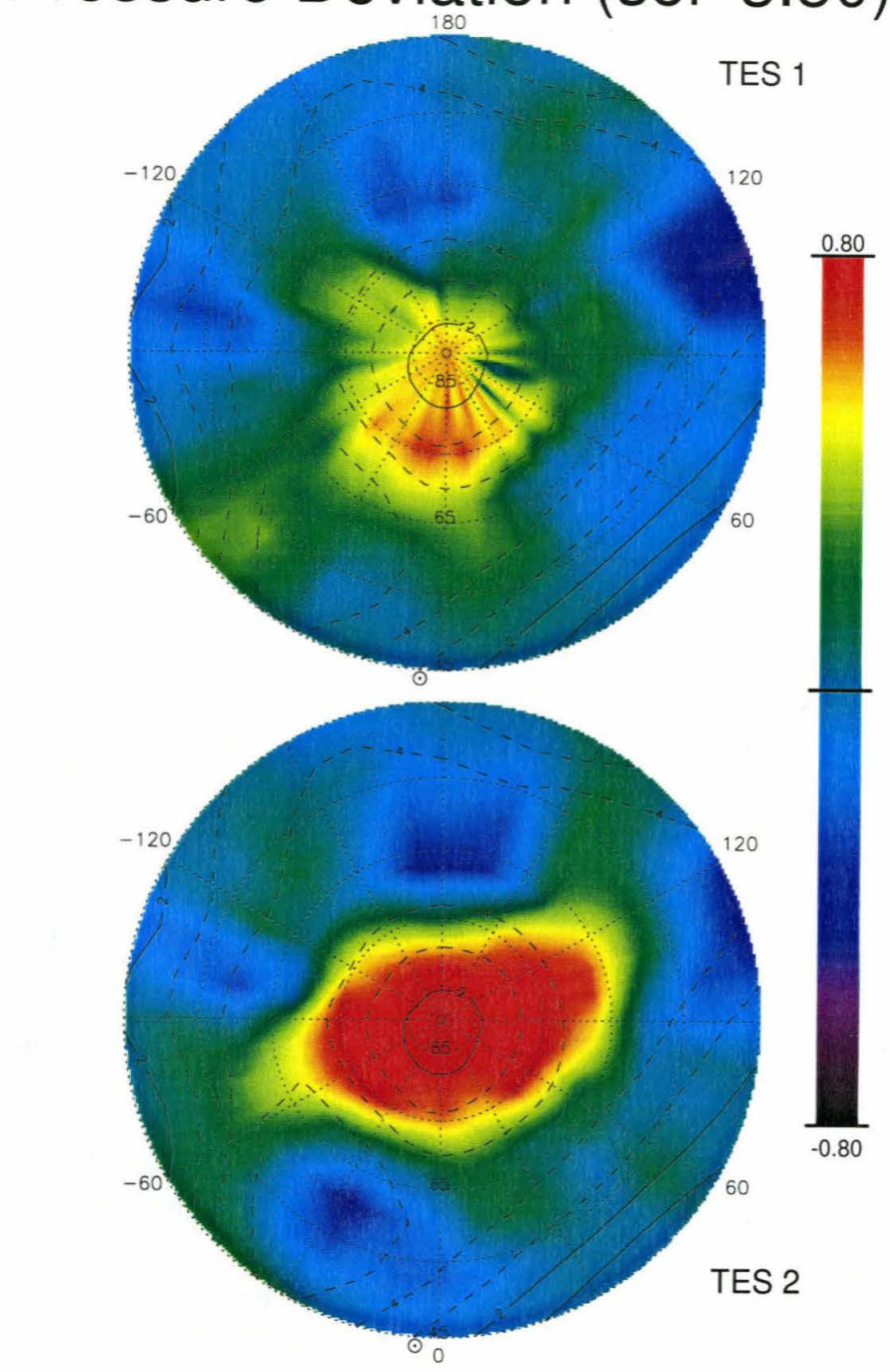

Figure 5.8: Polar Assimilation Maps at Sol 3.5
Northern hemisphere polar projection of assimilation model surface pressures. This is the same data as 5.7, but a polar projection. The figure covers the northern hemisphere down to 45N.

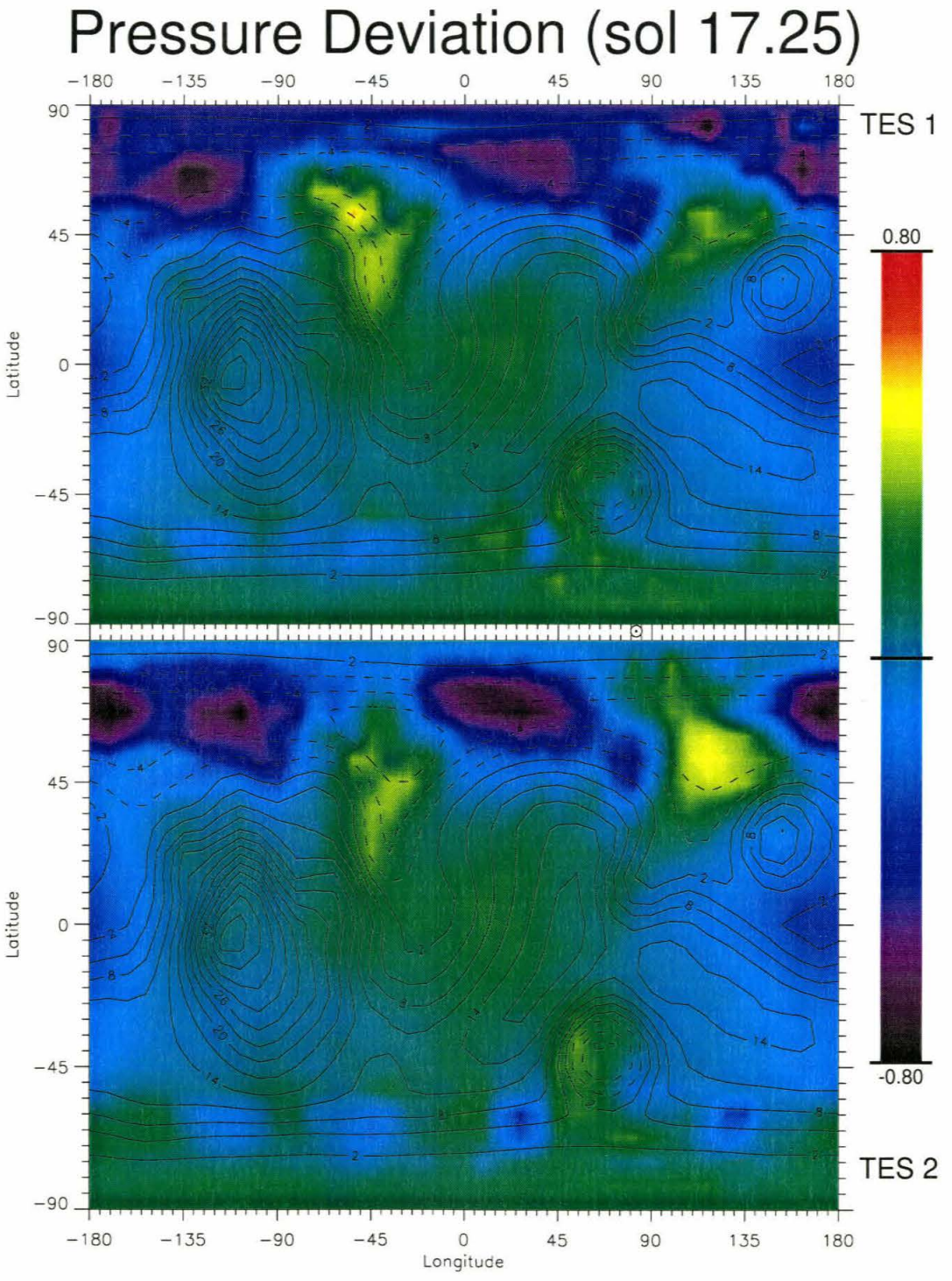

Figure 5.9: Assimilation Pressure Maps at Sol 17.25
Surface pressure deviation for each point in the assimilation model at sol 17.25 for two different initial conditions. The same mean pressure was subtracted from each map. These are the same two model runs as in Figure 5.7.

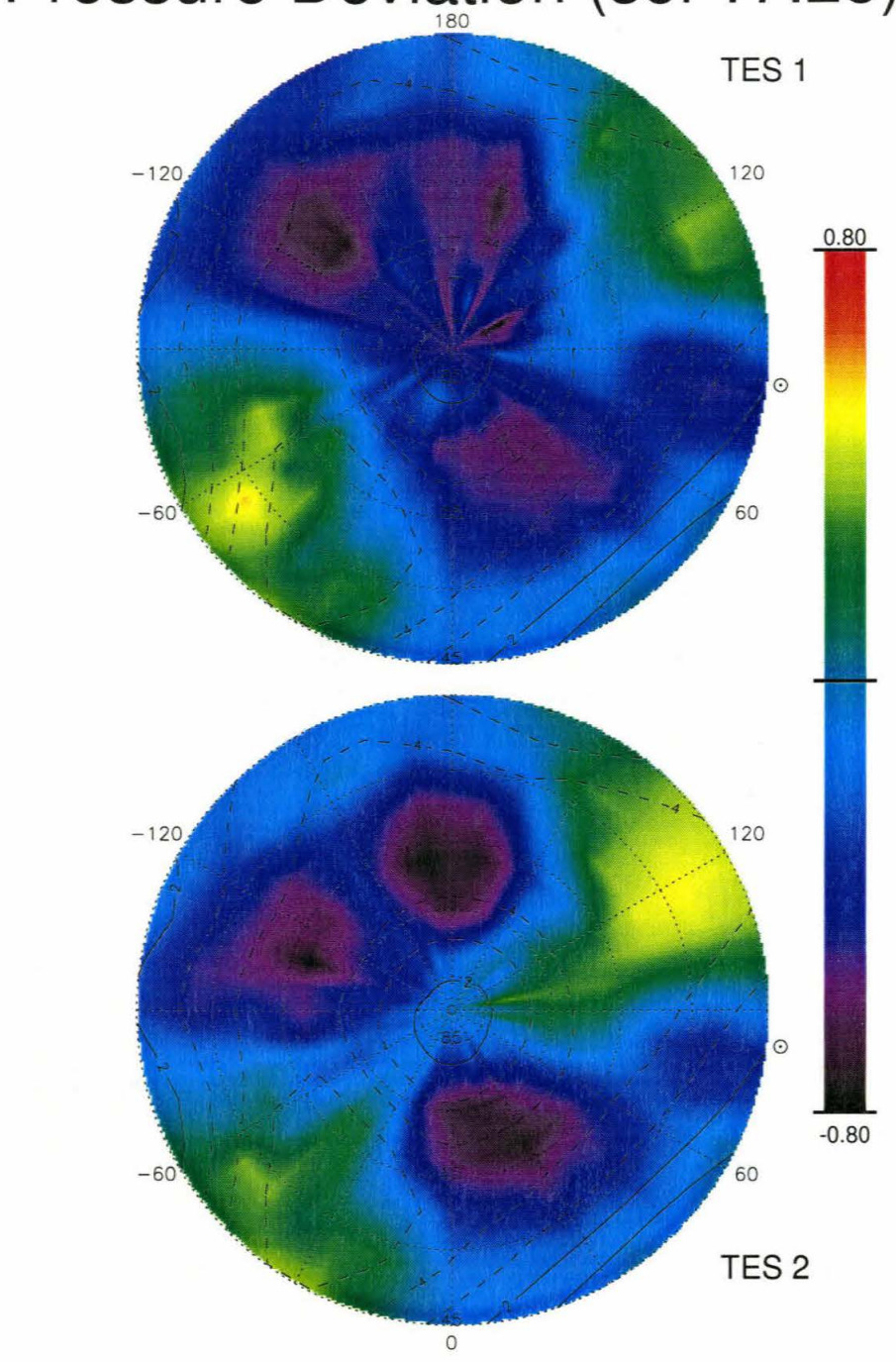

Figure 5.10: Polar Assimilation Maps at Sol 17.25
Northern hemisphere polar projection of assimilation model surface pressures. This is the same data as 5.9, but a polar projection. The figure covers the northern hemisphere down to 45N.

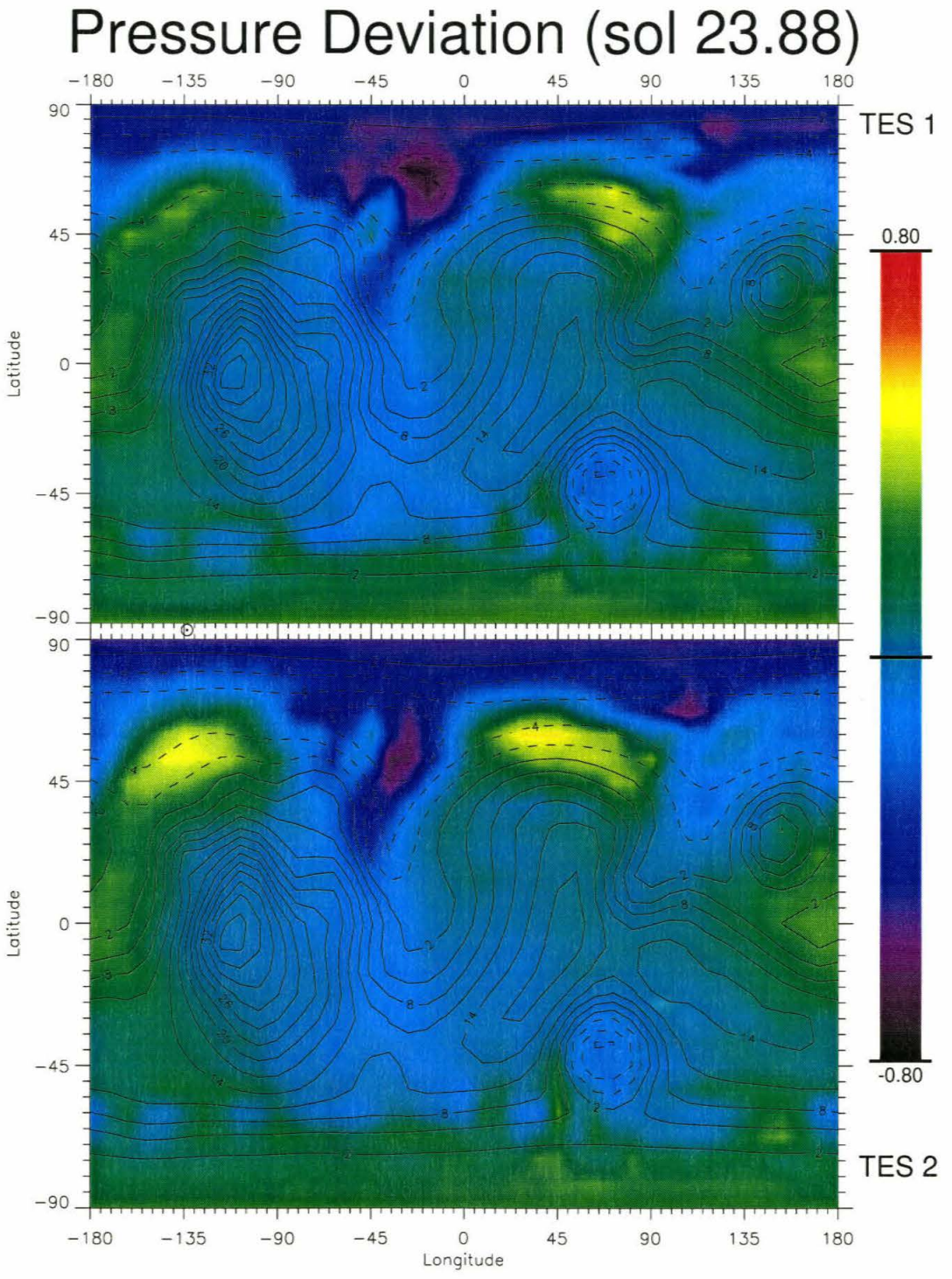

Figure 5.11: Assimilation Pressure Maps at Sol 23.88
Surface pressure deviation for each point in the assimilation model at sol 23.88 for two different initial conditions. The same mean pressure was subtracted from each map. These are the same two model runs as in Figure 5.7.

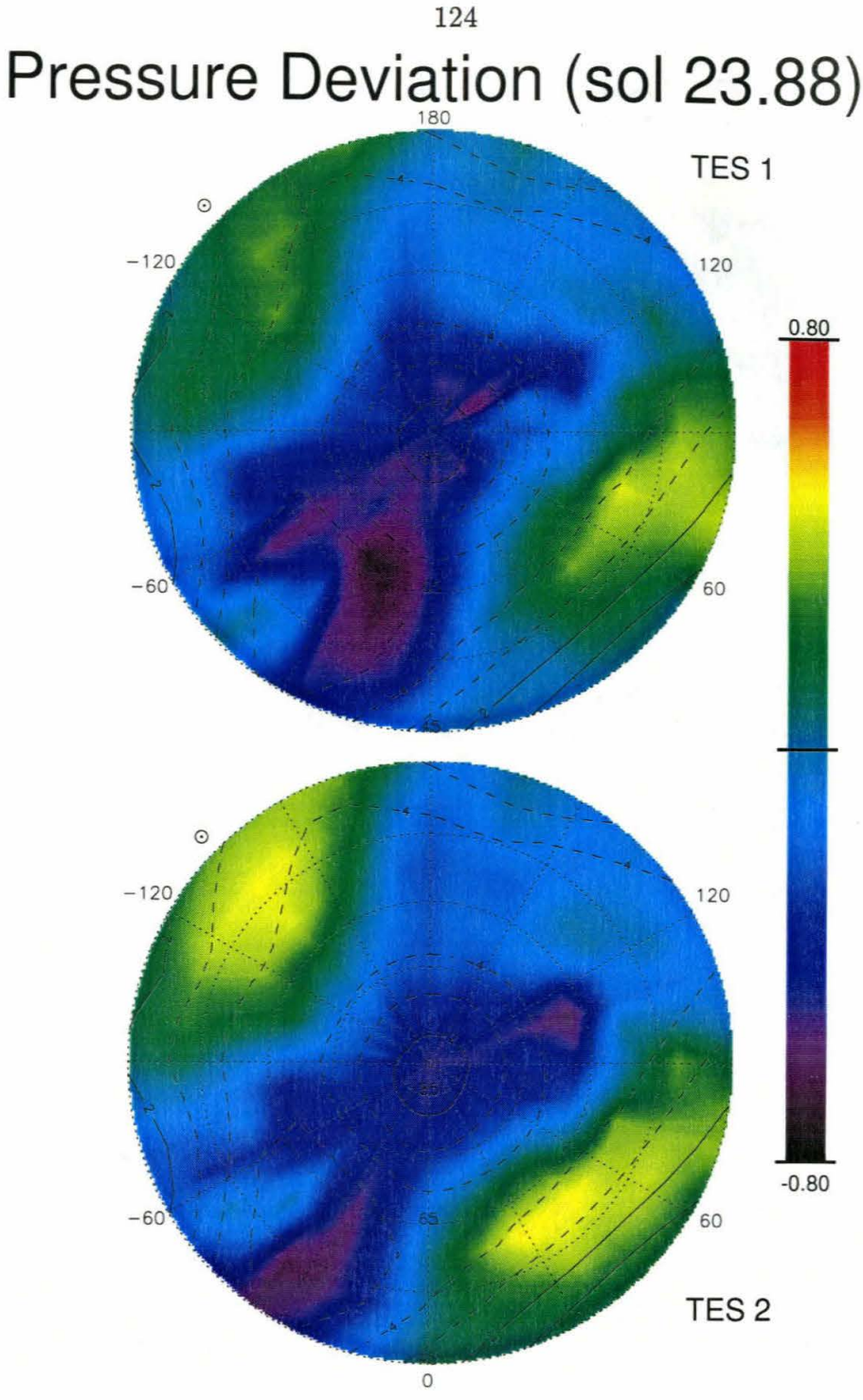

Figure 5.12: Polar Assimilation Maps at Sol 23.88
Northern hemisphere polar projection of assimilation model surface pressures. This is the same data as 5.11, but a polar projection. The figure covers the northern hemisphere down to 45N.

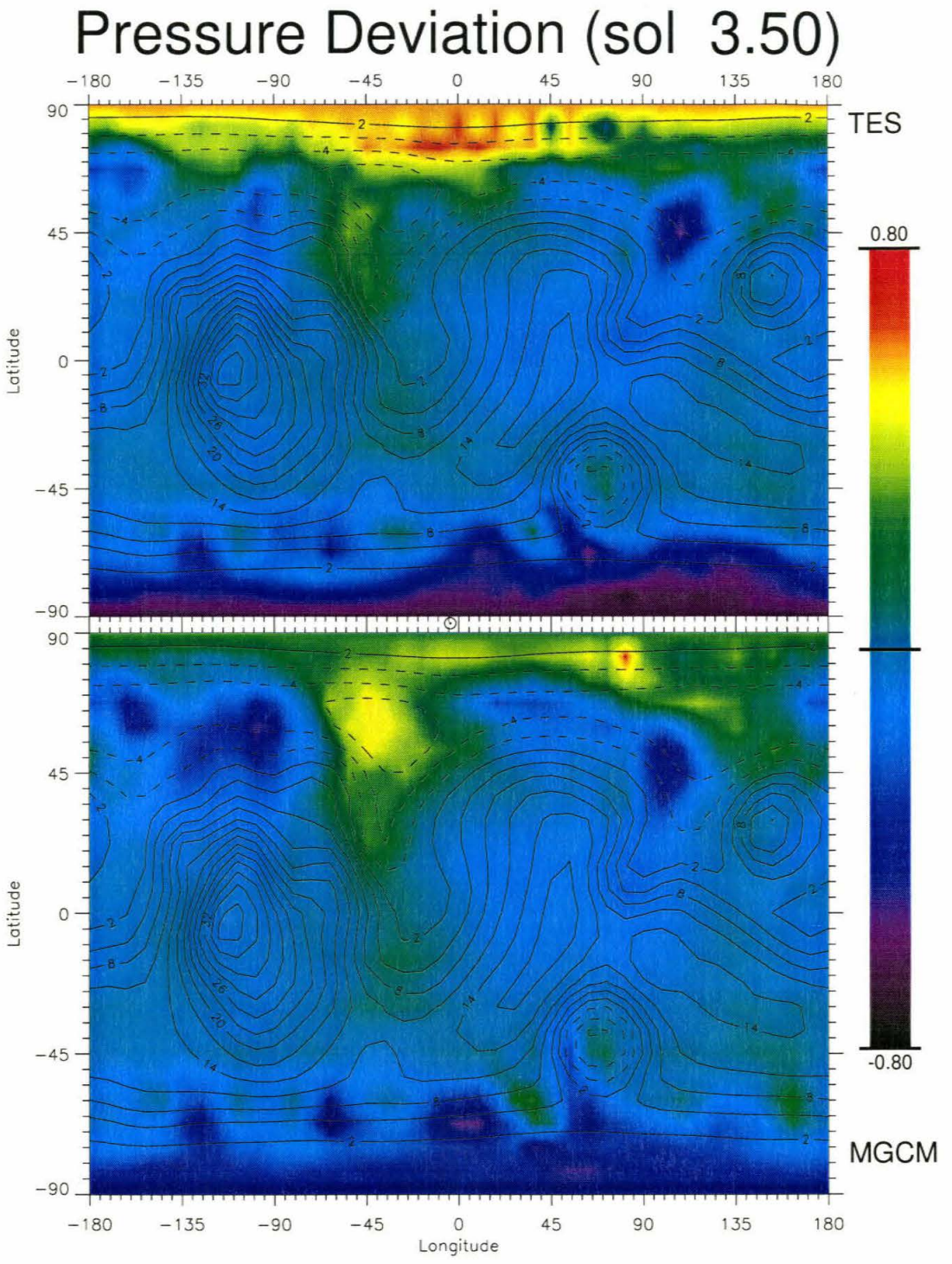

Figure 5.13: Pressure Maps at Sol 3.5

Surface pressure deviation for each point of the model at sol 3.5, with and without assimilation. The maps are processed identically to Figure 5.7. The lower map is the MGCM running with no assimilation with the same initial condition as the case with assimilation.

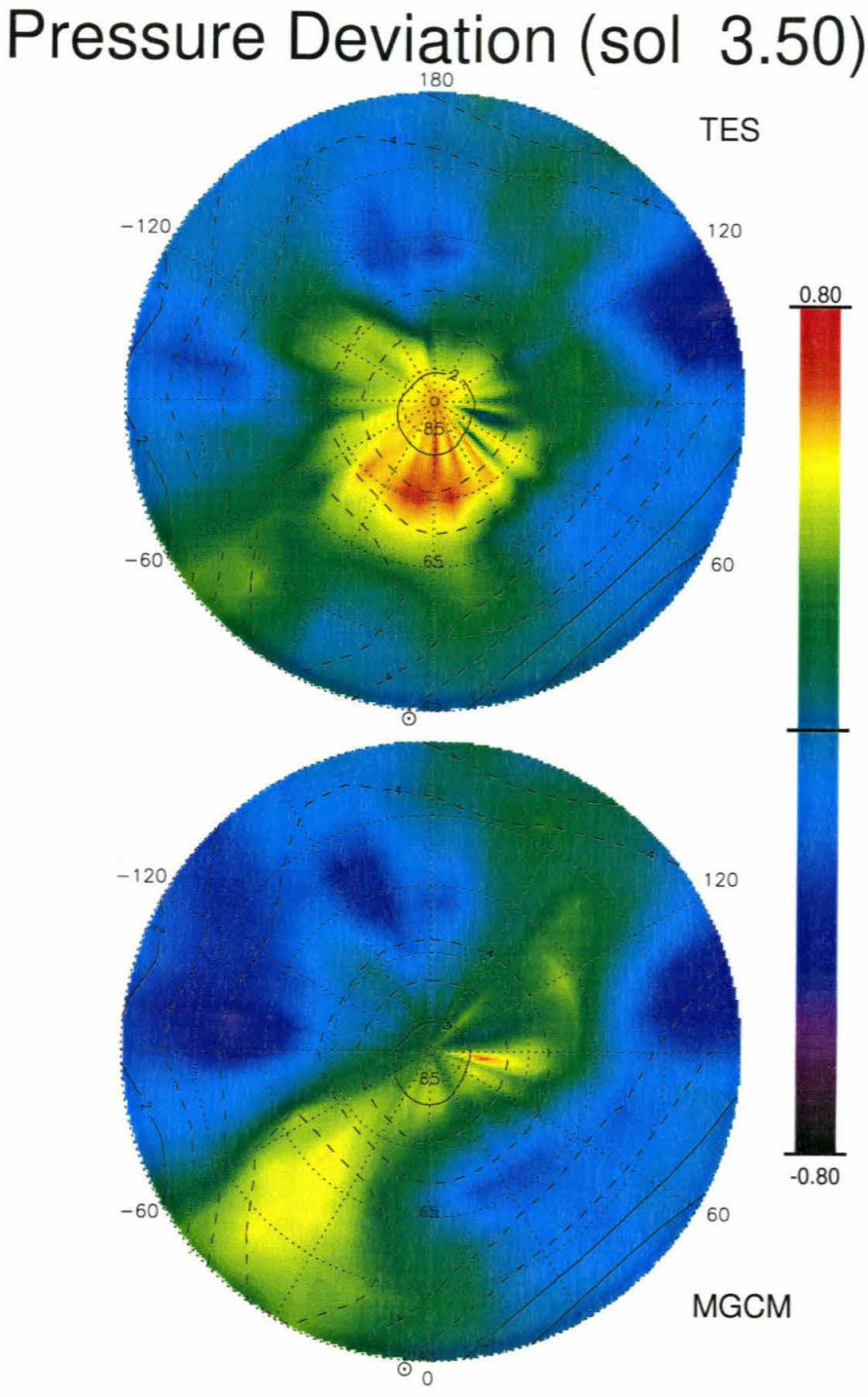

Figure 5.14: Polar Pressure Maps at Sol 3.5
Northern hemisphere polar projection of assimilation model surface pressures. This is the same data as 5.13, but a polar projection. The figure covers the northern hemisphere down to 45N.

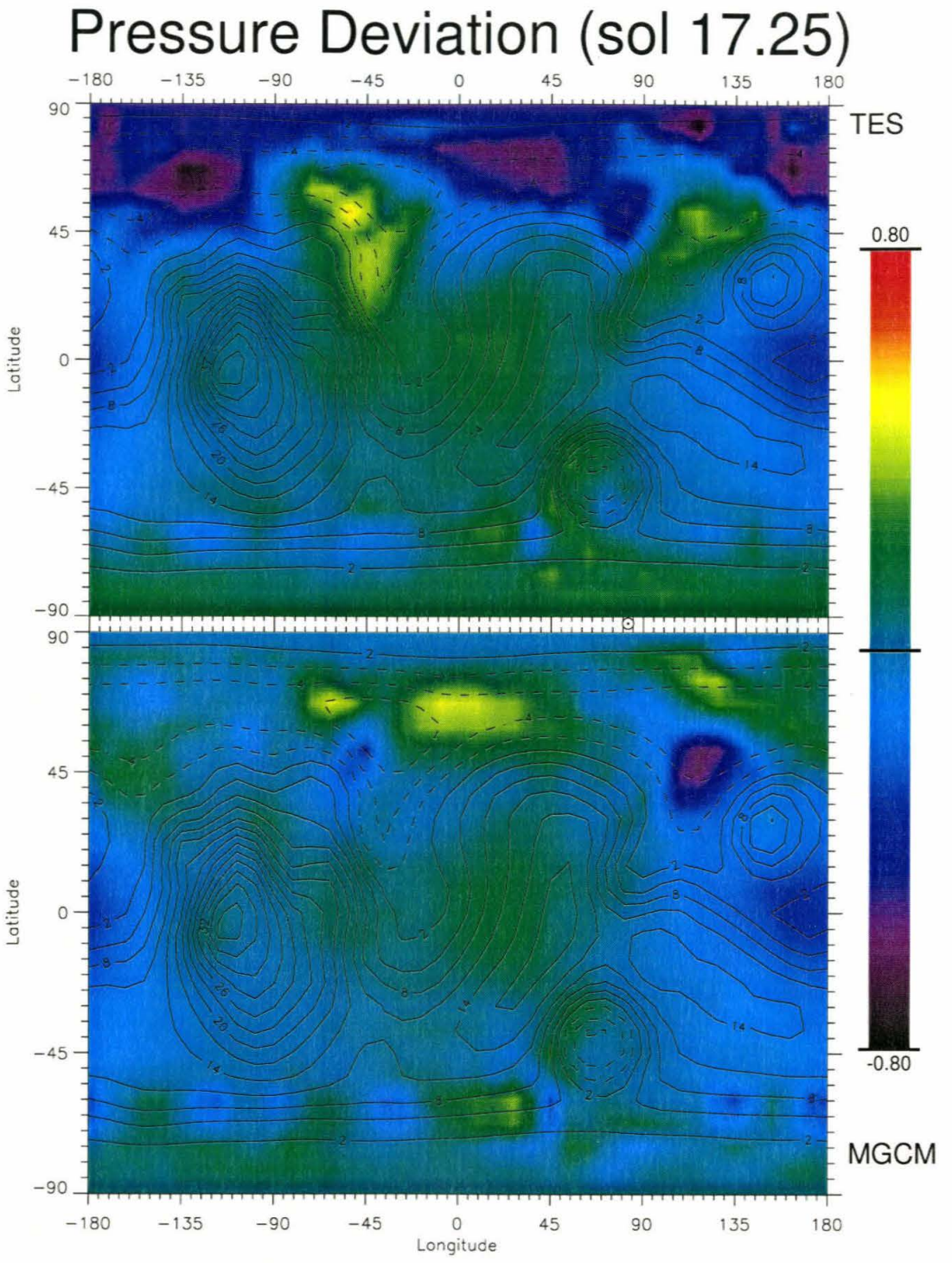

Figure 5.15: Pressure Maps at Sol 17.25

Surface pressure deviation for each point of the model at sol 17.25, with and without assimilation. The maps are processed identically to Figure 5.9. The lower map is the MGCM running with no assimilation with the same initial condition as the case with assimilation.

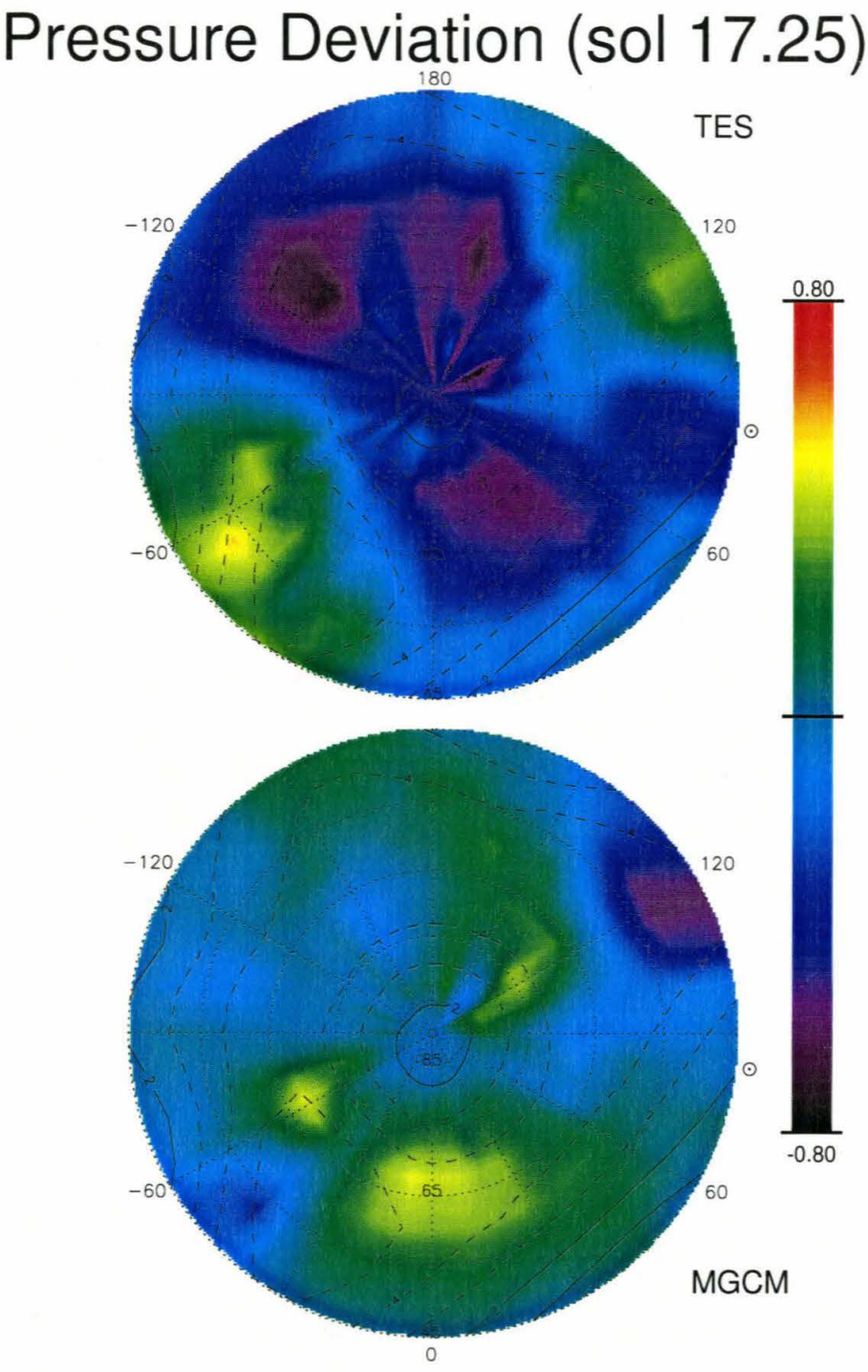

Figure 5.16: Polar Pressure Maps at Sol 17.25
Northern hemisphere polar projection of assimilation model surface pressures. This is the same data as 5.15, but a polar projection. The figure covers the northern hemisphere down to 45N.

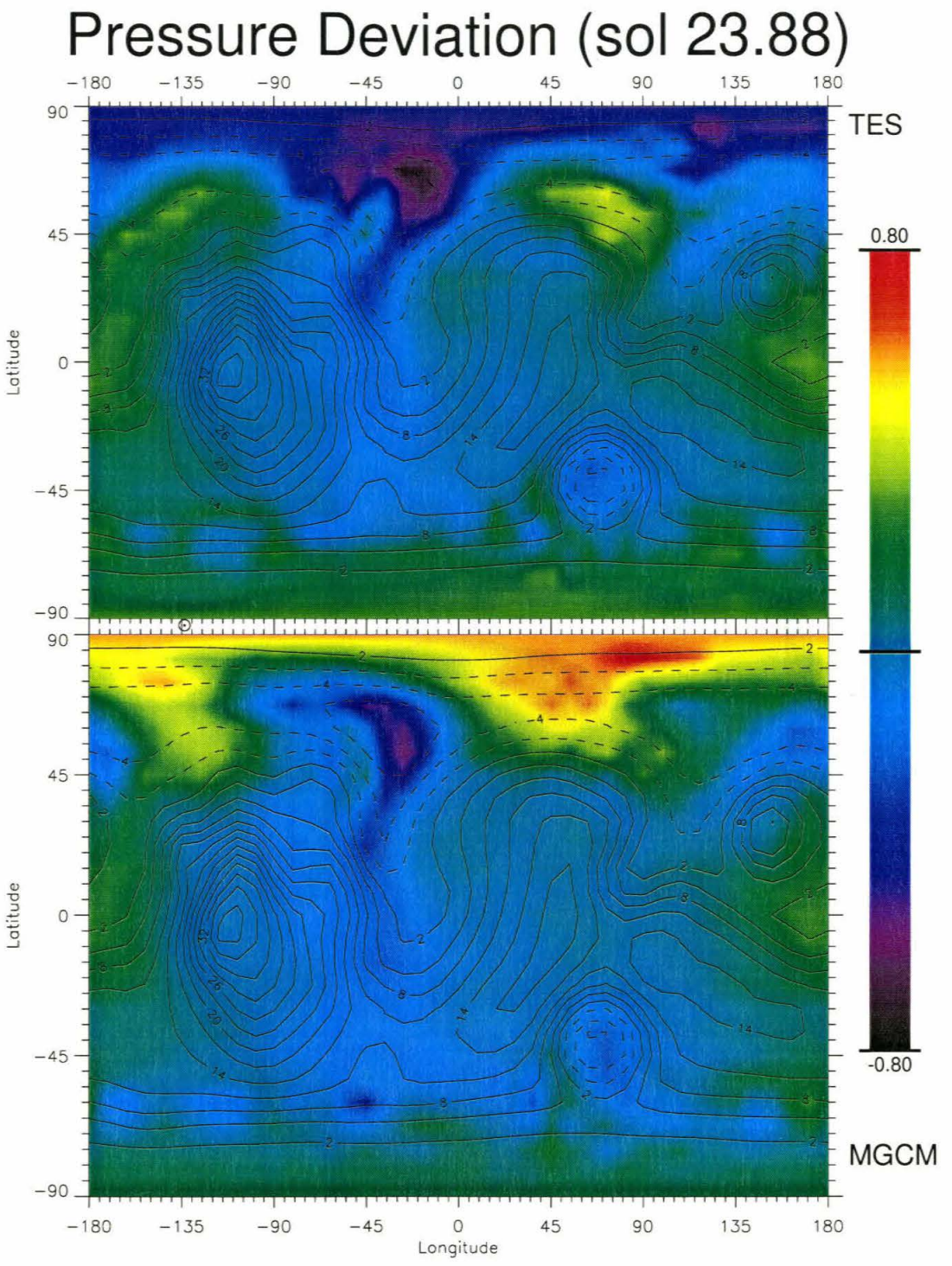

Figure 5.17: Pressure Maps at Sol 23.88

Surface pressure deviation for each point of the model at sol 23.88, with and without assimilation. The maps are processed identically to Figure 5.11. The lower map is the MGCM running with no assimilation with the same initial condition as the case with assimilation.

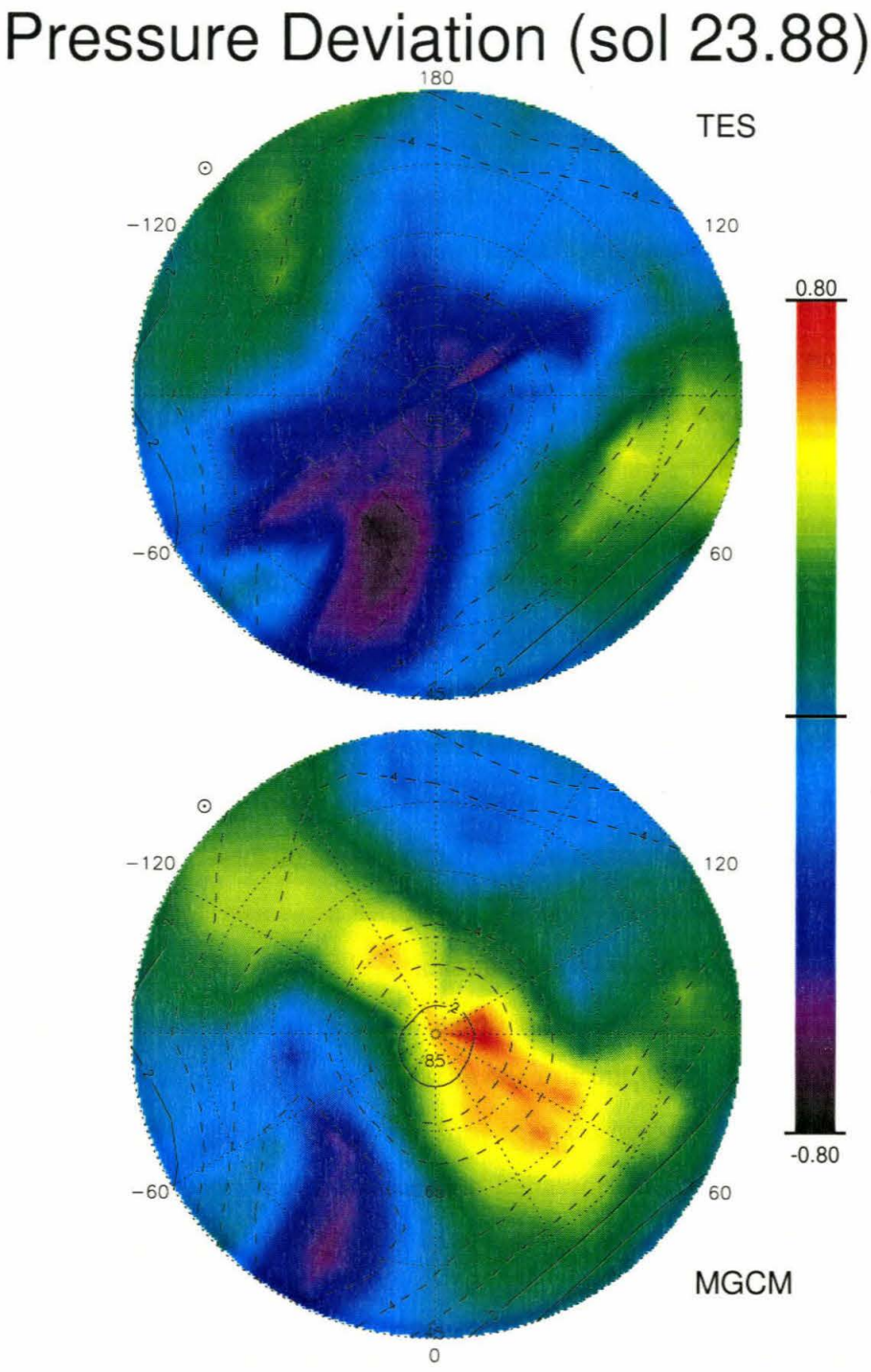

Figure 5.18: Polar Pressure Maps at Sol 23.88
Northern hemisphere polar projection of assimilation model surface pressures. This is the same data as 5.17, but a polar projection. The figure covers the northern hemisphere down to 45N.

similation (Figures 5.13 through 5.18).

One of the recognized features of the northern storm belt is that it tends to have two "modes." In the first mode, the storms have a low zonal wavenumber ($m \leq 2$) while the second mode is characterized by higher wavenumbers ($m \approx 3$ to 5) [*Zurek et al.,* 1992, *Collins and James,* 1995]. In the pressure maps, m can be determined by counting the number of pressure maximums (or minimums) around the planet in the storm belt. These are not nice evenly spaced waves and m is more a measure of the number of storms and their mean size than actual wave characteristics. Modeling (as well as the Viking Lander data) shows that the atmosphere tends to change between the two modes occasionally [*Collins et al.,* 1996, *Barnes et al.,* 1993]. This is also visible in the assimilation results when viewed in a polar projection. At sol 3.5 (Figure 5.8), $m \approx 4$, whereas at later times m decreases to 3 (Figure 5.10, for the minimums, there are only really two obvious maximums) and finally 2 (Figure 5.12). It is not clear whether the high m regime at sol 4 is primarily due to the assimilation model not yet being fully decoupled from its initial conditions or is actually in the data. The change in model occurs as storms move faster than the one ahead of them and catch it. One occurrence can be seen in Figure 5.10 between longitudes -120 and -180. There appear to be several locations, notably around longitude 90E where the storms slow down and allow the following ones to catch up. When this occurs, the later one either fades out or merges with the one in front. At $m \approx 2$, the waves appear to be sufficiently strong and spread out that they do not catch each other. Both the MGCM and assimilation models show this behavior, although it appears to occur sooner in the assimilated model (but at much the same time in both assimilation runs).

There is a difference in speed between the MGCM and assimilation waves with hints that the assimilation waves have a higher phase speed, but this is difficult to quantify. Due to the sparse data, the location (and possibly the number) of the storms is corrected by the close passes (red profiles in Figure 4.1). Then they travel for a sol or two under the influence of the MGCM. At that point, the next orbit starts and it causes the storms to jump and be reset, but since both assimilation

cases end up similar, it appears that the jump actually reflects the atmosphere of Mars. Unfortunately, the discontinuous motion makes it difficult to determine the actual speed of the features.

A second difference between the assimilation waves and those of the MGCM are that the assimilated waves are much stronger. This is especially true in the low m modes. The low pressures are much lower while the high pressures are about the same (see Figure 5.15 or 5.16). This is less clear at sol 24, but is still the case. At sol 3.5 (Figures 5.7, 5.8, 5.13 and 5.14), it is much harder to define a pattern in the assimilation results (possibly due to the assimilation not having completely corrected the model), but again, some of the waves appear to be stronger than in the case without assimilation. One interesting change is that the waves are initially between longitudes $\sim$ 45N and $\sim$ 70N (in both the data and the MGCM) and then the assimilation forces them further north. While components are still as far south as 45N, they tend to be more in the range of 55N to 80N. In some cases, the southern tails to the storms are partly due to the interaction of the tides with the storms (although the tail at 90E is persistent and probably topographically controlled, as is the one around -45E) [*Hollingsworth et al.,* 1996].

Initially the models, both with assimilation and without (Figure 5.13), start with a high zonal wavenumber ($m \approx 6$) storm belt in the southern hemisphere. It is weaker than the northern belt. By sol 17 (Figure 5.15), the waves have slowed down and are generally no longer traveling and seem to weaken further as time goes on. This is possibly due to the changing of the seasons and the fast radiative response of the atmosphere. In the assimilation case (Figure 5.11), this weakening is also accompanied by an increase in the wavenumber (up to 8) not seen in the MGCM. The southern storm belt has been known from previous modeling work [*Zurek et al.,* 1992], as has been the weakness, but it appears from the data assimilation that they are even weaker than predicted by the model–at least during the latter part of the available data. At sol 3.5 (Figure 5.7), they appear to be stronger than in the MGCM, but this might be more of an artifact of the differing polar pressure regimes than the actual strength of the baroclinic waves.

The final interesting feature in the pressure maps are the polar regions. Initially, the assimilation results indicate that the pressure in the northern polar regions are too low (Figures 5.8 and 5.14). Then as time progresses the bias changes and by sol 17, the polar pressure for the assimilation run is lower than the MGCM (Figure 5.16). Finally this trend continues and by sol 24 (Figure 5.18), the difference is well over 1 mbar. Although the change is less dramatic, the southern polar region does the opposite. It starts out low and ends up higher than predicted by the model. The most likely cause of these effects is the condensation of the polar cap. While the assimilation does not directly affect the mass flow to the cap, it can move the atmospheric mass to the polar regions. Unfortunately, the data does not sample the northern polar regions (north of 80N) so the pressure changes are due to the global effects of the assimilation gains. Thus while somewhat questionable, the changes probably reflect the exact seasonal cap formation pattern for the 1997 Martian Fall as opposed to the generic fall implemented in the MGCM.

5.4.2 Zonal Mean Climatology

While the weather is interesting, there is also the climatology or mean behavior of the atmosphere to consider. Initially we hoped to be able to look at the changing of the seasons with the assimilation, but since we only have 25 sols worth of data, we are a bit limited in what is possible. So we will just study the climatology for the period from $L_s \approx 198$ to $L_s \approx 216$. One of the more useful methods of studying the climatology is to look at the zonal mean values (the longitudinal averaged value) [*Zurek et al.,* 1992, *Haberle et al.,* 1993]. In this case we have taken the zonal mean at each pressure level over the period of the data assimilation.

Figures 5.19 through 5.21 are contour maps of the three dynamical fields over the 25 sols of data assimilation. The solid red region at the bottom is the zonal mean surface pressure–a proxy for the topography. Dashed contours represent negative values and several contours in each map have been color coded for ease of reference. The three zonal mean fields shown in the figures are those for the MGCM run without

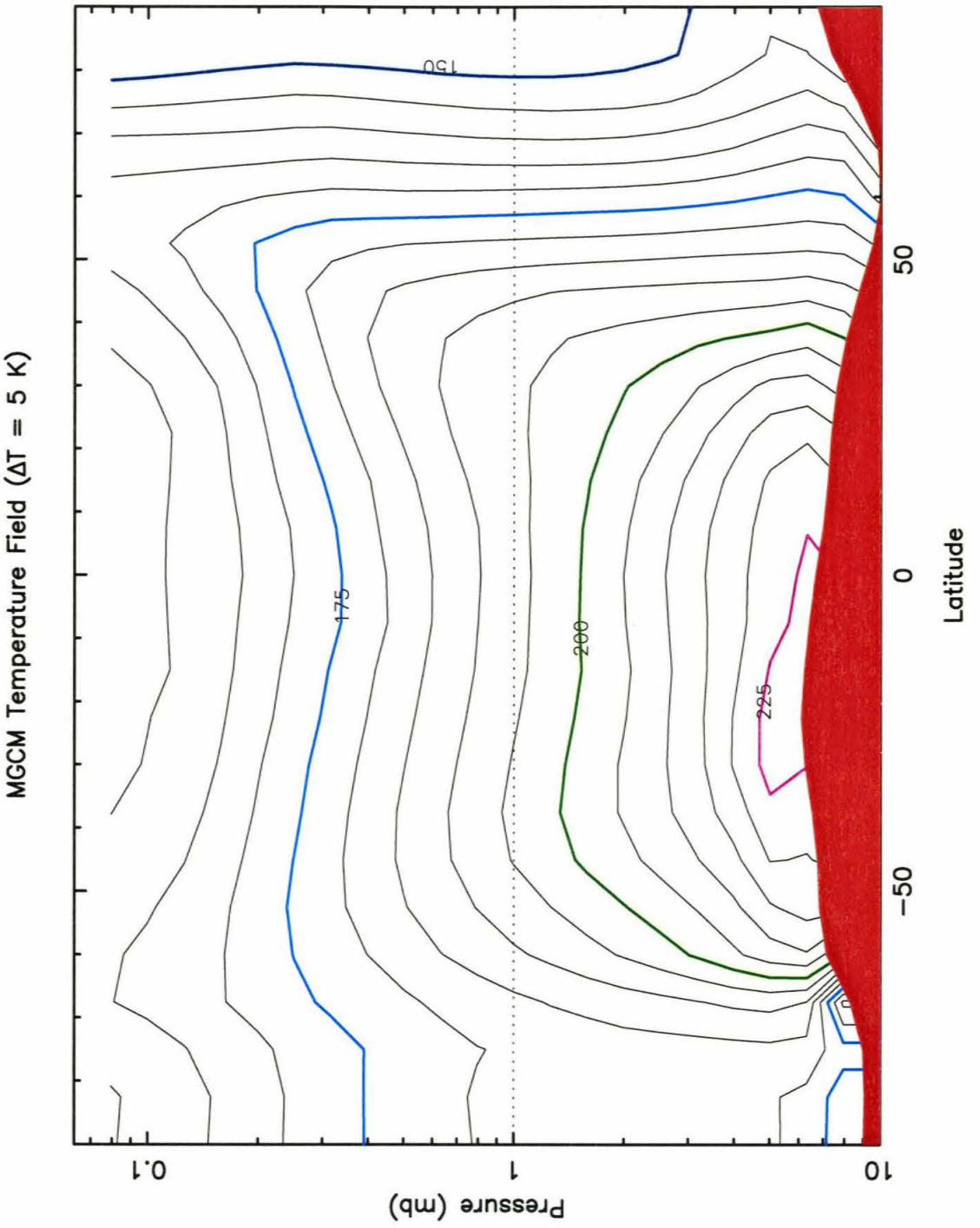

Figure 5.19: Zonal Mean Temperature Field of the MGCM
Contour plot of the zonal mean temperature field averaged over the 25 sols covered by the TES data. The contour interval is 5K. The 150K contour is indicated in blue, the 175K contour in light blue, the 200K contour in green, and the 225K contour in magenta.

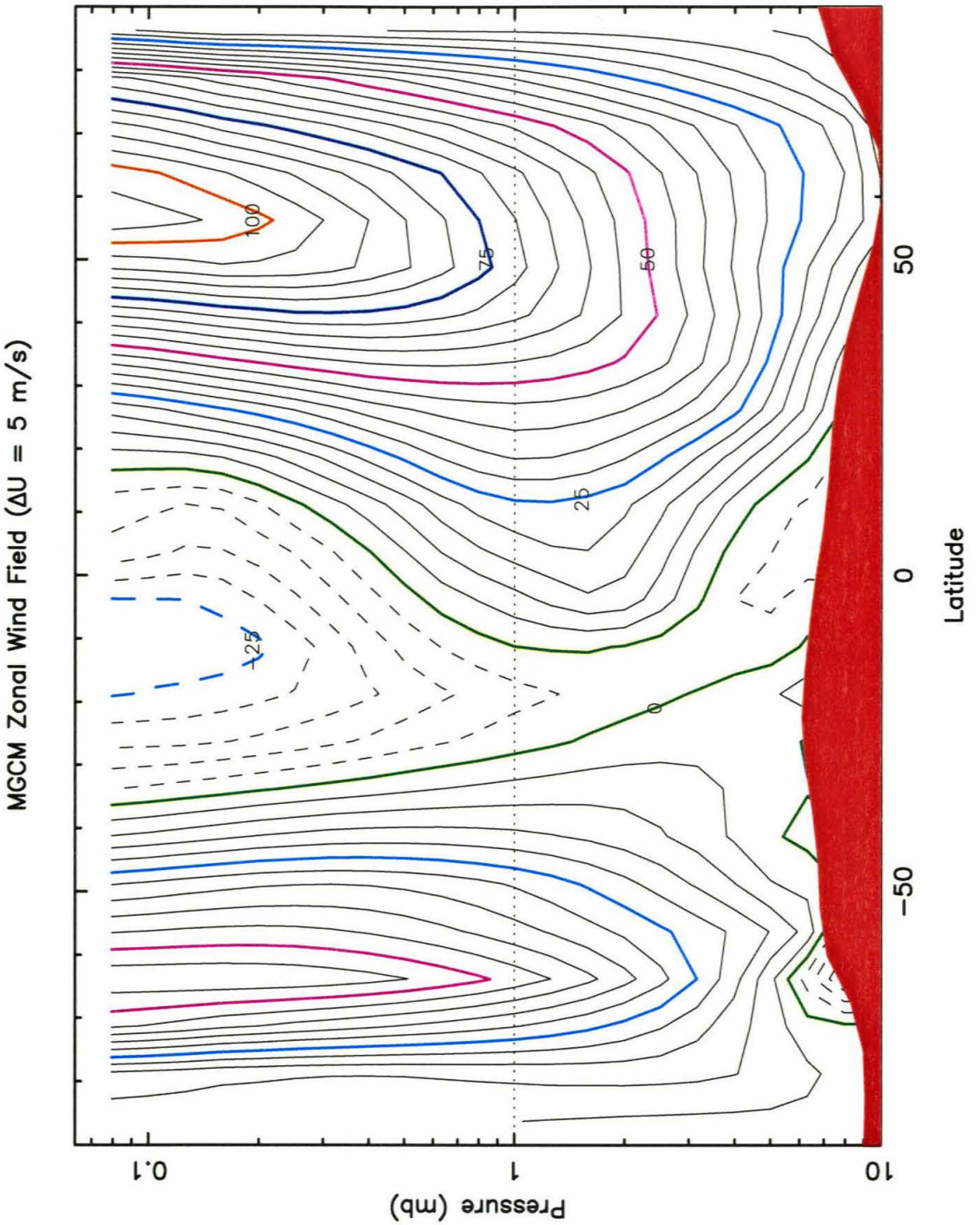

Figure 5.20: Zonal Mean Zonal Velocity of the MGCM
Contour plot of the zonal mean zonal velocity (U) averaged over the 25 sols covered by the TES data. The contour interval is 5 m/s. Negative (or easterly) velocities are dashed contours, and positive (or westerly) velocities are solid contours. The stationary (0 m/s) contour is in green, ±25m/s are both in light blue, 50 m/s is in magenta, 75 m/s is in blue and 100 m/s is in orange.

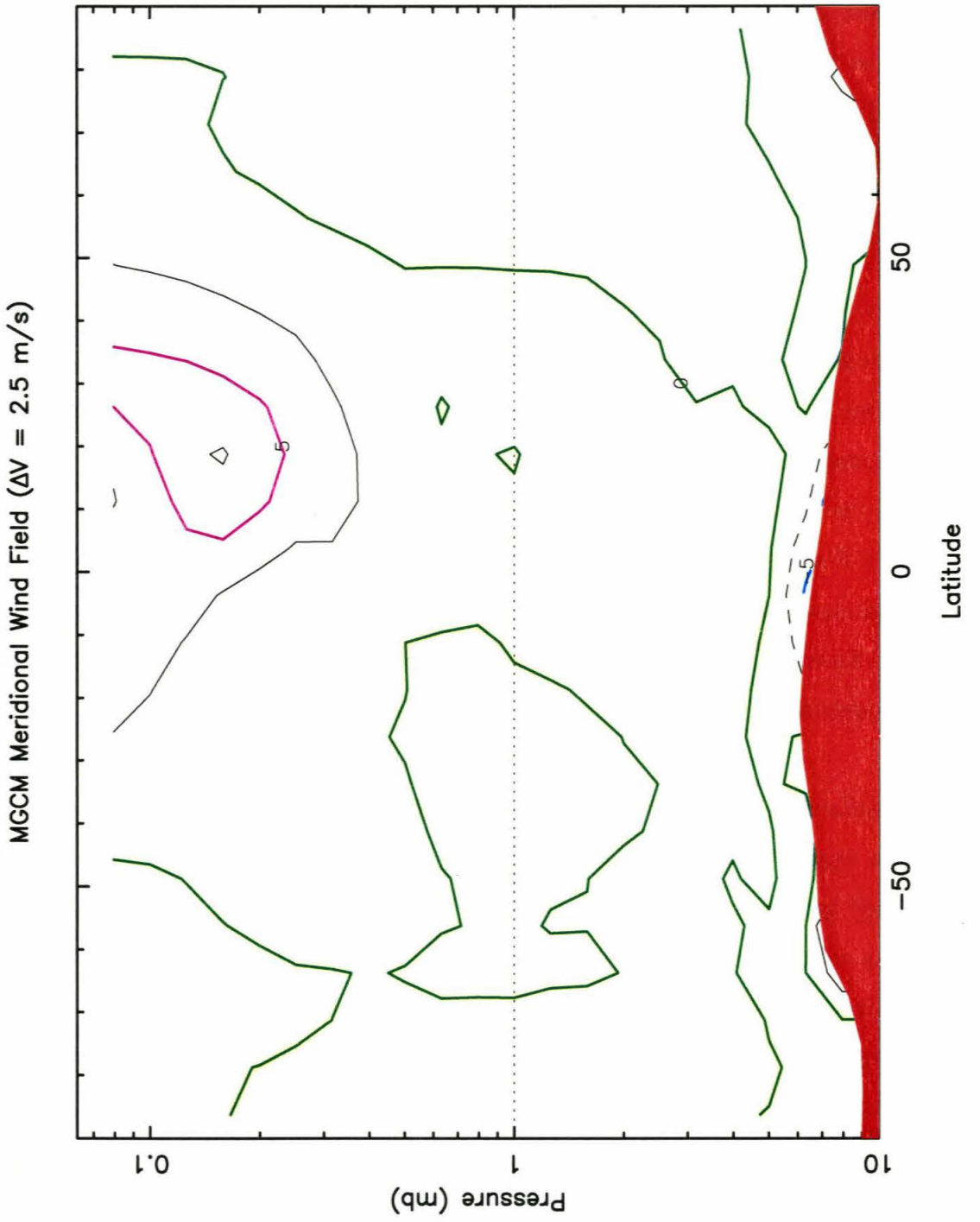

Figure 5.21: Zonal Mean Meridional Velocity of the MGCM
Contour plot of the zonal mean meridional velocity (V) averaged over the 25 sols covered by the TES data. The contour interval is 2.5 m/s. Negative (or southward) velocities are dashed contours; positive (or northward) velocities are solid. The stationary (0 m/s) contour is in green, +5 m/s is in magenta and −5 m/s is in blue.

data assimilation. Given the coarse resolution of the contours, the fields from the data assimilation runs look very similar. Many of the implications of the zonal mean values have already been discussed [*Haberle et al.,* 1993]. The current results are very similar to the low dust Fall equinox cases (since they are only 20 to 30 degrees of L_s away).

In order to understand the differences in the climatology driven by the assimilation, it is easiest to look at the differences between the zonal mean values with and without the assimilation. Figures 5.22 through 5.24 show contour maps of the differences. Note that the contour intervals have changed. Again, the dashed values are negative and the color coded and labeled contours indicate the ranges. The sign of the differences is chosen so that negative contours represent places where the assimilation results are lower than the MGCM and positive contours are places where they are higher. Of course, in the case of negative velocities, higher values are actually less negative and thus slower. The differences between the cases where the TES data is assimilated into two different initial conditions are very similar, and especially when looking at the patterns of the differences between the assimilation results and the MGCM. All the features being discussed here appear in both TES models, and are thus probably actual results of the data assimilation process.

The mean meridional wind field (Figure 5.21) is very small. The meridional winds at different longitudes are much larger but, at this season, they average almost to zero. Furthermore, the difference field is generally very small (as seen by the green 0 contours over most of the map—Figure 5.24). The only apparent feature, appearing in both assimilation runs, is the slight strengthening of the poleward surface wind north of 50N. This is possibly due to the strengthened baroclinic waves generating stronger winds around themselves. It is unclear whether this is actually statistically significant and not just due to noise.

There are several interesting and significant differences between the MGCM temperature field and that from the assimilation process (Figure 5.22). First, there is a consistent pattern above the south pole. Near the surface, there is a slight ($\lesssim$ 0.5K) warming. Above that there is a slight cooling. And finally, at high altitudes, again a slight warming. The net effect of this is to reduce the vertical extent and magnitude

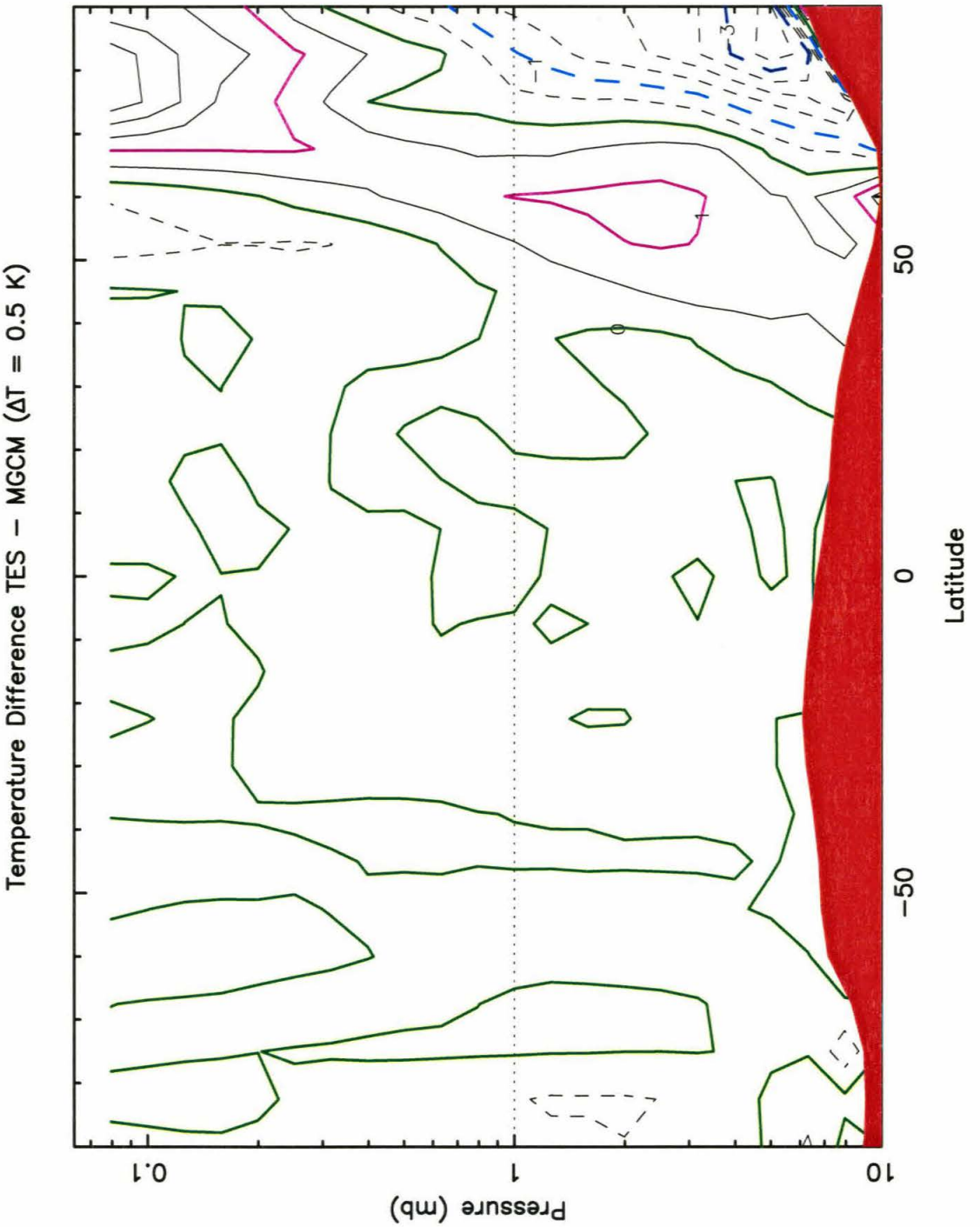

Figure 5.22: Zonal Mean Temperature Difference
The difference between the zonal mean temperature of the assimilated results and those of the MGCM without assimilation. The contour interval is 0.5K. Negative differences (colder assimilation results) are in dashed contours, while warmer assimilation results are in solid contours. The neutral (0K) contour is in green, +1K is in magenta, -1K is in light blue, and -3K is in blue.

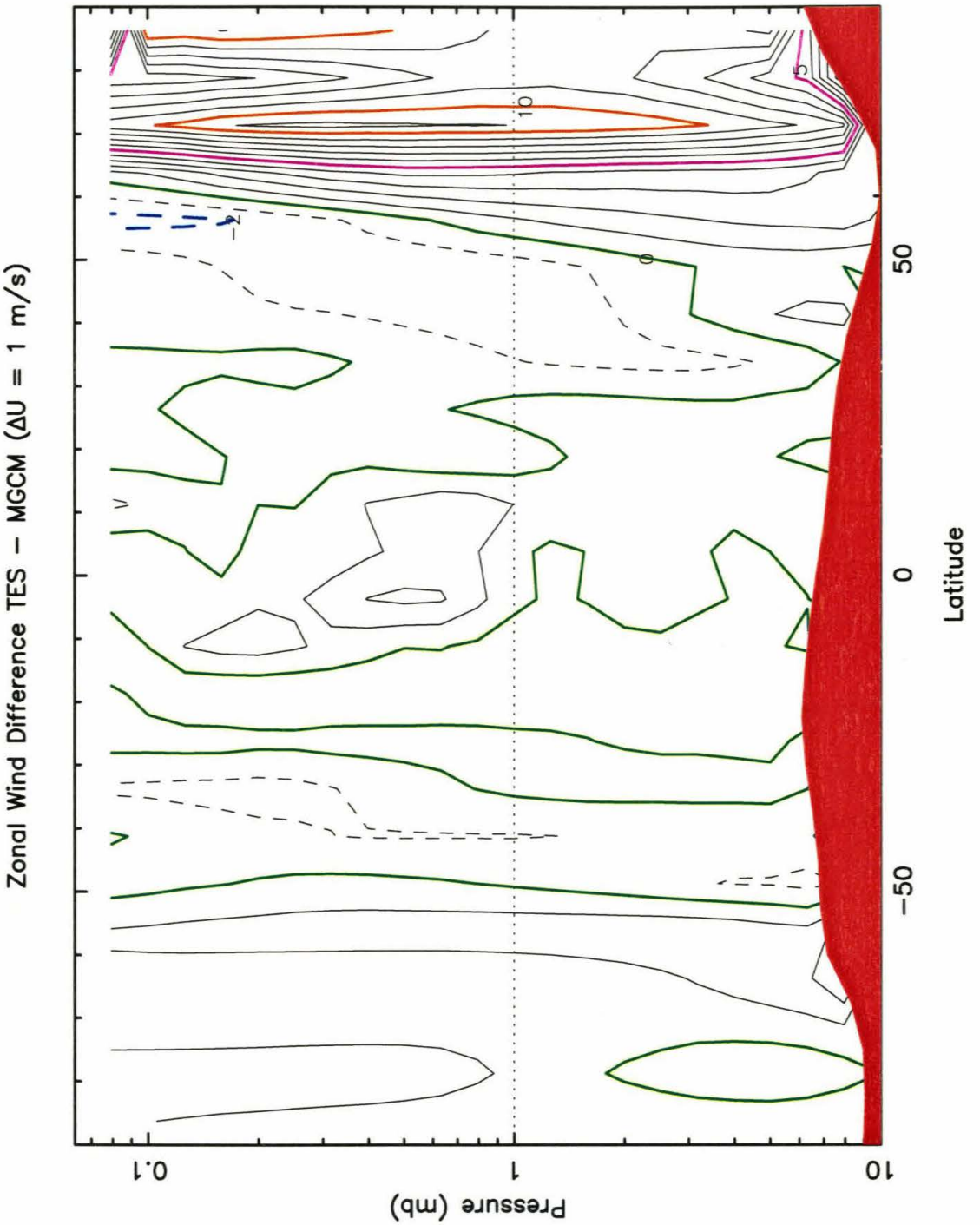

Figure 5.23: Zonal Mean Zonal Wind Difference
The difference between the zonal mean of the zonal winds of the assimilated results and those of the MGCM without assimilation. The contour interval is 1 m/s. Negative differences (more easterly flow) are in dashed contours while positive differences are in solid contours. The neutral (0 m/s) contour is in green, -2 m/s in blue, +5 m/s in magenta, and +10 m/s in orange.

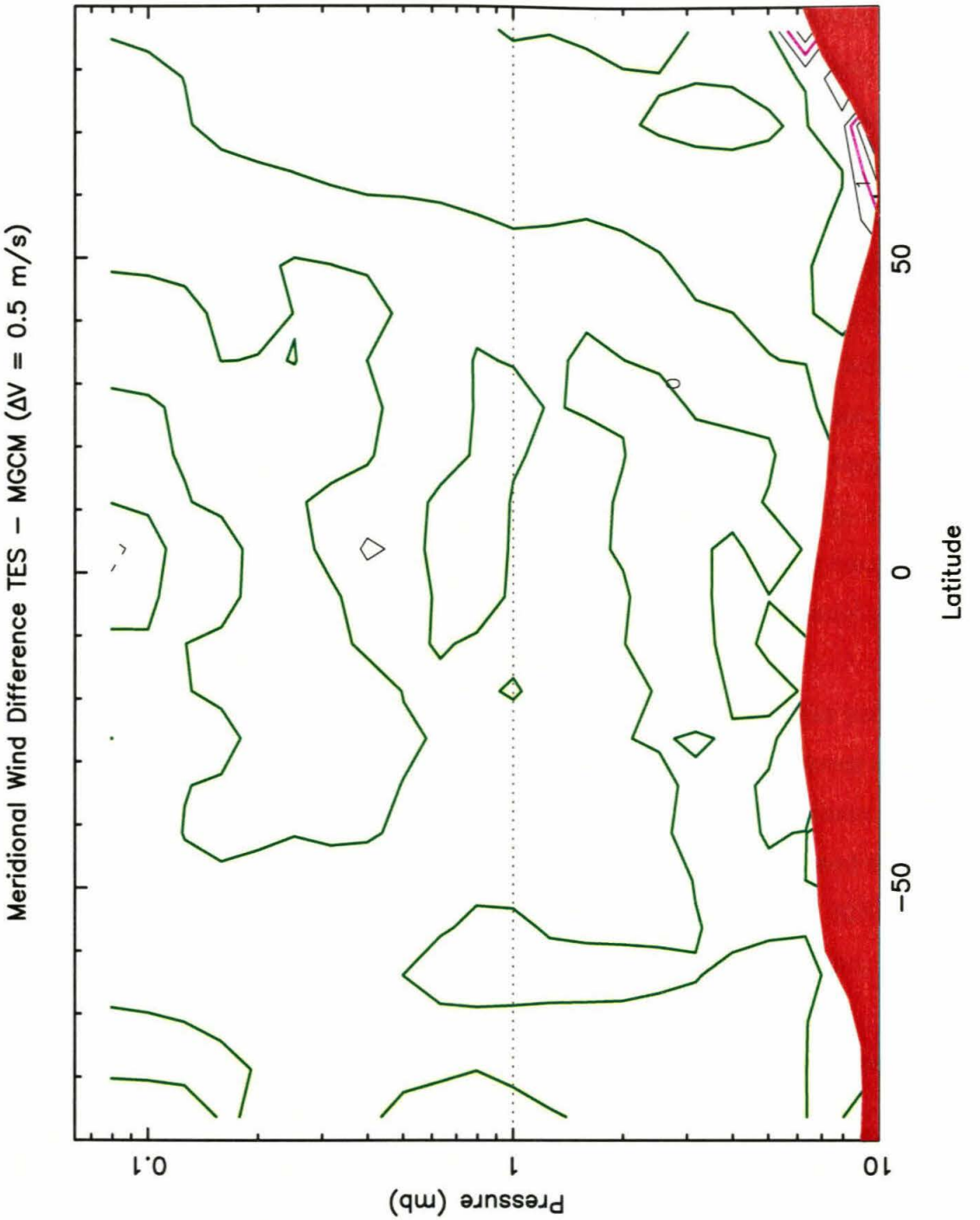

Figure 5.24: Zonal Mean Meridional Wind Difference
The difference between the zonal mean meridional winds from the assimilated results and those of the MGCM without assimilation. The contour interval is 0.5 m/s. Negative differences (more southern flow) are in dashed contours, while positive differences are in solid contours. The neutral (0 m/s) contour is in green and +1 m/s in magenta.

of the 1 mbar polar warming (see Figure 5.19) that occurs poleward of 80S. The temperature inversion still exists, but it is not as pronounced. The other changes are in the northern hemisphere, north of the midlatitude region. At the north pole, the assimilation implies that the real atmosphere is more stable than the MGCM. This can be seen in pattern of differences between the MGCM and assimilation results (Figure 5.22). There is a significant region starting at the surface and extending up to $\sim$ 1 mbar and, at the surface, extending as far south as 65N that the assimilation indicates is colder than the MGCM. In places, the temperatures are almost 4K colder. Above this, the assimilation results indicate that the model needs to be warmed up. While not quite as large, the required warming is above 2K at 0.1 mbar. This results in a temperature minimum below 145K around 1 mbar. This is very close to the CO_2 condensation temperature. Also, with the low altitude cooling, it appears that most of the atmosphere below there is at or near the condensation temperature. This implies that the condensation (being the Fall, the northern seasonal polar cap is growing) is occurring at relatively low altitudes.

Just south of the polar regions (around 60N), there is a reversal of the polar trend. The assimilation results indicate a cooler upper atmosphere and a warmer condition below 1 mbar. While both trends are relatively modest, combining them with the polar changes, they result in almost vertical isotherms between 80N and 60N. The net result is to suppress the near surface temperature inversion between 80N and 60N from the MGCM as well as the cooling above the temperature maximum, resulting in an almost isothermal atmosphere between the surface and 0.1 mbar.

Some of the changes in the zonal (or U component) winds are the largest climatological changes (Figure 5.23). The net effect of all the meaningful changes is to increase the width of the two midlatitude westerly (or positive) jets while decreasing the width of the easterly high altitude jet at the equator. The assimilation also decreases the small easterly surface flow at 60S. The large acceleration of the northern jet shown in the difference map occurs on the poleward shoulder and broadens the jet poleward by 5 to 10 degrees. Since none of the acceleration occurs in the center of the jet, the change only latitudinally widens the central region of the jet and does not

strengthen it. The core of the jet is actually a bit slower in the assimilation results model—the -2 m/s contour above 0.2 mbar at 55N. This broadening of the northern westerly jet is consistent with the increase in $-\partial T/\partial y$ (in Figure 5.22) via the thermal wind equation.

The westerly acceleration at the equator above 1 mbar slightly compresses the easterly jet at that location, causing the equatorial bulge of the Northern hemisphere jet to be more extensive vertically. This is a modest displacement, and, at the most, moves the top of the westerly flow upwards by maybe 0.1 mbar. The compression of the southern hemisphere surface jet is likewise small and is accompanied by a slight broadening of the core of the westerly jet at altitude. At the same time, the northern boundary of the jet is slightly displaced southward, sharpening the horizontal velocity gradient at that location. The poleward side of the westerly jet is, like the northern one, broadened, but to a much lesser degree.

The major effects of the data assimilation on the climatology occur in the northern midlatitude and polar regions. Unfortunately this is where the data is the worst, so while the direction of the effects is probably correct, the magnitude of the effects is uncertain. While the midlatitude effects may be related to be baroclinic waves, many of the processes seem to be associated with the polar regions. There are much more modest effects in the southern hemisphere. Due to the better data sampling, while smaller, they are probably still real.

5.5 Conclusions

The assimilation of TES data during the aerobraking phase using the techniques described in chapter 4 has a significant impact on the results of the MGCM. Despite the short length of the data stream, it successfully changes both the climatology and the weather of the Martian atmosphere. By using a convergence test and by starting with several different initial conditions, it is possible to verify that the assimilation process is driving the model to a distinct state, presumably that of the Martian atmosphere during the 1997 northern hemisphere Fall.

There are three important modifications that the assimilation process imposes on the MGCM. First, it intensifies the baroclinic waves in the northern hemisphere. Even with a dust load corresponding to the observations, the amplitude of the waves is larger than predicted by the MGCM. This may be partly due to the large known inter-annual variability of the waves [*Zurek et al.,* 1992], but the data also implies a more northerly location of the storm track than estimated from Viking observations.

The second modification due to the data is that the midlatitude westerly zonal jets are significantly wider than the ones predicted by the model for this season. This is especially true of the northern jet. The widening is primarily poleward, but there is some compression of the equatorial easterly jet. The broader northern jet may be due to larger mass flow to the growing seasonal cap, but should also allow for the transport of dust and possibly energy further north as well.

The final change implied by the assimilation process is that the temperature above the northern polar cap (up to about 1 mbar) is much closer to the CO_2 condensation temperature than predicted, but is then warmer above that. This implies that the condensation forming the seasonal polar cap is probably occurring in the lower atmosphere, and possibly all the way to the surface.

While the assimilation technique successfully assimilated the TES data, it requires significant amount of tuning to perform optimally. The tuning is partly due to the problematic orbit during this period (at least for data assimilation), but may be partly inherent in the method itself. This is seen in the search for the optimum f value (which may vary even over this short time-scale) as well as in the need to filter the observations. It is important to verify that the assimilation results are actually meaningful and reflective of the actual state of the Martian atmosphere. Fortunately this can be verified, at least to some extent, by assimilating the data using different initial conditions, but the use of a different Mars GCM is needed to truly verify the results. It is also necessary to ignore some parts of the atmosphere due to the poor performance (or inability to overcome the model) in those regions. Despite all these issues, the assimilation of the TES data allows the MGCM to create this Martian Fall and not a random Martian Fall.

Bibliography

Barnes, J. R., Time Spectral Analysis of Midlatitude Disturbances in the Martian Atmosphere, *J. Atmos. Sci., 37,* 2002-2015, 1980.

Barnes, J. R., R. M. Haberle, J. B. Pollack, H. Lee, J. Schaeffer, Mars Atmospheric Dynamics as Simulated by the NASA Ames General Circulation Model. 3. Winter Quasi–Stationary Eddies, *J. Geophys. Res., 101,* 12753-12776, 1996.

Barnes, J. R., J. B. Pollack, R. M. Haberle, C. B. Leovy, R. W. Zurek, H. Lee, J. Schaeffer, Mars Dynamics as Simulated by the NASA Ames General–Circulation Model. 2. Transient Baroclinic Eddies, *J. Geophys. Res., 98,* 3125-3148, 1993.

Collins, M., S. R. Lewis, P. L. Read, F. Hourdin, Baroclinic Wave Transitions in the Martian Atmosphere, *Icarus, 120,* 344-357, 1996.

Collins, M., I. N. James, Regular Baroclinic Transient Waves in a Simplified Global Circulation Model of the Martian Atmosphere, *J. Geophys. Res., 100,* 14421-14432, 1995.

Haberle, R. M., J. B. Pollack, J. R. Barnes, R. W. Zurek, C. B. Leovy, J. R. Murphy, H. Lee, J. Schaeffer, Mars Atmospheric Dynamics as Simulated by the NASA Ames General–Circulation Model. 1. The Zonal–Mean Circulation, *J. Geophys. Res., 98,* 2093-3123, 1993.

Hollingsworth, J. L., R. M. Haberle, J. R. Barnes, A. F. C. Brider, J. B. Pollack, H. Lee, J. Schaeffer, Orographic Control of Storm Zones on Mars, *Nature, 380,* 413-416, 1996.

Kieffer, H. H., B. M. Jakosky, C. W. Snyder, M. S. Matthews (Eds.), *Mars,* 1498 pp., The University of Arizona, Tucson, 1992.

Pollack, J. B., R. M. Haberle, J. Schaeffer, H. Lee, Simulations of the General–Circulation of the Martian Atmosphere. 1. Polar Processes, *J. Geophys. Res., 95,* 1447-1473, 1990.

Smith, M. D., J. C. Pearl, B. J. Conrath, P. R. Christensen, MGS TES Results: Characterization of Atmospheric Dust, *Bull. Amer. Astro. Soc., 30,* 1031, 1998.

Tillman, J. E., R. M. Henry, S. L. Hess, Frontal Systems During Passage of the Martian North Polar Hood over the Viking Lander 2 Site Prior to the First 1977 Dust Storm, *J. Geophys. Res., 84,* 2947-2955, 1979.

Zurek, R. W., J. R. Barnes, R. M. Haberle, J. B. Pollack, J. E. Tillman, C. B. Leovy, Dynamics of the Atmosphere of Mars, in *Mars,* edited by H. H. Kieffer, B. M. Jakosky, C. W. Snyder, and M. S. Matthews, pp. 298-320, The University of Arizona, Tucson, 1992.