ABSTRACT

Title of Document: MARS WEATHER AND PREDICTABILITY: MODELING AND ENSEMBLE DATA ASSIMILATION OF SPACECRAFT OBSERVATIONS

Steven J. Greybush, Ph.D., 2011

Directed By: Professor Eugenia Kalnay, Department of Atmospheric and Oceanic Science, The University of Maryland, College Park

Combining the perspectives of spacecraft observations and the GFDL Mars General Circulation Model (MGCM) in the framework of ensemble data assimilation leads to an improved understanding of the weather and climate of Mars and its atmospheric predictability.

The bred vector (BV) technique elucidates regions and seasons of instability in the MGCM, and a kinetic energy budget reveals their physical origins. Instabilities prominent in the late autumn through early spring seasons of each hemisphere along the polar temperature front result from baroclinic conversions from BV potential to BV kinetic energy, whereas barotropic conversions dominate along the westerly jets aloft. Low level tropics and the northern hemisphere summer are relatively stable. The bred vectors are linked to forecast ensemble spread in data assimilation and help explain the growth of forecast errors.

Thermal Emission Spectrometer (TES) temperature profiles are assimilated into the MGCM using the Local Ensemble Transform Kalman Filter (LETKF) for a 30-sol evaluation period during the northern hemisphere autumn. Short term (0.25 sol) forecasts compared to independent observations show reduced error (3–4 K global RMSE) and bias compared to a free running model. Several enhanced techniques result in further performance gains. Spatially-varying adaptive inflation and varying the dust distribution among ensemble members improve estimates of analysis uncertainty through the ensemble spread, and empirical bias correction using time mean analysis increments help account for model biases. With bias correction, we estimate a predictability horizon of about 5 sols during which temperature, wind, and surface pressure forecasts initialized from an assimilation analysis are superior to a free running model forecast.

LETKF analyses, when compared with the UK reanalysis, show a superior correspondence to independent radio science temperature profiles. Traveling waves in both hemispheres share a correspondence in phase, and temperature differences between the analyses are generally less than 5 K. Assimilation of Mars Climate Sounder (MCS) temperature profiles reveals the importance of vertical distributions of dust and water ice aerosol in reducing model bias. A strategy for assimilation of TES and MCS aerosol products is outlined for future work.

# MARS WEATHER AND PREDICTABILITY: MODELING AND ENSEMBLE DATA ASSIMILATION OF SPACECRAFT OBSERVATIONS

By

Steven J. Greybush

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Eugenia Kalnay, Chair
Professor Kayo Ide
Professor Takemasa Miyoshi
Professor Ross J. Salawitch
Dr. Ross N. Hoffman
Professor Brian R. Hunt, Dean's Representative

# Acknowledgements

I wish to express my gratitude to the following individuals:

My advisor, Eugenia Kalnay, for her inspiring ideas, direction, and mentoring;

Committee Members Kayo Ide, Takemasa Miyoshi, Ross Hoffman, Brian Hunt, and Ross Salawitch, for their great suggestions and guidance;

The "Mars Team," especially Matthew Hoffman, Janusz Eluszkiewicz, and John Wilson, for their patient and helpful advice;

The Mars atmosphere community for their insightful comments, in particular Tim McConnochie; Dave Hinson; Francois Forget; Istvan Szunyogh; Ming Cai; Armin Kleinboehl, Nick Heavens, David Kass, Tim Schofield, and the MCS team; and Peter Read, Luca Montabone, Stephen Lewis and the UK team;

The UMD weather chaos group, the EnKF community, Matus Martini, Steve Penny, Cathy Sabol, Javier Amezcua, Anthony Santorelli, Edward Nowottnick, Seung-Jae Lee, and Ji-Sun Kang for many stimulating discussions;

David Kuhl and Gyorgyi Gyarmati for their support in managing our computing cluster, and the AOSC staff for their helpful assistance;

And finally the UMD faculty and graduate students, friends, past teachers and mentors, my brother Nicholas, and especially my parents James and Leona Greybush for a supportive environment that encouraged intellectual curiosity, academic achievement, and love of learning.

This research was supported by NASA grant NNX07AM97G.

# Table of Contents

# List of Tables

# List of Figures

km (Level 7) u-wind speed. Middle: zonal mean u-wind speed. Bottom: zonal mean temperature.

**Figure 2.1:** Bred vector amplitude time evolution. SH amplitudes are in blue, NH in red. Raw BV amplitudes from year 1 are in light shading. 10-day smoothed amplitudes from year 1 are in dark shading. The six seasons are identified by background colors that match the colors in Fig. 1.3, and begin with NH early autumn ($L_s$ 120).

**Figure 2.2:** Bred vector daily growth rate seasonal evolution for two Martian years, illustrating interannual variability. SH amplitudes are in blue, NH in red. 30-day smoothed amplitudes BV amplitudes from year 1 are in light shading; smoothed amplitudes from year 2 are in dark shading.

**Figure 2.3:** Latitude-pseudopressure plots of bred vector structure as a temporal average for each Martian "season." Shading represents zonal mean BV amplitude; contours denote zonal mean temperatures. Plots are arranged so that a comparison of symmetry between NH and SH can be analyzed.

**Figure 2.4:** 2-day time average plots following the NH Autumnal Equinox ($L_s$ 180°). Top Left: temperature transient eddy (contoured) from smoothed climatology and temperature BV (shaded) at approximately 3.5 km above the surface (Level 20). Top right: similar to top left, except at Level 25 (about 0.5 km above the surface). Bottom left: vertical cross section through 60° S latitude with temperature BV (shaded) and control (contours). Bottom right: vertical cross section through 60° N latitude with temperature BV (shaded) and control (contours).

**Figure 2.5:** Time series over one Martian year of bred vector (rows 1 and 3) and transient eddy (rows 2 and 4) temperature squared amplitude ($K^2$; vertical axis) and wave number (color of dots) for each hemisphere at Level 25 (near surface). Seasons are indicated by color panels. Note that the BV instability generally precedes the initiation of seasonal eddy activity.

**Figure 2.6:** Histograms of NH (left) and SH (right) BV zonal wavenumber (dark blue, wave 1; cyan, wave 2; yellow, wave 3; red, wave 4) at Level 25 (corresponds to Figure 2.5) for various BV amplitude bins (< 1 $K^2$, 1–2 $K^2$, 2–3 $K^2$, 3–4 $K^2$, and > 4 $K^2$; top row), and for times of rapid growth (by 1 $K^2$ in 0.25 sol; bottom row, left bars) and rapid decay (by 1 $K^2$ in 0.25 sol; bottom row, right bars) of instabilities.

**Figure 2.7:** Bred Vector energy diagram, based on Lorenz (1967), illustrating the transformation of energy among control potential energy ($P_c$), bred vector potential energy ($P_b$), control kinetic energy ($K_c$), and bred vector kinetic energy ($K_b$). $Q_c$ and $Q_b$ are the heating terms, and $D_c$ and $D_b$ the frictional dissipation terms for the control and bred vector, respectively. A budget analysis is performed for $K_b$ (circled), and the barotropic (BT; green) and baroclinic (BC; blue) conversion terms have been computed.

**Figure 2.8:** Zonal mean BV quantities, plotted as a 10-sol mean starting at the NH autumnal equinox ($L_s$ 180°), Hour 00. Upper left: temperature BV (shaded) with control temperature (contours). Upper right: BV kinetic energy (shaded) with control *u*-component wind (contours). Lower left: baroclinic term of the BV KE tendency equation (shaded) with control u-component wind (contours). Lower right: barotropic term of the BV KE tendency equation (shaded) with control *u*-component wind (contours).

**Figure 2.9:** Same quantities as Figure 2.8, but for a cross section through longitude 345°, plotted as a 2-sol mean starting at the NH autumnal equinox ($L_s$ 180°), Hour 00. Top left panel is time mean signed bred vector rather than magnitude.

**Figure 2.10:** NH late autumn ($L_s$ 180°–240°) time averaged temperature BV amplitude ($K^2$, shaded, top) and BV KE baroclinic term ($m^2\ s^{-3}$, shaded, bottom) near 3 km altitude (Level 20), plotted with surface topography (positive: solid contours, negative: dashed contours).

**Figure 2.11:** Comparison of bred vectors (left panels) with ensemble spread from data assimilation (right panels) just after the NH autumnal equinox ($L_s$ 180°) of MY 24. The top row shows temperature bred vector [K] using assimilation analyses as a control run (left) and temperature analysis ensemble spread [K] (right) at model Level 20, or approximately 3 km altitude, at a snapshot in time ($L_s$ 186.7°). The bottom row shows zonal mean temperature BV amplitude (left) and ensemble spread (right) as a time mean over 30 sol ($L_s$ 180.5°–198.4°). Contours are analysis temperature [K].

**Figure 3.1:** Left panel: vertical coverage of TES and MCS retrievals relative to MGCM vertical levels. Right panel: sample horizontal coverage of TES observations in a 0.25 sol window. Shading represents the passage of time (successive orbits).

**Figure 3.2:** Schematic illustrating the components of ensemble data assimilation for Mars: the Mars Global Circulation Model (MGCM), Thermal Emission Spectrometer (TES) observations, and the Local Ensemble Transform Kalman Filter (LETKF) assimilation system.

**Figure 3.3:** Schematic illustrating the 4d-LETKF, where observations (binned together as red line segments for the first 6 hour assimilation cycle) are compared to forecasts (solid blue line) at the correct hour (3–9) rather than just to the 6 hour forecast as in 3d-LETKF.

**Figure 3.4:** Performance of the MGCM-LETKF, evaluated by comparing RMSE differences of 0.25-sol forecasts from assimilation analyses, as well as a free run forecast, with TES observations. RMSE values in the legend are from sols 10–20 of assimilation (after spin-up).

**Figure 3.5:** Impact of improving data assimilation methods on zonal mean temperature bias (left, colors) and error standard deviation (right, colors) of 0.25-sol forecasts from data assimilation compared to TES observations for sols 10–20 of assimilation. Contours are forecast ensemble mean temperature. Row 1: free run

forecast (no assimilation, just model climatology). Row 2: baseline data assimilation configuration, with fixed inflation and dust, no bias correction. Row 3: adaptive inflation and varying dust among the ensemble members, no bias correction. Row 4: Same as row 3, but with empirical bias correction.

**Figure 3.6:** Improvements to the ensemble spread through adaptive inflation and ensemble dust distributions. Top four panels are time averaged temperature [K] ensemble mean (contours) and spread (shaded). Upper left: fixed dust, fixed inflation. Upper right: varying dust, fixed inflation. Middle left: fixed dust, adaptive inflation. Middle right: varying dust, adaptive inflation. Lower row: covariance inflation (%) estimated adaptively for fixed dust (left) and varying dust (right) after 30 sols assimilation.

**Figure 3.7:** Performance sensitivity of the MGCM-LETKF to localization length scale, evaluated by comparing RMSE differences of 0.25-sol forecasts from assimilation analyses with TES observations.

**Figure 3.8:** Global RMSE of temperature forecasts compared to TES observations. The black curve is from a freely running model with no data assimilation. The blue (without bias correction) and green (with empirical bias correction) curves are from 0.25-sol forecasts based on data assimilation analyses. The red and gold curves are 10-sol forecasts that are initialized from a data assimilation analysis at the onset, but have no knowledge of observations afterwards. The red curve is without bias correction, and the gold curve applies a bias correction field, calculated from the mean analysis increment over the previous 10 days of assimilation, every 0.25 sols.

**Figure 3.9:** 5-day forecast RMSE as compared to the most recent analysis from data assimilation for surface pressure (left column), zonal mean temperature (middle column), and zonal mean u-wind (right column). Colors are RMSE, whereas contours are time mean analysis quantity. The top row are forecasts without empirical bias correction, the middle row are forecasts with empirical bias correction, and the last row are a free run forecast without any observation data.

**Figure 4.1:** Schematic comparing vertical resolution of the Oxford-LMD Mars model versus the GFDL MGCM. The Oxford-LMD model is purely sigma coordinate, whereas the MGCM transitions to pressure coordinates at higher altitudes.

**Figure 4.2:** Preliminary comparison of temperature fields in the lowest model level. This comparison takes place at a snapshot in time, on Mars Year 24, Mars Day 565, Mars Hour 00, or $L_s$ 182.3°. Upper left: UK Reanalysis temperature field on its own coordinates. Upper right: UK Reanalysis temperature field interpolated horizontally to the MGCM grid. Lower left: MGCM-LETKF temperature field on its own coordinates. Lower Right: UK Reanalysis column dust opacity, normalized to a surface pressure of 700 Pa.

**Figure 4.3:** Zonal mean temperature bias (UK – LETKF) and RMS difference between the UK reanalysis and the LETKF reanalysis.

**Figure 4.4:** Hovmoller diagram of traveling waves (temperature perturbations at ~3 km altitude) at 60°S (left panels) and 60°N (right panels) in the UK (top panels) and LETKF (bottom panels) reanalyses. Gray lines are the same for both reanalyses, and represent a subjective depiction of the peaks of the traveling waves by the author.

**Figure 4.5:** Location of the Radio Science profiles over the entire MGS mission (top panel), and those used to evaluate the reanalyses during MY 25 NH autumn (bottom panel). The profiles are generally located in the NH sub-arctic, with the latitude shifting gradually southward over the course of 20 sols (colors are $L_s$ value).

**Figure 4.6**: Left panel: Spaghetti plot of temperature profiles from radio science (black), LETKF reanalysis (red), and UK reanalysis (blue). Right: bias and RMSE of the LETKF and UK reanalyses, taking the radio science data as truth. The uncertainties of the RS profiles are denoted by the green curve. Column mean values are 4.56 K RMSE and 2.03 K bias for the LETKF and 5.79 K RMSE and 2.36 K bias for UK.

**Figure 4.7:** Sample Mars Climate Sounder (MCS) climatology for Days 540–544 ($L_s$ 170°–172°) of Mars Year 29. Upper left: dust aerosol concentration [$10^{-3}$ $km^{-1}$]. Upper right: water ice concentration [$10^{-3}$ $km^{-1}$]. Lower left: temperature [K]. Lower right: Uncertainty values for each observation type (temperature in K, dust opacity in $10^{-4}$ $km^{-1}$, water ice opacity $10^{-5}$ $km^{-1}$).

**Figure 4.8:** MGCM-LETKF assimilation of MCS version 3 profiles using adaptive inflation, shown as an average over sols 10–20. Left: Bias (MCS minus forecast). Right: error standard deviation. Contours: analysis ensemble mean. Top row: no bias correction; middle row: with empirical bias correction; bottom row: freely running model.

**Figure 4.9:** Bias (observation minus forecast) of MGCM-LETKF 0.25 sol forecasts from assimilation compared with MCS temperature observations (contours and experimental period same as Figure 4.8) for various MGCM configurations. (A) Control assimilation from Figure 4.8 with spatially fixed dust opacity (0.2–0.5), original dust height, and original MGCM physics. (B) Same as (a), but with new MGCM physics (see text). (C) Same as (b), but with dust opacities determined based on TES climatology and dust height determined from model tracers. (D) Same as (c), but with radiatively active water ice clouds.

**Figure 4.10:** Seasonal evolution of the $CO_2$ polar ice caps (shaded) and TES dust opacities (contours) through a MGCM simulated Mars Year 24. Each successive panel (latitude vs. longitude) represents the passage of 60 sols. Shading intervals represent accumulated $CO_2$; units approximately correspond to depth, in cm. Contours represent column dust opacity normalized to a surface pressure of 6.1 hPa; purple and pink contours correspond to high opacities.

# Chapter 1: Introduction to the Weather and Climate of Mars

## *1.1 Fundamental Differences between Earth and Mars*

The study of planetary atmospheres shares a synergistic relationship with the study of meteorology on Earth. Several features of Martian weather would be familiar to the terrestrial meteorologist: a Hadley circulation (albeit extending to polar latitudes), diurnal temperature variations (nearly 100 K), traveling weather systems, and even thin water ice clouds. The Martian sol (24 hours, 39 minutes) is similar in length to an Earth day, and the axial tilt of Mars is 25 degrees (Table 1.1). A year lasts 687 Earth days (668 sols). The Martian orbit is more elliptic than the Earth's, with a 44% variation in solar insolation between aphelion and perihelion. Mars has a similar Rossby radius of deformation, but about half the planetary radius of the Earth,

**Table 1.1: Fundamental differences, including orbital parameters and atmospheric constituents, between Mars and the Earth.**

| Variable | Mars | Earth |
|---|---|---|
| **Radius** | **3396 km** | **6378 km** |
| **Gravity** | **3.72m $s^{-2}$** | **9.81m $s^{-2}$** |
| **Solar Day** | **24 hours 39 minutes (1 sol)** | **24 hours** |
| **Year** | **686.98 Earth days** | **365.24 Earth days** |
| **Obliquity (Axial Tilt)** | **25 °** | **23.5 °** |
| **Primary Atmospheric Constituents** | **Carbon Dioxide (95%), Nitrogen, Argon** | **Nitrogen (78%), Oxygen (21%), Argon** |
| **Surface Pressure** | **600 Pa** | **101,300 Pa** |
| **Deformation Radius** | **920 km** | **1100 km** |
| **Surface Temperature** | **140–300 K** | **230–315 K** |

meaning that a smaller number of waves are expected in a given latitude circle. Other features would seem alien to the earthbound observer. The primary atmospheric constituent is carbon dioxide (95%), with surface pressures of only $10^2$–$10^3$ Pa. During winter a significant fraction of the atmosphere (~30 %) deposits on the poles as $CO_2$ ice. Powerful local and regional dust storms are common; during some Martian years they grow to envelop the entire planet (Figure 1.1) and persist for dozens of days.

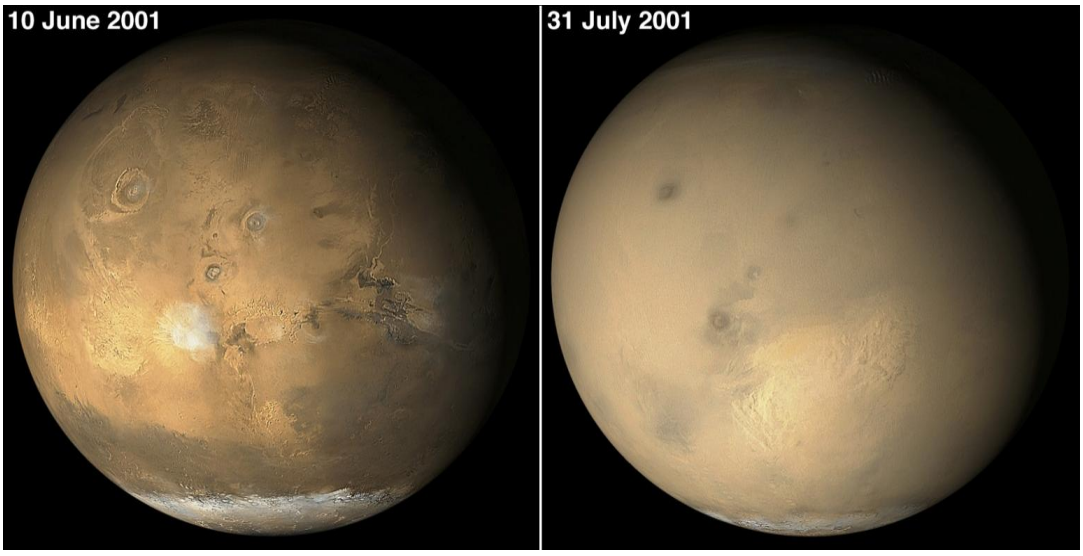


**Figure 1.1: Mars before (left) and during (right) the 2001 global dust storm. Visible image courtesy of NASA / JPL / Malin Space Science Systems.**

The topography of Mars is more extreme than that of the Earth (Figure 1.2). There is a nearly 5 km dichotomy in elevation between the northern hemisphere (NH) and southern hemisphere (SH). The northern latitudes are dominated by low flat plains, Vastitas Borealis, that some have hypothesized may have been an ocean early in Mars' existence (Baker et al., 1991). The SH is high in elevation, with the exception of the deep Hellas and Argyre impact basins, and is heavily cratered. In the tropics is the Tharsis plateau, with the extinct volcano Olympus Mons topping 26 km

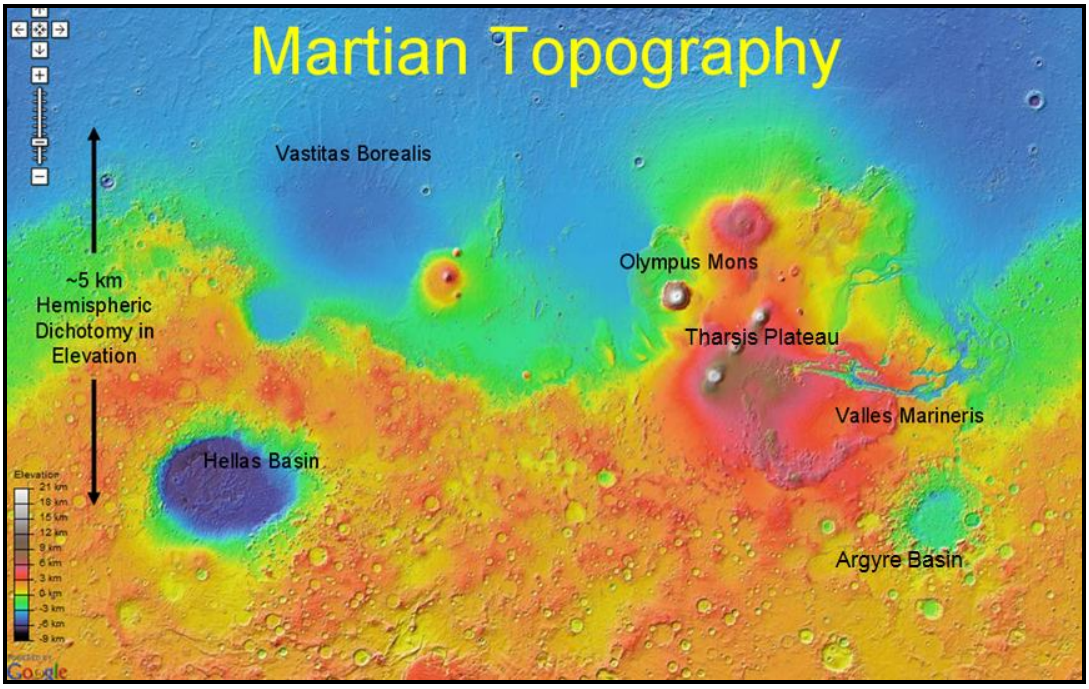


**Figure 1.2: The topography of Mars (courtesy of Google Mars), with major regions labeled. High resolution topography was obtained from the Mars Orbital Laser Altimeter (MOLA).**

in elevation, and the Valles Marineris canyon system extending 4000 km with a depth of up to 7 km.

The atmosphere of Mars provides an excellent laboratory to test our understanding of meteorological theories. Fundamental principles of geophysical fluid dynamics still apply, but with different parameters and geophysical processes being important. The Martian atmosphere also serves as a testbed for numerical weather prediction models and data assimilation systems built for the terrestrial atmosphere; techniques and insights developed here can be reapplied to the Earth. Finally, an accurate understanding of the temperature and density structure of the upper atmosphere of Mars is essential to support aerobraking and aerocapture of future robotic space missions.

In the remainder of this chapter, I explain conventions for keeping time on Mars (Section 1.2), describe the contributions of spacecraft to explaining features of Mars weather (Section 1.3), introduce the GFDL Mars Global Circulation model (MGCM) (Section 1.4) and use it to demonstrate general features of the circulation (Section 1.5), and list the scientific questions to be addressed by this dissertation (Section 1.6).

## *1.2 Keeping Time on Mars*

The Martian sol (24 hours, 39 minutes, 35 seconds) is slightly longer than an Earth day. When I refer to (Mars) days in this paper, we mean a Martian sol, and (Mars) hours are 1/24 of a sol. A Martian year lasts 686.98 Earth days, or 668.5991 sols. A commonly used seasonal time index is the solar (areocentric) longitude $L_s$ which is 0° at the northern spring equinox, 90° at the northern summer solstice, 180° at the northern autumnal equinox, and 270° at the northern winter solstice.

Figure 1.3 shows a schematic of the Martian orbit. Mars is at its perihelion (closest to the sun) at $L_s$ 250.87°. Due to the eccentricity of the orbit, the number of sols between $L_s$ 0° and $L_s$ 90° (northern spring equinox and northern summer solstice) is greater than from $L_s$ 180° to 270° by 51 sols. Along with a 44% variation in solar radiation between perihelion and aphelion, northern winter is less severe than the long southern winter.

We define six seasons on Mars, each 60° of $L_s$ in length. Starting from $L_s$ 0°, we term them NH "late spring," "summer solstice," "early autumn," "late autumn," "winter solstice," and "early spring." With no large thermal reservoir analogous to the Earth's oceans, there is less of a seasonal lag between extremes in radiation and

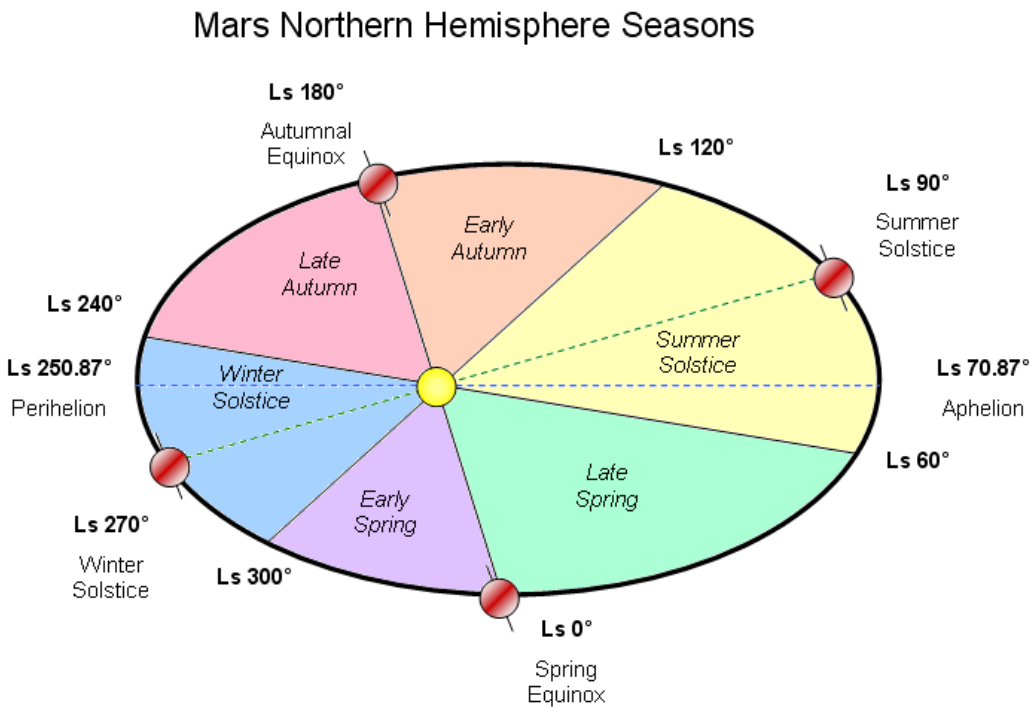


**Figure 1.3: Schematic of the orbit of Mars and Northern Hemisphere seasons, as referred to in this paper. Blue lines correspond to the perihelion and aphelion (closest and furthest distances to the sun). Seasons are determined by solar longitude ($L_s$), which progresses from 0° to 360° over the course of a Mars year (668 sols). Figure is not to scale.**

maxima in annual temperatures. Therefore, it makes sense to have seasons symmetric about the solstices, rather than the four seasons on Earth that match the lagged terrestrial annual cycle of temperatures. Our analysis of instabilities and traveling waves in Chapter 2 further informs our choice of seasons.

For designing a system that incorporates measurements by Earth-operated spacecraft of a Martian climate, time keeping and conversions are very important. Convention in the Mars community labels Mars years consecutively, starting year 0 on May 24, 1953 which is prior to all spacecraft observation. For example, Mars Year (MY) 24 $L_s$ 0° corresponds to July 14, 1998. Table 1.2 indicates conversions between the equinoxes of each recent Mars year and the corresponding Earth dates,

**Table 1.2: Selected Date Conversions from Mars to Earth.**

| Earth Date | Mars Year, $L_s$ | Model Year, Day, Hour |
|---|---|---|
| 1996-08-26 20:07:09 | MY 23 $L_s$ 0.00° | Year 0, Day 183, Hour 16.12 |
| 1997-09-12 21:25:16 | MY 23 $L_s$ 180.00° | Year 0, Day 555, Hour 12.10 |
| 1998-07-14 19:06:12 | MY 24 $L_s$ 0.00° | Year 1, Day 184, Hour 5.99 |
| 1999-07-31 20:21:13 | MY 24 $L_s$ 180.00° | Year 1, Day 556, Hour 1.91 |
| 2000-05-31 18:34:04 | MY 25 $L_s$ 0.00° | Year 2, Day 184, Hour 20.32 |
| 2001-06-17 19:47:07 | MY 25 $L_s$ 180.00° | Year 2, Day 556, Hour 16.21 |
| 2002-04-18 17:41:14 | MY 26 $L_s$ 0.00° | Year 3, Day 185, Hour 10.32 |
| 2003-05-05 20:21:47 | MY 26 $L_s$ 180.00° | Year 3, Day 557, Hour 7.62 |
| 2004-03-05 16:24:12 | MY 27 $L_s$ 0.00° | Year 4, Day 185, Hour 23.92 |
| 2005-03-22 19:50:11 | MY 27 $L_s$ 180.00° | Year 4, Day 557, Hour 21.97 |
| 2006-01-21 16:19:05 | MY 28 $L_s$ 0.00° | Year 5, Day 186, Hour 14.69 |
| 2007-02-07 18:29:28 | MY 28 $L_s$ 180.00° | Year 5, Day 558, Hour 11.51 |
| 2007-12-09 16:26:11 | MY 29 $L_s$ 0.00° | Year 6, Day 187, Hour 5.65 |
| 2008-12-25 18:01:31 | MY 29 $L_s$ 180.00° | Year 6, Day 559, Hour 1.90 |
| 2009-10-26 15:22:02 | MY 30 $L_s$ 0.00° | Year 7, Day 187, Hour 19.47 |
| 2010-11-12 17:15:44 | MY 30 $L_s$ 180.00° | Year 7, Day 559, Hour 16.01 |
| 2011-09-13 14:27:29 | MY 31 $L_s$ 0.00° | Year 8, Day 188, Hour 9.43 |
| 2012-09-29 16:07:34 | MY 31 $L_s$ 180.00° | Year 8, Day 560, Hour 5.76 |
| 2013-07-31 13:50:13 | MY 32 $L_s$ 0.00° | Year 9, Day 188, Hour 23.68 |

using the Mars24 algorithm described by Allison (1997) and Allison and McEwen (2000).

The GFDL Mars Global Circulation Model (MGCM), described in Section 1.4, uses its own convention based on model days and hours. As there is no standard solution in the community for dealing with Martian leap years, and there are design constraints on the length of model parameters, the MGCM assumes a year of exactly 668 days. Model year 0, day 0, hour 0 (midnight on the Martian prime meridian of 0° longitude) corresponds to the 1996 perihelion (February 20, 1996 at 02:48:29 UTC), and is prior to any spacecraft observations that will be used in our project. Time

conversions are always made first between Mars Year and $L_s$ and Earth date, and then to model year/day/hour, ensuring that spacecraft observations are always used at the correct time.

### *1.3 Spacecraft Exploration of Mars*

In addition to the theoretical insights that arise from consideration of the fundamentals of geophysical fluid dynamics, much of our insights into Mars weather and climate are the result of spacecraft exploration of the Red Planet (Figure 1.4).

In 1971, the Mariner 9 probe became the first spacecraft to successfully orbit Mars. However, the probe found a global dust storm in progress (Anderson and Leovy, 1978), which obscured much of the surface; after the storm abated, the spacecraft was able to continue its objective to return photographs of the planet's surface to Earth.

The two Viking lander missions of the late 1970s (landed 22.697° N, 48.222° W and 48.269° N, 225.990° W) provided a wealth of insights to Mars weather, especially the long duration (1976–1982, or more than two Mars year) surface pressure, temperature, and wind observations (Hess et al., 1977), which are unrivaled to this day.

In 1997, the Mars Global Surveyor spacecraft began orbiting Mars on a multi-year exploration, which lasted until communications were lost in 2006. Its most prominent contribution to atmospheric science was measurements from the Thermal Emission Spectrometer (TES), whose nadir radiances enabled the retrieval of temperature profiles and dust opacities (Conrath et al., 2000; Smith et al., 2001). The Mars Orbital Camera (MOC) provided wide camera angle visible imagery of the

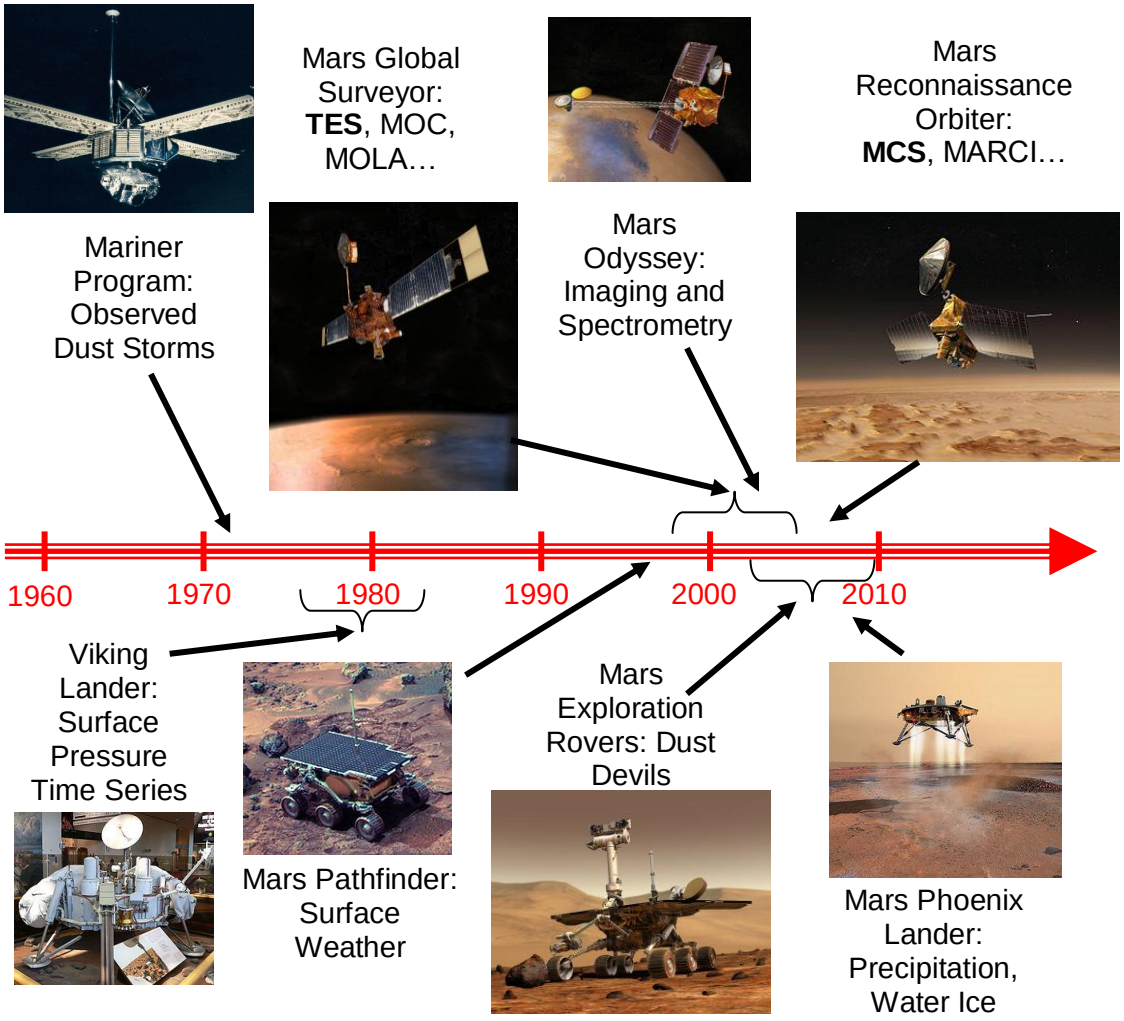


**Figure 1.4: Timeline of spacecraft exploration of Mars that had a significant impact on our understanding of Mars weather and climate. Mars Global Surveyor (MGS; 1997–2006) and Mars Reconnaissance Orbiter (MRO; 2006–present) observations are used in this study. Images courtesy of NASA.**

surface (Figure 1.5), which was particularly helpful for observing the progress of dust storms and water ice clouds and assessing their annual repeatability (Cantor et al., 2002). The Mars Orbiter Laser Altimeter (MOLA) provided high resolution altimetry data of the planet (Kreslavsky and Head, 2000), which was used to improve the representation of topography (Figure 1.2) and surface roughness in Mars models.

The Mars Pathfinder's Sojourner Rover (19.30 N, 33.52 W) captured a brief surface meteorological record of surface pressure, temperature, and wind for three

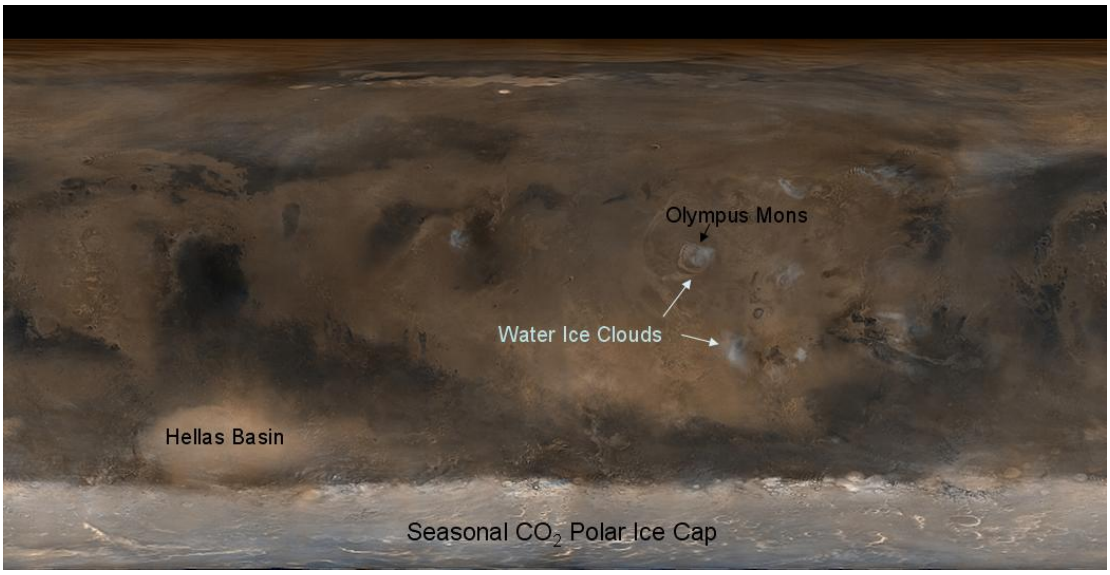


**Figure 1.5: Visible image of Mars from the Mars Global Surveyor (MGS) Mars Orbital Camera (MOC), depicting a seasonal polar cap, varying surface albedo, and water ice clouds. Image courtesy of NASA / JPL / Malin Space Science Systems and Bruce Cantor.**

months in 1998, and in situ atmospheric vertical structure was obtained during descent (Schofield et al., 1997). Winds rotated in a clockwise manner each day, which is associated with the thermal tide (the atmospheric response to solar heating); daily variations of 3–4 K temperature and 2–3 Pa were associated with synoptic weather systems.

The 2001 arrival of Mars Odyssey, still operational as of 2011, provided an additional source of visible and IR imagery from the Thermal Emission Imaging System (THEMIS), which has been used to ascertain dust and water ice optical depths at a later local time than TES (Smith et al., 2003). The European Space Agency's Mars Express is also currently in orbit, and has been used to map the surface and subsurface, assess atmospheric composition (Formisano et al., 2004), and study the evolution of the polar ice caps (Langevin et al., 2005).

The durable Spirit (landed 1.95° S, 34.47° E) and Opportunity (landed 14.57° S, 175.48° E) rovers, arriving in 2004, have become the longest operating surface explorers of Mars. Spirit became trapped in loose soil in 2009 after a 7.7 km journey; as of early 2011, Opportunity has traversed 26 km and is still roving. During their missions "cleaning events," where strong winds removed a layer of dust from the solar panels, helped extend the duration of the missions (Vaughan et al., 2010). The rovers were also able to image dust devils (Greely et al., 2006) as well as track changes in visible optical depth due to atmospheric dust (Lemmon et al., 2004).

The 2008 Phoenix lander survived several months in the Martian Arctic (landed 68.219° N, 234.251° E), taking meteorological observations of surface pressure, temperature, and wind speed (Taylor et al., 2008), until it was enveloped in a coating of $CO_2$ ice during the approach of winter. The onboard LIDAR detected evidence of snow falling from cirrus-like water ice clouds at night (Whiteway et al., 2009), as well as near-surface fog (Moores et al., 2011). Visual evidence was also provided of water ice just below the surface (Smith et al., 2009), which sublimated into the atmosphere over the course of a few sols. Surface winds, which reached a maximum of 16 m $s^{-1}$, were linked to dust lifting (Holstein-Rathlou et al., 2010). Brief surface pressure drops (on the order of 20 seconds) were attributed to dust devils, which peaked in frequency around noon (Ellehoj et al., 2010).

The Mars Reconnaissance Orbiter (MRO) arrived in March 2006, and is still operational as of early 2011. The Mars Climate Sounder instrument (MCS) is a limb sounder located aboard the spacecraft (McCleese et al., 2007). The MCS retrievals (Kleinboehl et al., 2009) have been used to provide improved temperature constraints

for models up to 80 km altitude, and first-ever vertical distributions of dust and ice clouds (McCleese et al., 2010). The MARCI camera has been used to examine changes in the daily weather on Mars (Malin et al., 2011).

### *1.4 The GFDL Mars Global Circulation Model*

Like terrestrial weather and climate models, Mars global circulation models (GCMs) numerically integrate the equations of motion for the atmosphere, albeit with a radiation scheme, boundary conditions, and physical parameterizations appropriate to Mars. Notable Mars GCMs include the Oxford-LMD model (Forget et al., 1999; Lewis et al., 1999), the NASA AMES model (Haberle et al., 1993), MarsWRF (Richardson et al., 2007), and the GFDL model (Wilson et al., 2002), which we use in this study.

The Geophysical Fluid Dynamics Laboratory (GFDL) Mars Global Circulation Model (MGCM) was developed by John Wilson (Wilson and Hamilton, 1996). Originally based on the GFDL SKYHI terrestrial GCM, the MGCM has more recently been adapted to the GFDL Flexible Modeling System. The present version of the MGCM uses a finite volume dynamical core (Lin, 2004). For simplicity we have selected a 36 × 60 latitude-longitude horizontal grid for our studies, corresponding to a 300 km horizontal resolution. The model uses a hybrid sigma vertical coordinate that follows the terrain near the surface, and transitions to purely pressure in the upper atmosphere. Figure 1.6 illustrates the 28 vertical levels used in this study, which extend from the surface (~600 Pa) to ~90 km altitude (~0.03 Pa).

The model includes a budget for gaseous and condensed $CO_2$, which simulates the cycling between the atmosphere and the seasonal polar ice caps and matches the

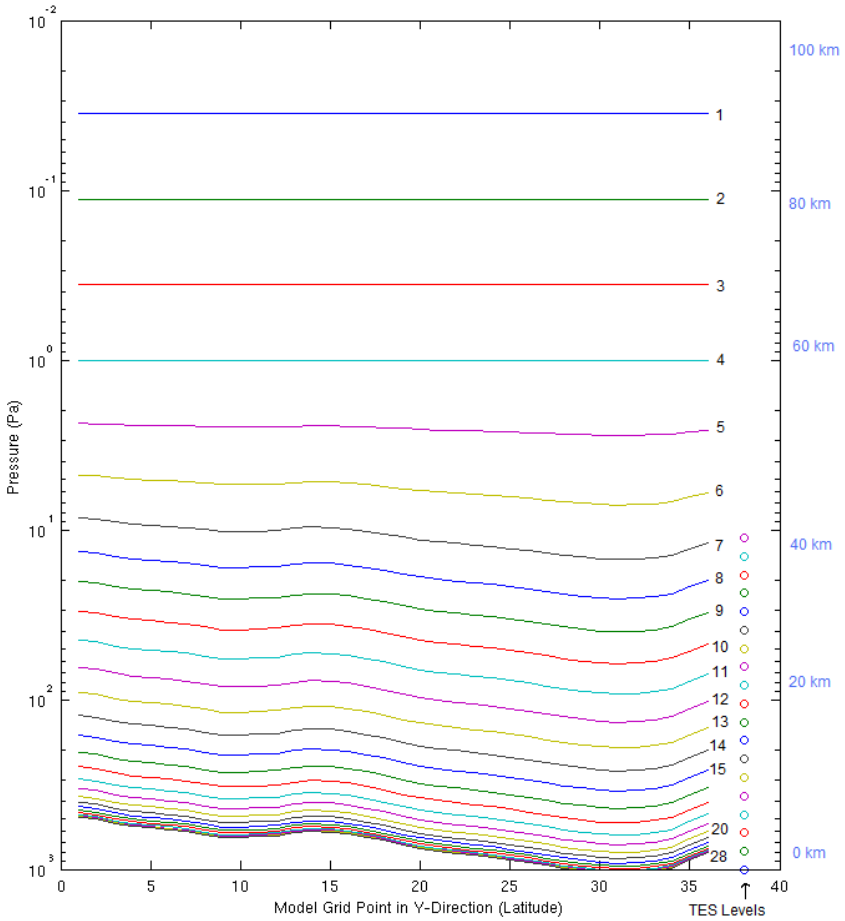


**Figure 1.6: Vertical levels employed by the GFDL Mars Global Circulation Model (MGCM).**

annual features of surface pressure changes observed by the Viking lander barometers (Hourdin et al., 1993). The radiation scheme considers solar and IR radiative transfer for gaseous $CO_2$ and suspended dust aerosol, with dust optical properties based on Ockert-Bell et al. (1997) in the visible and Wolff and Clancy (2003) in the IR part of the spectrum. Surface input fields consist of topography, albedo, thermal inertia, emissivity, and spatially-varying surface roughness (Heavens et al., 2008). Physics include a 12-layer soil model tuned to fit observed TES surface temperatures (Wilson et al., 2007), a boundary layer scheme, a mass-conserving water budget, and, recently, water ice cloud microphysics (Montmessin et al., 2004) and a

parameterization for momentum deposition from the breaking of topographically generated gravity waves at high altitudes (Garner, 2005).

The model has several options to represent dust aerosol, which is represented by a distribution of (typically 5) particle size bins from 1.5–15 μm. The simplest configuration is achieved through specifying a global opacity value (normalized to a surface pressure of 6.1 hPa). Dust is distributed in the vertical according to the Conrath profile (Conrath, 1975; Lewis et al., 1999), where mixing ratio decreases exponentially with pressure above a specified altitude, although recent Mars Climate Sounder (MCS) observations show that vertical distribution of dust can be more complex (McCleese et al., 2010). Most recently, the MGCM can input temporally and spatially varying maps of dust opacity, such as interpolated maps of MY 24 TES opacity. Here, the model injects dust into the boundary layer to match the observed opacity. Finally, the model can self-consistently lift, advect (using 3-dimensional tracer fields spanning several size bins), and deposit dust, with the dust field exhibiting feedbacks with the meteorological fields.

There are two physical mechanisms for dust lifting: convection and wind stress. Convective lifting is primarily found in tropical regions as the result of dust devils (Balme et al., 2003; Fisher et al., 2003), and is parameterized in the model based upon static stability of the boundary layer. Dust lifting from surface wind stress due to mid-latitude traveling waves is linked with regional and larger storms, especially during NH late autumn and early spring. The strength of this parameter, which depends upon surface winds (Ginoux et al., 2001), must be carefully tuned; feedbacks with meteorological fields are nonlinear, and too large a stress lifting

parameter can result in a runaway dust storm. Wilson and Kahre (2009) used finite surface dust reservoirs to help explain interannual variability of dust storm activity. Simulation of dust storms remains a significant challenge to the Mars community.

## *1.5 Mars Model Climatology*

Figure 1.7 describes the diurnal cycle of near surface temperatures (model Level 28, or ~0.3 km altitude) at consecutive intervals of 0.25 sol at the Northern Winter Solstice ($L_s$ 270°). The sub-solar point is therefore around 25° south latitude.

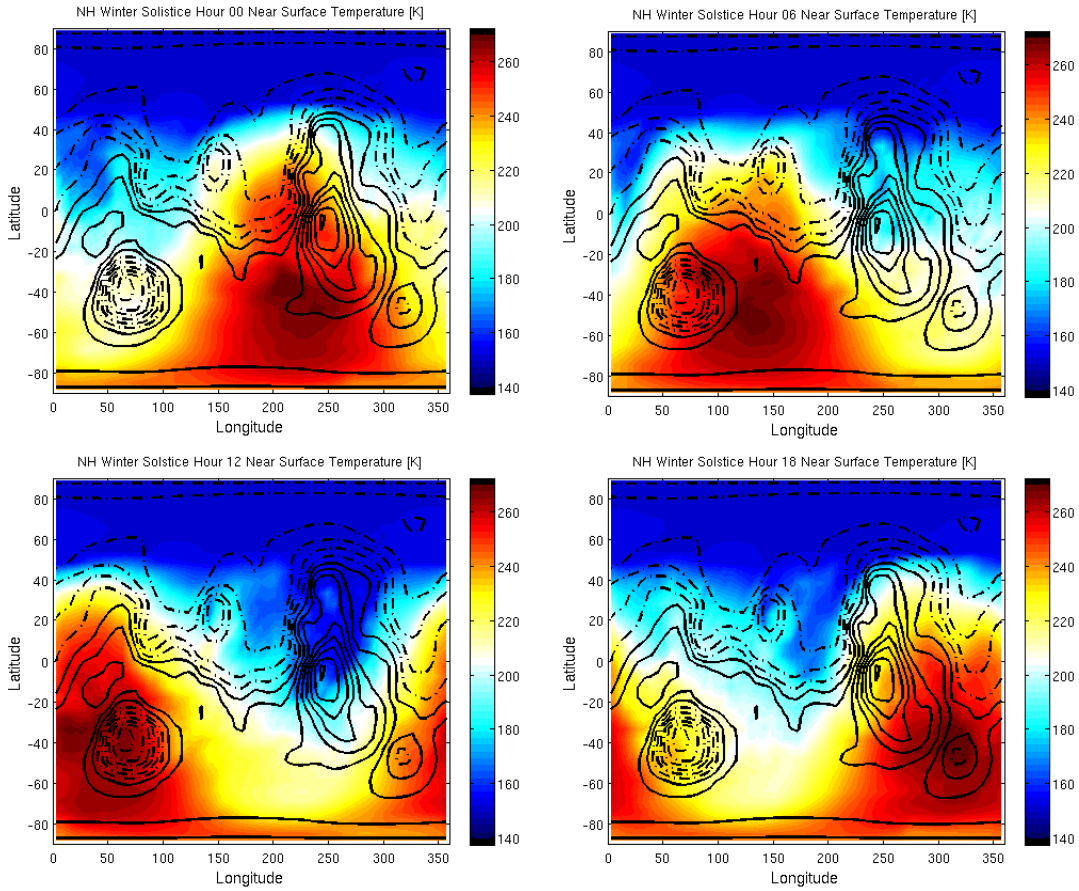


**Figure 1.7: MGCM near-surface temperature field (shaded, K) at the Northern Winter Solstice ($L_s$ 270°) as it evolves through a daily cycle (in 0.25-sol increments). Contours are topography, with high elevations having solid contours, and lower elevations having dashed contours.**

A tongue of warm temperatures tracks the sub-solar point as it moves westward across the surface of Mars. Unlike the Earth, these warm temperatures (230–270 K) extend all the way to high latitudes during the long summer days. Daily contrasts in temperatures on this desert planet can be extreme, as much as 80 K. In the northern polar night, temperatures are limited to around 145 K, the freezing point of carbon dioxide at surface pressures. The diurnal tide dominates the temperature field despite the extreme topography.

Figure 1.8 shows zonal mean climatologies of temperature and zonal wind for the time periods immediately following the NH spring equinox ($L_s$ 180°–210°) and winter solstice ($L_s$ 270°–300°). At equinox there is relative symmetry between the hemispheres, with a surface temperature maxima near the equator, and slightly cooler

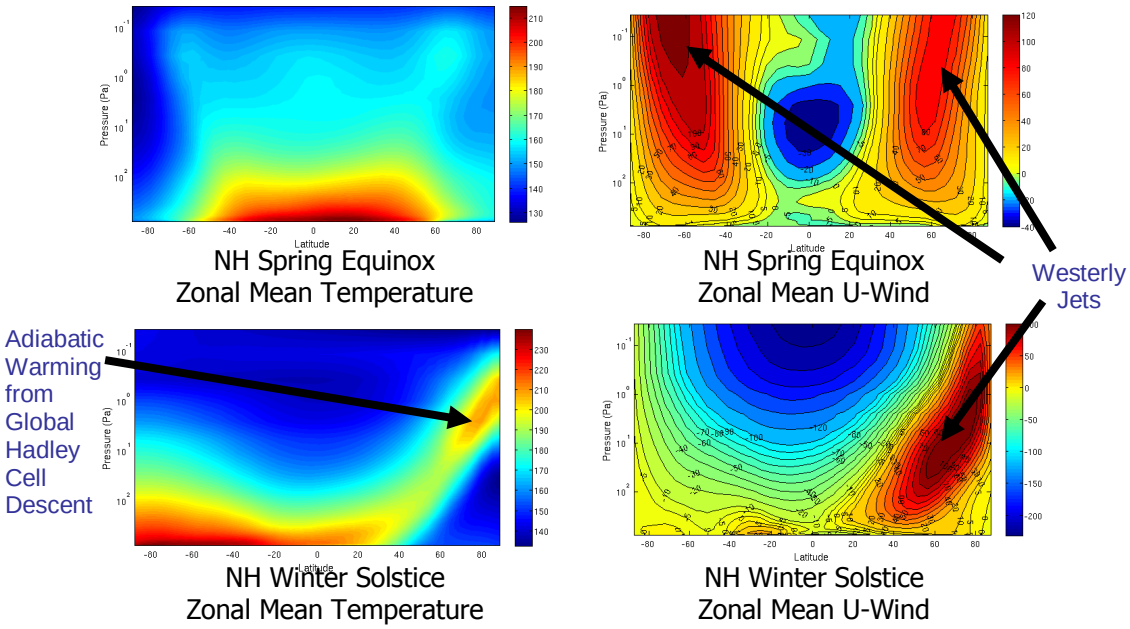


**Figure 1.8: Zonal mean climatology of the MGCM near the NH spring equinox ($L_s$ 0°, top row) and NH winter solstice ($L_s$ 270°, bottom row) of temperature (left panels) and zonal wind (right panels).**

conditions in the SH which is heading into its winter. Strong 80–100 m $s^{-1}$ westerly jets dominate mid-latitude flow above 10 Pa in each hemisphere. A weaker 30 m $s^{-1}$ easterly jet is present in the tropics at around 10 Pa (40 km altitude). The picture becomes highly asymmetric during NH winter. The westerly jet is present in the winter hemisphere with its maxima at 20–60 km, and the easterly tropical jet has strengthened considerably at high altitudes. Near surface temperatures are warmest not in the tropics but in the mid-latitudes and polar regions of the summer hemisphere which experience the longest daylight hours. A warming inversion is present in the winter pole at 40–60 km altitude. As there is no ozone layer on Mars, this warming aloft is not analogous to the stratospheric inversion in the terrestrial atmosphere. Rather, this polar warming (Wilson, 1997; McDunn et al., 2011) is the result of adiabatic warming from the descending branch of a global Hadley circulation. These zonal mean plots agree well with the output of the UK-LMD GCM (Read and Lewis 2004, Figure 3.19 and 3.21).

Figures 1.9 and 1.10 explore the impact of our choice in horizontal resolution upon the fidelity of the simulations. Our choice of resolution in this project, 5 × 6 degrees (36 × 60, or ~300 km) is compared to a higher resolution (2 × 2.4 degrees, or ~120 km) MGCM simulation for a snapshot in the NH autumn ($L_s$ 165°). Given the extremes of Martian topography (Section 1.1), the most noticeable differences are in the near surface features. The coarser resolution is able to resolve the major topographic features (Tharsis plateau, Hellas and Argyre basins), but not in as much detail as the high resolution run, particularly for the volcanic peaks (Figure 1.9, top panels). Surface temperatures (middle panels) are dominated more by the diurnal tide

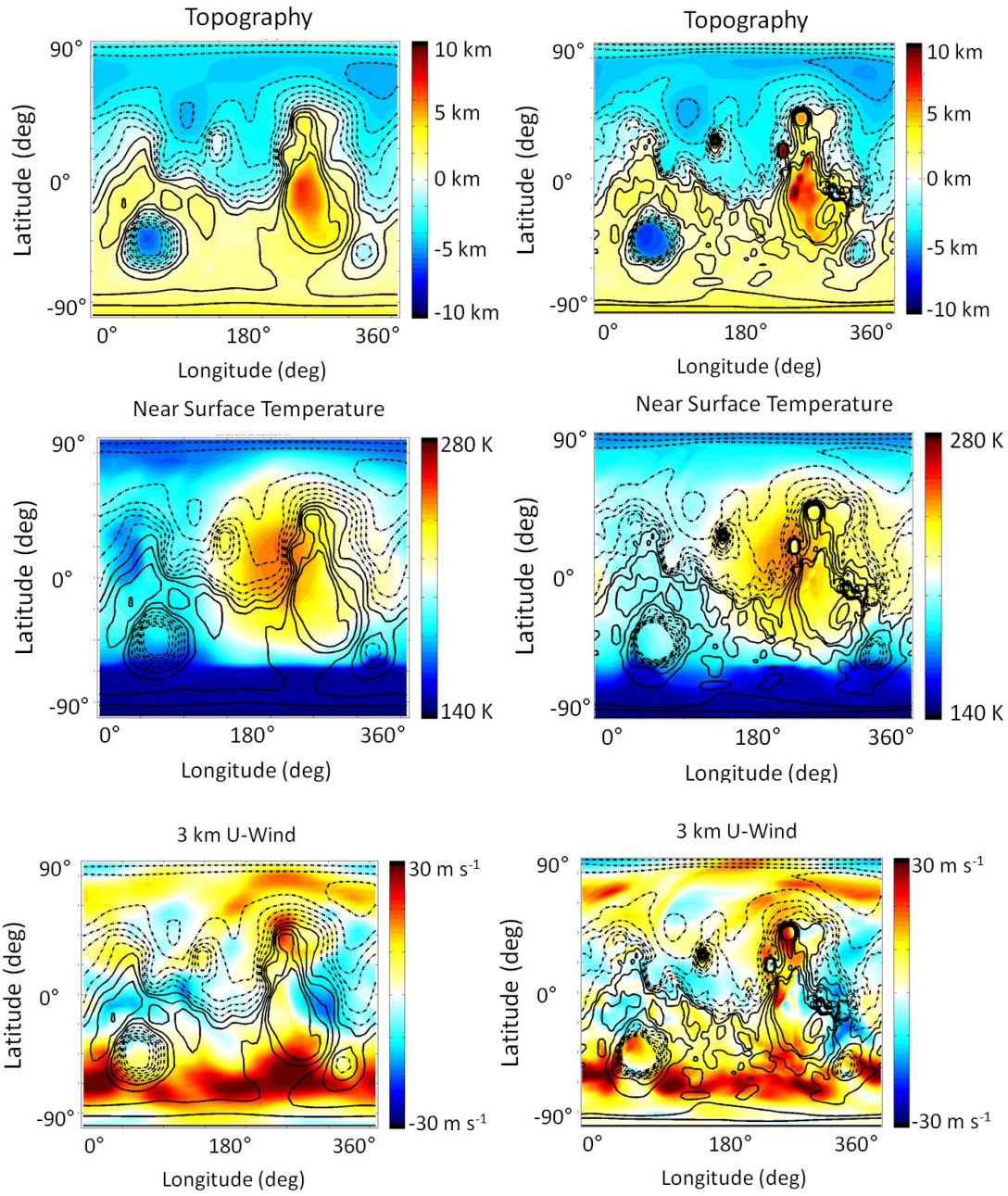


**Figure 1.9: Impact of MGCM resolution on near surface fields (snapshots from NH autumn). Left columns employ 5 × 6 degree (~300 km) horizontal resolution, right columns employ 2 × 2.4 degree (~120 km) resolution. Top: model surface topography. Middle: model near surface (Level 28) temperature. Bottom: model ~3km (Level 25) zonal wind speed.**

and surface characteristics. However, topography plays a very important role in near surface winds (bottom panels), as upslope and downslope winds dominate in the diurnal cycle. Mesoscale resolution may also be necessary to simulate local dust

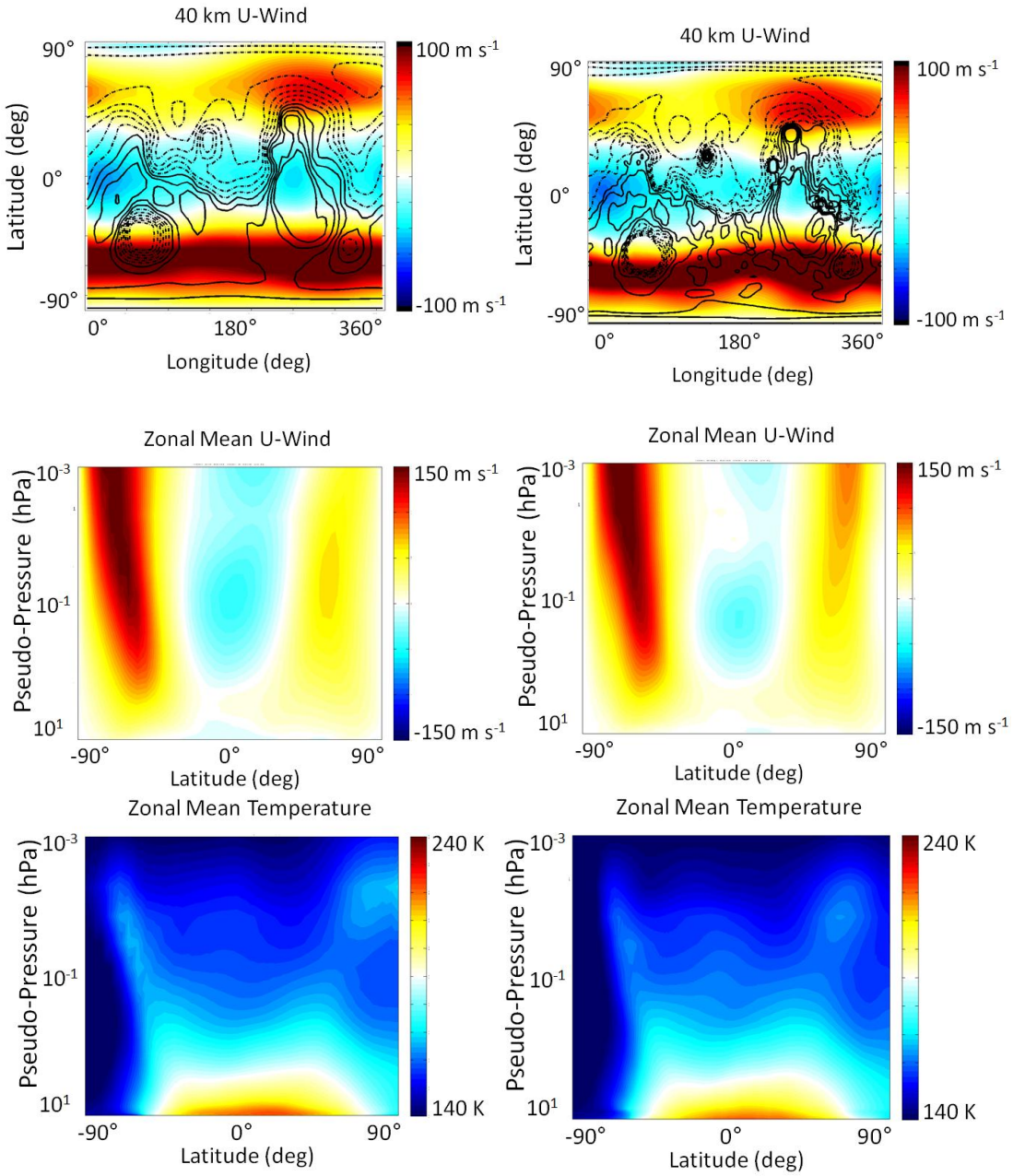


**Figure 1.10: Impact of MGCM resolution on upper level and zonal mean fields (snapshots from NH autumn). Left columns employ 5 × 6 degree (~300 km) horizontal resolution, right columns 2 × 2.4 degree (~120 km) resolution. Top: ~40 km (Level 7) u-wind speed. Middle: zonal mean u-wind speed. Bottom: zonal mean temperature.**

lifting events (Rafkin, 2011). Nonetheless, local wind maxima on a scale larger than 300 km are captured well at both resolutions. Figure 1.10 shows that upper level features agree well between coarse and high resolution runs. For example, at 40 km a

NH jet maxima at 280° longitude and tropical jet maxima at 30° longitude are captured in both (top panels). Features of the zonal mean u-wind (middle panels) and temperature (bottom panels) also generally agree. Therefore, in order to enable a multitude of ensemble experiments without excessive runtime, we use a 5 × 6 degree horizontal resolution for this study. As better computational resources become available, we plan to switch to a higher resolution for future studies.

### *1.6 Scientific Objectives and Outline*

This chapter has demonstrated the wealth of insights provided by both spacecraft instruments and general circulation models toward the weather and climate of Mars. However, spacecraft remote sensing observations are irregular in space and time, and model simulations may diverge from the real atmosphere. A synthesis of these two sources of information is needed to further advance our understanding of the atmosphere of Mars. Data assimilation provides the optimal framework for combining sparse observations with the full fields and forecasting capabilities of a model. A reanalysis is a sequence of analyses from a data assimilation system that represent our best estimate of the state of the atmosphere for a historical record spanning several years. A team of scientists (see Acknowledgements) is preparing a Mars reanalysis of the Thermal Emission Spectrometer (TES) and Mars Climate Sounder (MCS) datasets using the GFDL MGCM and the Local Ensemble Transform Kalman Filter (LETKF) data assimilation system. This dissertation represents an important step in creating this reanalysis, building upon the experiments of M. Hoffman et al. (2010) to include the first assimilation and validation with real spacecraft observations for this system.

The following scientific questions are addressed in this research. (Q1) Considering the unique characteristics of Mars and its observing systems, what is an optimal design of an ensemble data assimilation system for Mars? (Q2) How do analyses from assimilation compare to other reanalyses and independent observations? (Q3) How can insights from data assimilation aid remote sensing and model development for Mars, and improve understanding of traveling waves, water ice clouds and dust distributions? (Q4) What are the locations, seasonal evolution, and physical origins of instabilities in the Martian atmosphere? (Q5) What is the predictability horizon for Mars numerical weather prediction?

Chapter 2 uses the technique of bred vectors to investigate atmospheric instabilities (Q4, Q5) and their physical origins (Q3), and links instabilities to errors in data assimilation systems (Q3). Chapter 3 describes the implementation of an ensemble data assimilation for TES temperatures and advanced techniques that improve the performance of the system (Q1), and uses forecasts based on the system to create Mars weather forecasts (Q5). Finally, Chapter 4 compares the analyses with the UK reanalysis and independent Radio Science profiles (Q2), explores assimilation of MCS observations (Q1, Q3), and presents a strategy for dust assimilation (Q3).

# Chapter 2: Elucidating Atmospheric Instabilities with Bred Vectors

## *2.1 Introduction to Bred Vectors*

Breeding (Toth and Kalnay, 1993), a technique for elucidating atmospheric instabilities, is based on the principle that, in a chaotic system, two states that are initially similar will diverge over time; stable perturbations will diminish, whereas unstable perturbations will grow in amplitude until they dominate the difference between the states. These patterns of differences, or bred vectors, are manifestations of physical instabilities in the model. In the context of forecasting, the bred vectors represent forecast errors due to uncertain initial conditions, and thus breeding can be helpful in understanding the performance of a data assimilation system (Chapter 3; Hoffman et al., 2010). The locations and preferred seasons of the instabilities, as well as physical explanations for their origins, can also give insights into the dynamics of the Martian atmosphere, since instabilities may play a role in wave regime transitions, or in the growth of dust storms.

Newman et al. (2002) applied the bred vector technique to the Oxford-LMD Mars GCM. This study reveals that there are seasons where instabilities are active (northern late autumn and winter) as well as relatively quiescent (northern summer), in contrast with the Earth where there are growing modes present in every season. Rogberg et al. (2010) examined predictability using forecasts based on the UK Reanalysis, and found similar seasonal variations in error growth.

In this chapter, the bred vector technique is applied to the atmosphere of Mars with the goal of characterizing spatial patterns and temporal evolution of dynamical

instabilities, as well as providing physical explanations for their origins. Section 2.2 provides a description of the Mars global circulation model used in this study, as well as the algorithm and configuration for the breeding technique. Section 2.3 presents results for a fixed dust (0.3 opacity) atmosphere, including time evolution of BV magnitude, seasonal BV patterns, and spectral analysis. Section 2.4 investigates the origin of the instabilities through the application of BV energy equations as well as other instability criteria. Section 2.5 examines the results of breeding in a data assimilation framework, comparing instabilities with ensemble spread. Finally, Section 2.6 concludes and compares the results of this study with previous work.

## *2.2 Methodology*

### *2.1.1 Atmospheric Model*

This study employs the NOAA-GFDL Mars Global Circulation Model (MGCM), developed by John Wilson (Wilson and Hamilton, 1996; Richardson and Wilson, 2002; Hinson and Wilson 2004). The model employs a finite volume dynamical core (Lin, 2004). Output dynamical variables consist of temperature ($T$), zonal component of the wind ($u$), meridional component of the wind ($v$), and the difference in pressure between model levels (*delp*), from which surface pressure can be diagnosed. For simplicity, a latitude-longitude grid has been used for this study, although the model runs most recently using a cubed-sphere configuration. Horizontal resolution is 5° latitude by 6° longitude, or 36 × 60 grid points.

The model in its original configuration uses a hybrid sigma-pressure vertical coordinate for its output / restart files. For this study, the default levels employed in Hoffman et al. (2010) were modified so that the model would be run strictly on

terrain-following sigma surfaces. This vastly simplifies the derivation of the perturbation kinetic energy equation (see Appendix A), and enables simpler diagnostics of dynamical processes. Table 2.1 describes the vertical levels employed; sigma values were selected so that the pressure values at each level, if one assumes a surface pressure of 610 Pa (a common reference value for Mars), are

**Table 2.1: MGCM vertical levels, sigma definitions, reference pressures, and reference heights. Level 1 is defined as a 0.002 Pa pressure level rather than a sigma level due to a model requirement. Reference pressures are computed assuming a surface pressure of 610 Pa. Reference heights are calculated using an atmospheric profile at a location with 610 Pa surface pressure.**

| Model Level | Sigma | Reference Pressure (Pa) | Reference Height (km) |
|---|---|---|---|
| 1 | | 0.002 | |
| 2 | 0.0001 | 0.061 | 81 |
| 3 | 0.0005 | 0.305 | 70 |
| 4 | 0.0013 | 0.793 | 61 |
| 5 | 0.0033 | 2.013 | 53 |
| 6 | 0.0077 | 4.697 | 46 |
| 7 | 0.0151 | 9.211 | 41 |
| 8 | 0.0250 | 15.25 | 36 |
| 9 | 0.0386 | 23.55 | 32 |
| 10 | 0.0588 | 35.87 | 28 |
| 11 | 0.0880 | 53.68 | 25 |
| 12 | 0.1284 | 78.32 | 21 |
| 13 | 0.1815 | 110.7 | 18 |
| 14 | 0.2476 | 151.0 | 15 |
| 15 | 0.3252 | 198.4 | 12 |
| 16 | 0.4113 | 250.9 | 9.8 |
| 17 | 0.5012 | 305.7 | 7.7 |
| 18 | 0.5901 | 360.0 | 6.0 |
| 19 | 0.6734 | 410.8 | 4.5 |
| 20 | 0.7478 | 456.2 | 3.4 |
| 21 | 0.8113 | 494.9 | 2.5 |
| 22 | 0.8633 | 526.6 | 1.8 |
| 23 | 0.9045 | 551.7 | 1.3 |
| 24 | 0.9360 | 571.0 | 0.91 |
| 25 | 0.9594 | 585.2 | 0.64 |
| 26 | 0.9763 | 595.5 | 0.46 |
| 27 | 0.9883 | 602.9 | 0.33 |
| 28 | 0.9966 | 607.9 | 0.26 |

similar for both coordinate systems. There are 28 vertical levels, with roughly half the levels located in the lowest ~15 km, and with the highest level extending vertically in excess of 85 km above the surface, or less than 0.1 Pa in pressure.

The model includes a gaseous and condensed $CO_2$ cycle that allows for seasonal accumulation at the poles, as well as tracers for dust and water vapor. Initial experiments have been performed using a temporally and horizontally constant normalized dust opacity of 0.3 following the Conrath vertical distribution (Conrath 1975), which corresponds to a relatively low level of atmospheric dust. Future experiments will include a dust field updated by observations, or a dynamical dust distribution using surface wind stress and convective lifting parameterizations and allow radiative feedback between the dust and temperature fields.

### *2.2.2 Bred Vector Methodology*

The procedure for breeding is straightforward. First, a "nature run" is created by integrating the full dynamical model (the MGCM) for two Martian years. Second, an initial random perturbation is added to the nature run at the initial time. Third, the perturbed run evolves forward using the MGCM for a period of time known as the rescaling interval. The difference between the control run and the perturbed run is known as the bred vector (BV). Next, after the rescaling interval has elapsed, the bred vector is reduced in size so that it is the same size as that of the initial perturbation according to a specified norm. The process of running the full model in time, and then rescaling the bred vector, is repeated for the entire duration of two Martian years.

On Mars, time is commonly kept using solar (or areocentric) longitude $L_s$, a seasonal index that begins at 0° at NH spring equinox and passes through 360° over the course of a Mars year (Figure 1.3). The experiment begins on the NH summer solstice (solar longitude $L_s$=90°), using climatology for initial conditions, and the length of the experiment is two Martian years (668 sols, or Martian days each). The rescaling time interval is 6 Martian hours (0.25 sol), and the rescaling amplitude is 1 $K^2$, as given by the temperature-squared norm scaled by area (cosine latitude). All dynamical variables (*T*, *u*, *v*, *delp*) are rescaled by the same factor. Note that bred vectors are to first order only sensitive to the rescaling time interval and initial amplitude, and not to the norm. Here, rescaling only occurs when the BV size is larger than that of the original perturbation amplitude; if it is smaller, then the perturbed run is not modified. Bred vectors are kept "young" by adding a small random perturbation with 1% of the original perturbation size, to the perturbed run every rescaling interval. This ensures that the subspace spanned the unstable modes should be well sampled by the BVs. For each hemisphere a separate BV amplitude is computed and rescaling factor used so that the more active hemisphere does not dominate the signal. For example, in late NH autumn ($L_s$~200°), traveling waves are prevalent in the mid-latitudes of both hemispheres, but globally-scaled bred vectors would show only the SH to be dominant because the SH instabilities grow faster. A linearly weighted average of the factors is used between 20° S and 20° N latitude to avoid spatial discontinuities.

## *2.3 Breeding Results*

### *2.3.1 Experiment Configuration*

The breeding method is applied to the MGCM using a fixed dust distribution with a normalized opacity (at surface pressure of 610 Pa) of 0.3, which corresponds to an atmosphere with a relatively low level of dust. Vertical distribution follows the Conrath profile (Conrath 1975) where dust concentration decreases by an exponential function above a certain height, in this case around 30 km. Beginning from the NH summer solstice ($L_s$ 90°), breeding occurs for a period slightly in excess of a Martian year, concluding at $L_s$ 120° during the NH summer. A second set of bred vectors, referred to as Year 2, continues from the end of the first run and proceeds through the following NH summer to $L_s$ 120°.

For the purpose of comparison with the bred vectors, a dataset of atmospheric traveling waves is compiled for Year 1. A smoothed atmospheric climatology is created for each day of the Martian year by performing a temporal average at each synoptic time (Mars Hour 00, 06, 12, or 18) over a moving window of 21 sols (10 sols before and after a given date). Eddies are calculated as the difference between the atmospheric state on a given date and the smoothed climatology. Because the temporal averaging occurs at the same time of day, the thermal tide is removed. In addition, standing waves are included in the temporally-smoothed climatology and therefore not represented in the eddy state, which should be reflective of traveling wave activity (weather systems).

### *2.3.2 Bred Vector Time Evolution*

Figure 2.1 reveals the annual evolution of bred vector amplitudes for the first year of a run beginning at the NH summer solstice, or $L_s$ 90°. Given the sub-daily fluctuations in BV amplitudes (light lines), the darker lines represent values smoothed over 10 sols. Values are reported separately for each hemisphere. The first 30 sols of simulation ($L_s$ 90°–105°) show small bred vector amplitudes (0.1 $K^2$) compared to the initial perturbation (1 $K^2$). During spin-up time, perturbations that projected onto decaying modes vanish; the additional random noise added every 6 hours helps to ensure that all of the growing directions are captured by the bred vectors. However, by comparing this time period with the same season one year later, both years exhibit

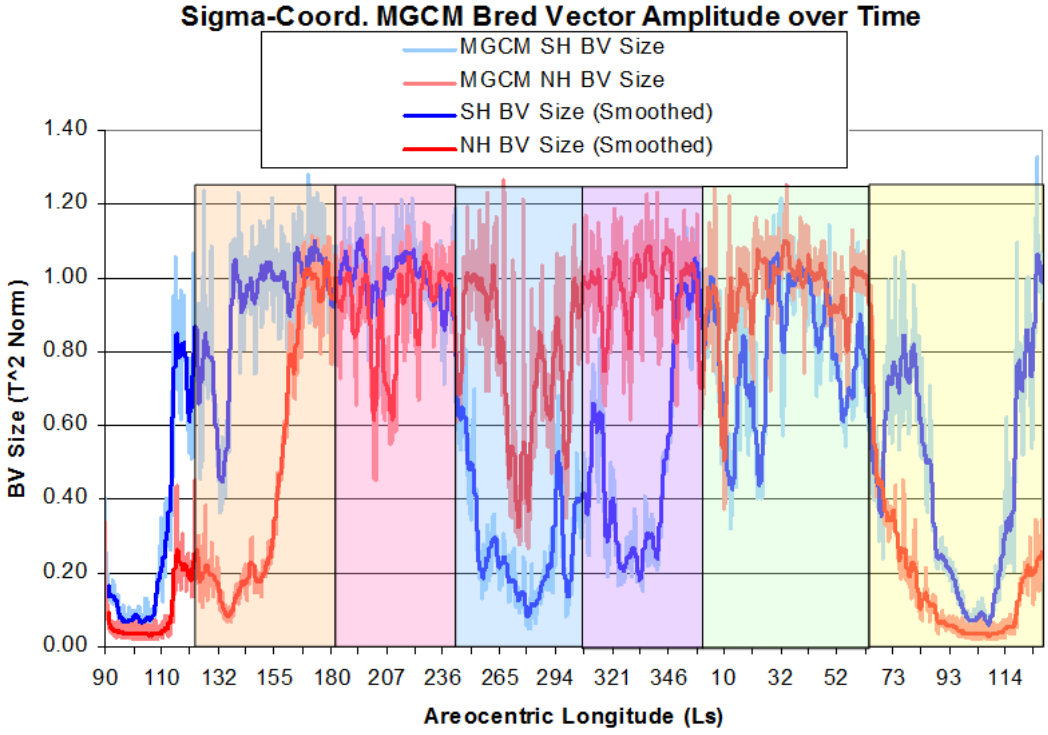


**Figure 2.1: Bred vector amplitude time evolution. SH amplitudes are in blue, NH in red. Raw BV amplitudes from year 1 are in light shading. 10-day smoothed amplitudes from year 1 are in dark shading. The six seasons are identified by background colors that match the colors in Fig. 1.3, and begin with NH early autumn ($L_s$ 120).**

a quiescent period around $L_s$ 100°, suggesting that the dynamics during this time of year are quite stable. During other times of the year, there exist instabilities in at least one hemisphere, particularly the winter hemisphere. Amplitudes occasionally exceed 1.2 $K^2$, indicating a perturbation growth of 20% in just 0.25 sol. Transitions between quiescent and active seasons occur over a short time, during which BV amplitudes can rise from 0.2 $K^2$ to 0.8 $K^2$ within a few $L_s$.

It is the growth rate rather than the absolute magnitude of the BV that is most important. We calculate growth rates based on the natural logarithm of the ratio of the amplitude at the end of the rescaling interval to the amplitude at the beginning of the interval. Comparisons between the Year 1 and Year 2 BV runs (Figure 2.2) show

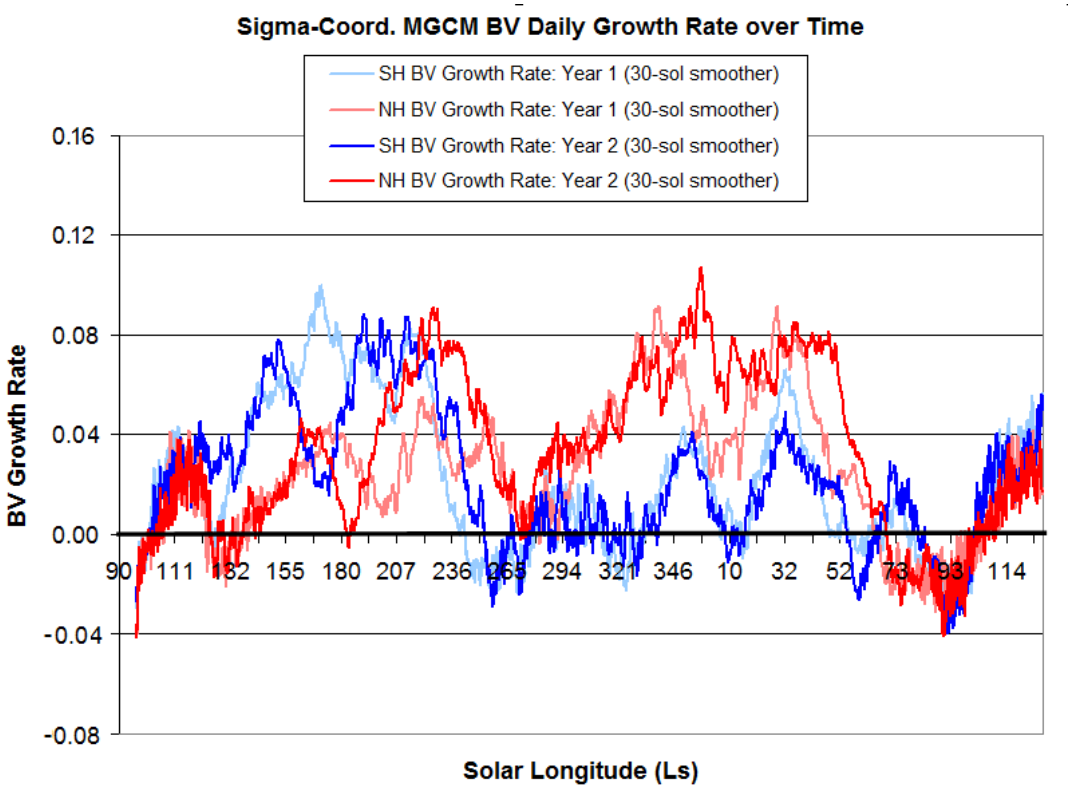


**Figure 2.2: Bred vector daily growth rate seasonal evolution for two Martian years, illustrating interannual variability. SH amplitudes are in blue, NH in red. 30-day smoothed amplitudes BV amplitudes from year 1 are in light shading, smoothed amplitudes from year 2 are in dark shading.**

remarkable consistency in bred vector activity from year to year, particularly in the time of transition from a quiescent period to an active one ($L_s$ 100°–120° and 340°–30° for SH, $L_s$ 270°–330° and 130°–160° for NH). As expected, the timing of shorter (1–10 day) quiescent periods or flare-ups does exhibit inter-annual fluctuations (see the period in the NH near $L_s$ 180° for an example). Maximum growth rates in a given rescaling period can exceed 2.0 per sol, although time averaged growth rates are significantly smaller, generally less than 0.1 per sol.

Figure 2.3 provides a view of zonal mean bred vector activity as a temporal mean over six Martian "seasons," each 60° in $L_s$ (Figure 1.3). The panels in Figure 2.3 are arranged so that the presence or lack of symmetry between hemispheres can be readily examined. Four patterns of instability are visible in Figure 2.3. The first is an area of large instability near the surface that is visible in the mid-latitudes and sub-polar regions. Its vertical extent is from 500–700 Pa pseudopressure, or only a few km in depth. This area is most active in the late spring and late autumn (top row panels), where the BV is largest at around 60° latitude in the spring hemisphere. During the NH winter hemisphere, the signal is located equatorward around 45° but is weaker due to dominant instabilities aloft. The zonal location of the BV follows a tight thermal gradient associated with the edge of the seasonal $CO_2$ ice cap. The instability largely disappears during the summer post-solstice period, as the seasonal ice cap has sublimated, eliminating the thermal gradient. The second pattern of instability is a belt that extends from just above the surface instability to around 1 Pa pseudopressure, or 60 km in height. This instability tends to follow the temperature gradient associated with the edge of the polar vortex. It is present during the winter

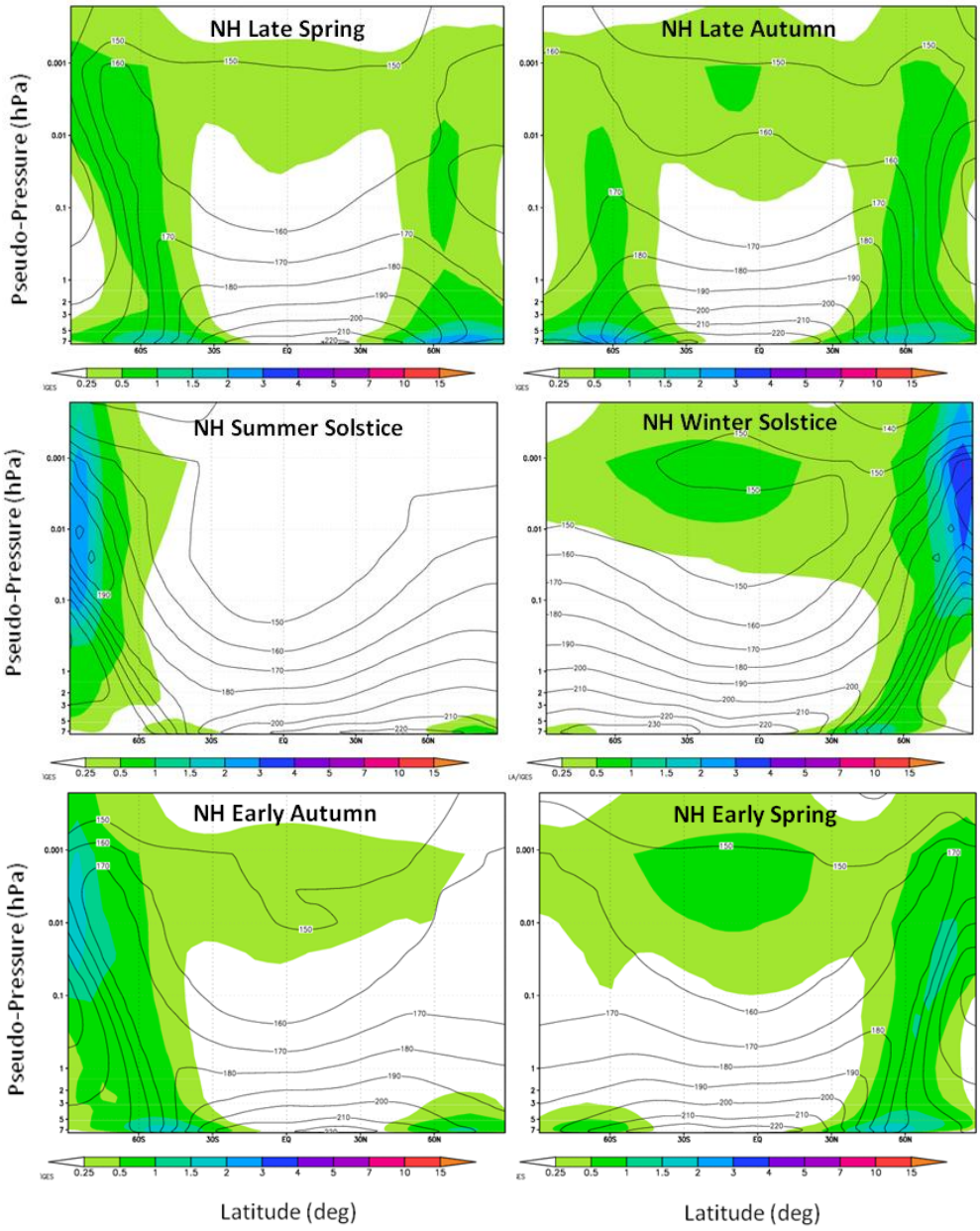


**Figure 2.3: Latitude-pseudopressure plots of bred vector structure as a temporal average for each Martian "season." Shading represents zonal mean BV amplitude; contours denote zonal mean temperatures. Plots are arranged so that a comparison of symmetry between NH and SH can be analyzed.**

solstice and transition seasons in the NH, and the transition seasons in the SH, and is absent during NH summer. Although weaker in magnitude than the surface instability, it covers a greater vertical extent. There is an asymmetry between the NH and SH during the solstice seasons (middle row panels), with the polar front

instability largely absent in the SH. It may be dominated by the third instability pattern, which is manifest as a maximum of BV activity located at the winter pole, from 0.1 to 10 Pa, or 40–80 km altitude. This pattern is only visible during the winter solstice season in the NH, and the winter solstice and following season in the SH. The final BV pattern is a relatively weak tropical instability at 0.3 Pa, or above 80 km. It is present during most seasons (SH winter solstice being the exception) and strongest during the NH winter, and may be associated with inertial instability (Wilson et al., 2002). There are asymmetries in BV between corresponding seasons (e.g., NH at $L_s$ 270° and SH at $L_s$ 90°), with stronger instabilities in the NH. These are due to enhanced thermal contrasts at perihelion (during NH winter) caused by the highly elliptical Martian orbit (Figure 1.3), as well as to the 5 km dichotomy in mean terrain height between the hemispheres.

### *2.3.3 Bred Vector and Eddy Zonal Wavenumber*

Figure 2.4 depicts a snapshot of bred vector activity at the NH autumnal equinox ($L_s$ 180°). The top panels compare the bred vectors with traveling wave activity (eddies) at Levels 20 and 25 (approximately 3.5 km and 0.6 km altitude, respectively). There is a correspondence between the eddies and the BVs, with wave 2 activity in the NH and SH, although the waves are out of phase with one another. This is because the BVs do not correspond to steady-state wave perturbations, but rather refer to the instabilities that give rise to changes in wave characteristics and amplitude. The bottom panels of Figure 2.4 show cross-sections through 60° N and 60° S latitude. The waves in the lowest 10 km of the atmosphere are largely distinct from those in the rest of the vertical column, with a higher wavenumber and more

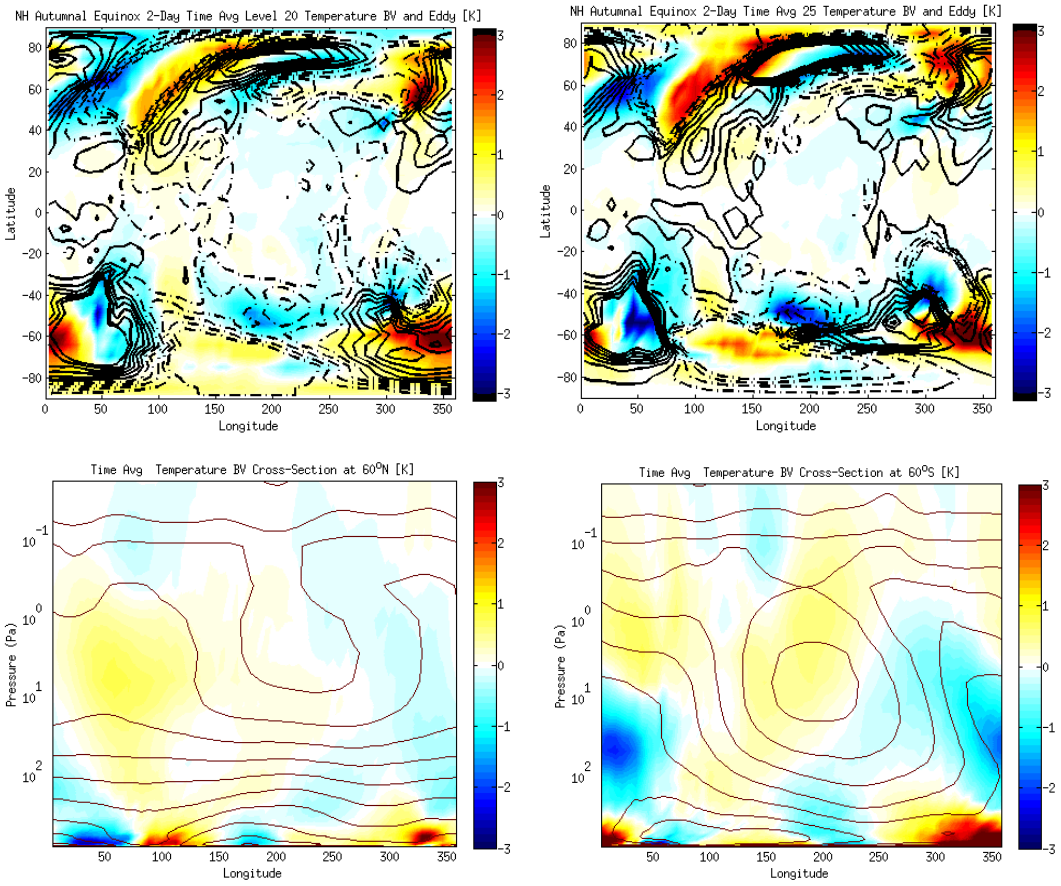


**Figure 2.4: 2-day time average plots following the NH Autumnal Equinox ($L_s$ 180°). Top Left: temperature transient eddy (contoured) from smoothed climatology and temperature BV (shaded) at approximately 3.5 km above the surface (Level 20). Top right: similar to top left, except at Level 25 (about 0.5 km above the surface). Bottom left: vertical cross section through 60° S latitude with temperature BV (shaded) and control (contours). Bottom right: vertical cross section through 60° N latitude with temperature BV (shaded) and control (contours).**

intense instability. The waves exhibit a westward tilt with height, hinting at baroclinic processes (this issue is explored further in Section 2.4.1). The manifestation of the bred vectors as wavelike features motivates the calculation of a dominant wavenumber for each level and hemisphere so that the evolution of wave regimes in time can be tracked.

Rows 1 and 3 of Figure 2.5 describe BV activity in each hemisphere at a model Level 25 (~ 0.6 km altitude), selected to illuminate the near surface cap-edge instability patterns described in the previous section. In order to summarize spatial

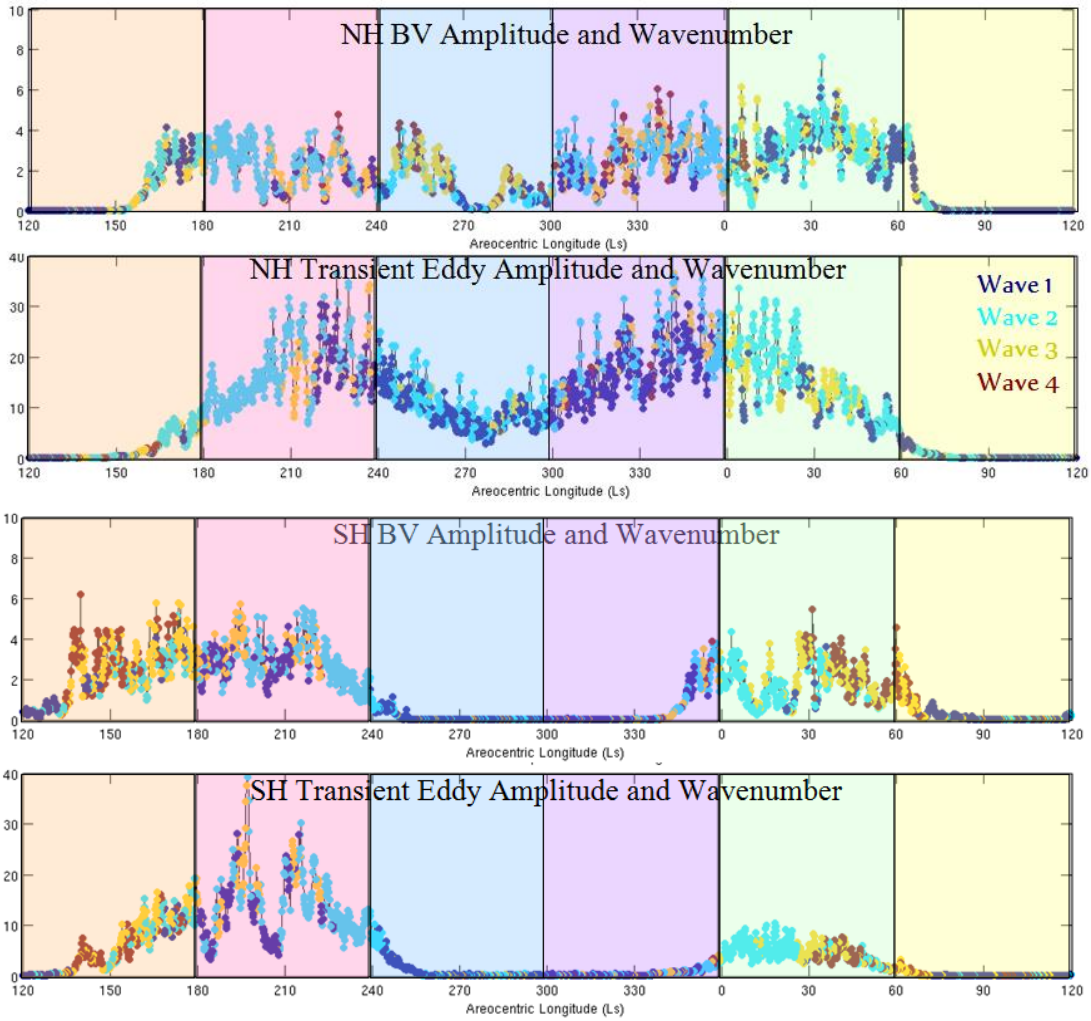


**Figure 2.5: Time series over one Martian year of bred vector (rows 1 and 3) and transient eddy (rows 2 and 4) temperature-squared amplitude ($K^2$; vertical axis) and dominant wave number (color of dots) for each hemisphere at Level 25 (near surface). Seasons are indicated by color panels. Note that the BV instability generally precedes the initiation of seasonal eddy activity.**

patterns as well as seasonal evolution of Bred Vector activity, zonal wavenumber regimes are assigned for each hemisphere and model level. Fourier analysis is performed at each latitude circle of 60 model grid points, and the absolute value of the Fourier coefficients corresponding to wavenumbers 1, 2, 3, and 4 are retained. To create a single regime assignment over all latitudes in the hemisphere, the mean Fourier amplitude for waves 1–4 is calculated for a specific hemisphere and level, and a regime assignment is made based on the dominant wave (largest mean amplitude of

the corresponding Fourier coefficient). The regime assignments are conveyed in Figure 2.5 using coloration: dark blue for wave 1, cyan for wave 2, yellow for wave 3, and red for wave 4. So that seasons can be shown in their entirety without any artifacts from spin-up, the time axis of the plot extends from $L_s$ 120° to $L_s$ 120° of the following year. Level 25 reveals waves 1, 2, 3, and 4 manifest at various times; for example, in the SH waves 3 and 4 dominate from $L_s$ 140°–180°, whereas there is a decaying wave 2 regime from $L_s$ 220°–240°. Wave 4 may be associated with transitions between wave regimes rather than a sustained pattern. Instability in the NH begins around $L_s$ 150°, undergoes a brief quieting around $L_s$ 270°, resumes in earnest at $L_s$ 300°, and extends through the NH spring until $L_s$ 70°. The presence of near surface wave activity is aided by the thermal gradient resulting from the seasonal polar ice cap. In the SH, activity begins at $L_s$ 345° and extends through $L_s$ 70°. There is then a quiescent period surrounding the solstice, from $L_s$ 70° to 120°, before SH spring activity continues through $L_s$ 240°. The upper levels (not shown) are most active near the respective winter solstice, with dominant wavenumbers of 1 and 2. There are two intervals (around $L_s$ 240° and $L_s$ 315°) that show activity in both hemispheres simultaneously, and may refer to the tropical instability.

Rows 2 and 4 of Figure 2.5 present an analysis of traveling wave (eddy) activity using the latitude-scaled, T-square norm. Note that these panels differ from the top panels; whereas the eddies are associated with traveling waves, BV refer to the waves' growth and decay (instability) rather than a steady-state (stable) solution. At Level 25, activity builds in a wave-2 regime from $L_s$ 165° to $L_s$ 210°. After plateauing with waves 1–3, amplitudes steadily decrease until $L_s$ 280°, when they

resume a climb to peak spring activity at $L_s$ 345°. This "solsticial pause" in wave activity is present in other GCMs as well (Read et al., 2011). An additional plateau is reached at $L_s$ 20° until the waves diminish at $L_s$ 65°. In the SH, autumn activity lasts from $L_s$ 0°–60°, and spring activity from 130°–250°. In upper levels (not shown), traveling waves reach a peak in both hemispheres around $L_s$ 240° with a wavenumber of 1. There are no appreciable traveling waves present in the atmosphere at any vertical level from $L_s$ 70°–130°, or the near entirety of Season 2. Seasons of traveling wave activity correspond with BV activity particularly well at low levels, with instability growth generally preceding the seasonal initiation of traveling waves. Hinson (2011) showed that wavenumbers 1–3 dominate in (coarse vertical resolution) Thermal Emission Spectrometer (TES) temperature observations, although Barnes (1984) and Tanaka and Arai (1999), using theoretical considerations, found instabilities in the basic flow to wave numbers 1 through 5 or more. TES data (Hinson 2011), as well as Viking surface pressure records (Barnes 1980), demonstrate that given wave regimes in the real atmosphere can be persistent over long durations of time; indications of regimes of a particular wavenumber are visible in Figure 2.5.

Figure 2.6 (top panels) shows histograms of the frequency of occurrence for each BV zonal wavenumber regime at Level 25 for various BV amplitudes. For waves with an amplitude > 1 $K^2$, wavenumber 2 is the most common, and wavenumber 4 the least common. The largest waves (BV amplitude > 4 $K^2$ at Level 25) feature a greater percentage of wavenumber 3 and 4 than for average BV amplitudes, particularly in the SH. The bottom panels of Figure 2.6 show wavenumber histograms for the fastest growing waves (increase in BV amplitude of

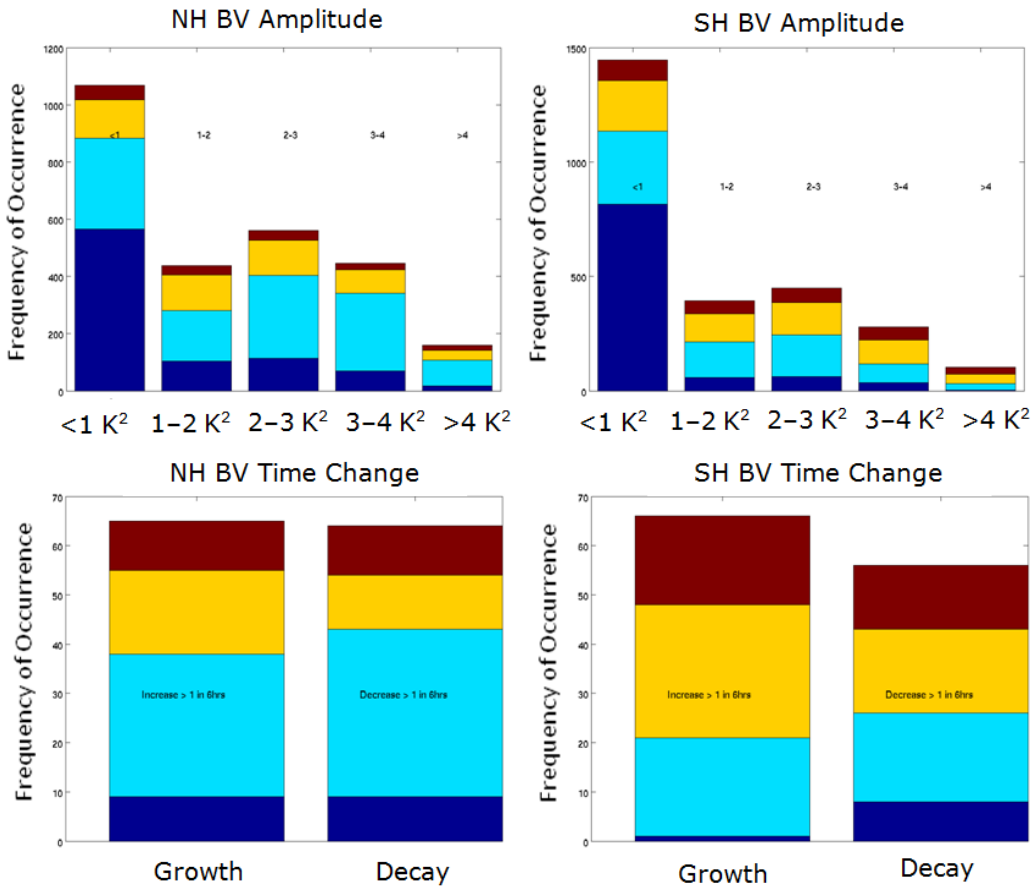


**Figure 2.6: Histograms of NH (left) and SH (right) BV zonal wavenumber (dark blue, wave 1; cyan, wave 2; yellow, wave 3; red, wave 4) at Level 25 (corresponds to Figure 2.5) for various BV amplitude bins (< 1 $K^2$, 1–2 $K^2$, 2–3 $K^2$, 3–4 $K^2$, and > 4 $K^2$; top row), and for times of rapid growth (by 1 $K^2$ in 0.25 sol; bottom row, left bars) and rapid decay (by 1 $K^2$ in 0.25 sol; bottom row, right bars) of instabilities.**

> 1 $K^2$ in 0.25 sol) and the most rapidly decaying waves (decrease in BV amplitude of 1 $K^2$ or more in 0.25 sol). In the NH, there is a larger percentage of wave 3 for growing waves than for decaying waves. In the SH, there are larger percentages of waves 2–4 for growing BV, and more wavenumber 1 for decaying waves. Therefore, wavenumber 3 and 4 regimes may be considered more unstable than wavenumber 1 regimes.

## *2.4 Diagnosis of Sources of Instability*

### *2.4.1 Barotropic and Baroclinic Instabilities*

Deriving an exact kinetic energy (KE) equation to the bred vectors can elucidate the origins of atmospheric instabilities. In classical instability studies (e.g., Orlanski and Katzfey, 1991), the basic flow is taken as a zonal average of the atmosphere, and does not satisfy the equations of motion. Here, we partition the perturbed run into the control (subscript *c*) and the bred vector (subscript *b*). By contrast, it is possible to derive exact KE equations for the BV since both the control and the perturbed run are both exact model solutions. This technique was pioneered by Hoffman et al. (2009) as applied to instabilities in the Earth's oceans.

The derivation of a kinetic energy tendency equation for the bred vector begins with the sigma-coordinate momentum equation

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v} \bullet \nabla \boldsymbol{v} + \dot{\sigma} \frac{\partial \boldsymbol{v}}{\partial \sigma} + f\boldsymbol{k} \times \boldsymbol{v} = -\nabla \Phi + \frac{\sigma}{p_s} \nabla p_s \frac{\partial \Phi}{\partial \sigma}, \qquad (2.1)$$

which is applied exactly to the control and perturbed atmospheric state. Here, $\boldsymbol{v}$ is "horizontal" velocity vector along sigma surfaces, $\dot{\sigma}$ is vertical motion, $f$ the Coriolis parameter, $\Phi$ geopotential height, and $p_s$ surface pressure. Taking the difference of the two equations creates a bred vector momentum equation:

$$\frac{\partial \boldsymbol{v}_b}{\partial t} + \boldsymbol{v} \bullet \nabla \boldsymbol{v}_b + \boldsymbol{v}_b \bullet \nabla \boldsymbol{v}_c + \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} + \dot{\sigma} \frac{\partial \boldsymbol{v}_b}{\partial \sigma} + f\boldsymbol{k} \times \boldsymbol{v}_b = -\nabla \Phi_b + \frac{\sigma}{p_s} \nabla p_s \frac{\partial \Phi}{\partial \sigma} - \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma}$$

(2.2) where bred vector (subscript *b*) equals perturbed run (no subscript) minus control run (subscript *c*). Finally, taking the dot product with the bred vector velocity, and manipulating terms, results in an equation that describes the change in kinetic energy of the BV, $K_b$, in an Eulerian framework.

$$\frac{\partial K_b}{\partial t}+\left[\mathbf{v}\bullet\nabla K_b+\dot{\sigma}\frac{\partial K_b}{\partial\sigma}\right]+\left[\nabla\bullet(\mathbf{v}_b\Phi_b)+\frac{\partial\dot{\sigma}_b\Phi_b}{\partial\sigma}\right]=-\left[\dot{\sigma}_b\alpha_b p_{sb}\right]-\left[\mathbf{v}_b\bullet\left((\mathbf{v}_b\bullet\nabla)\mathbf{v}_c+\dot{\sigma}_b\frac{\partial\mathbf{v}_c}{\partial\sigma}\right)\right]$$
$$-\frac{\Phi_b}{p_{sb}}\left(\frac{\partial p_{sb}}{\partial t}+\boldsymbol{v}_b\bullet\nabla p_{sb}\right)+\boldsymbol{v}_b\bullet\left(-\sigma\alpha\nabla p_s+\sigma_c\alpha_c\nabla p_{sc}\right) \qquad (2.3)$$

A detailed derivation of this equation can be found in Appendix A.

For this study, we focus on the baroclinic and barotropic conversion terms of the equation. The baroclinic conversion is $\left[\dot{\sigma}_b\alpha_b p_{sb}\right]$, where $\dot{\sigma}_b$ is the sigma coordinate vertical motion bred vector, $\alpha_b$ is the reciprocal of the density bred vector, and $p_{sb}$ is the surface pressure bred vector. This term refers to the conversion of available potential energy (due to thermal gradients) in the control solution to bred vector kinetic energy. The barotropic conversion is $\left[\mathbf{v}_b\bullet\left((\mathbf{v}_b\bullet\nabla)\mathbf{v}_c+\dot{\sigma}_b\frac{\partial\mathbf{v}_c}{\partial\sigma}\right)\right]$, with $\boldsymbol{v_b}$ referring to the bred vector velocity vector, $\boldsymbol{v_c}$ the control field velocity vector, $\sigma$ the vertical coordinate, and $\dot{\sigma}$ as before. This term refers to the conversion of kinetic energy from the control run into kinetic energy of the bred vector, and is typically found along sharp velocity gradients. Figure 2.7, which is based on the seminal diagram of Lorenz (1967), illustrates the transformation of energy between potential (P) and kinetic (K) energy bred vector (subscript *b*) and control (subscript *c*). As we have computed a budget for BV kinetic energy ($K_b$), the baroclinic term describes

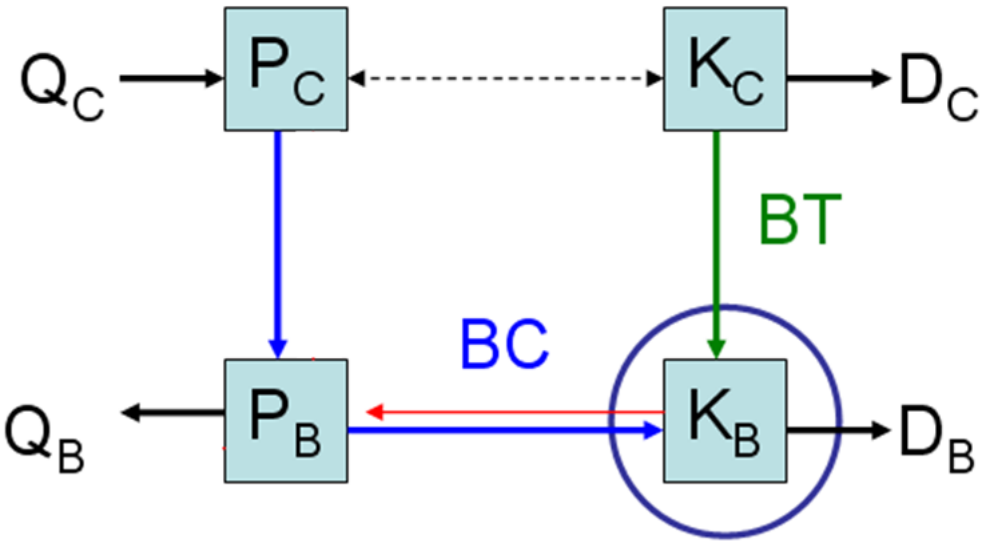


**Figure 2.7: Bred Vector energy diagram, based on Lorenz (1967), illustrating the transformation of energy among control potential energy ($P_c$), bred vector potential energy ($P_b$), control kinetic energy ($K_c$), and bred vector kinetic energy ($K_b$). $Q_c$ and $Q_b$ are the heating terms, and $D_c$ and $D_b$ the frictional dissipation terms for the control and bred vector, respectively. A budget analysis is performed for $K_b$ (circled), and the barotropic (BT; green) and baroclinic (BC; blue) conversion terms have been computed.**

energy conversion from $K_b$ to $P_b$, and the barotropic term describes energy conversion from $K_c$ to $K_b$.

Figure 2.8 shows a 10-sol time average zonal mean plot at the NH autumnal equinox ($L_s$ 180°) of temperature BV, $K_b$, and the baroclinic and barotropic KE tendency terms. The temperature BV is most prevalent in the lowest 10 km along the mid-latitude temperature gradient, and extends vertically along the polar vortex. The BV KE is also large in the low levels, extending vertically into the westerly jet stream (particularly in the SH and near the poles). There are three primary regions of bred vector kinetic energy conversion. The most prominent region is the cap-edge mid-latitude instabilities near the surface. Large positive values of the baroclinic conversion term are found in lower levels along the temperature gradients, in similar

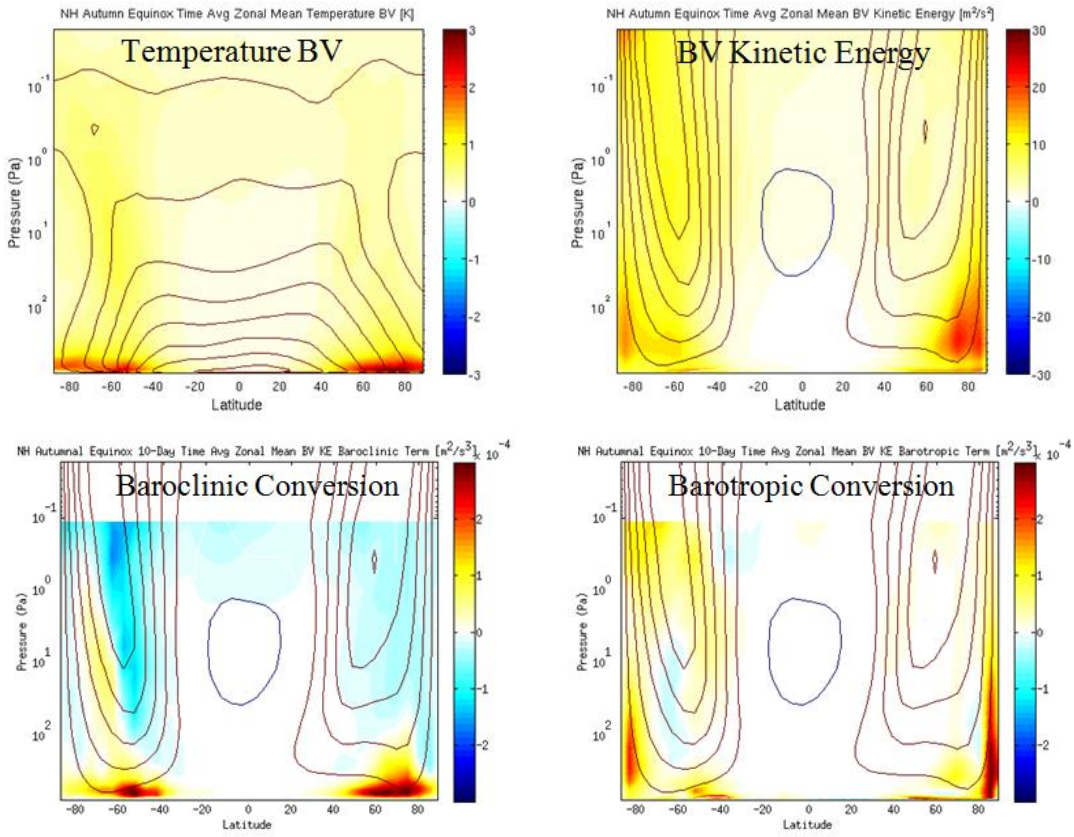


**Figure 2.8: Zonal mean BV quantities, plotted as a 10-sol mean starting at the NH autumnal equinox ($L_s$ 180°), Hour 00. Upper left: temperature BV (shaded) with control temperature (contours). Upper right: BV kinetic energy (shaded) with control u-component wind (contours). Lower left: baroclinic term of the BV KE tendency equation (shaded) with control u-component wind (contours). Lower right: barotropic term of the BV KE tendency equation (shaded) with control u-component wind (contours).**

locations to the temperature BV. Here, BV kinetic energy is generated baroclinically from potential energy (blue arrows in Figure 2.7). This is similar to the Earth, where baroclinic instabilities are present in regions with strong temperature gradients. The second region is the jet maxima aloft. This region is barotropically unstable, meaning that kinetic energy of the control run (from the winds in the jet stream) is transformed into bred vector kinetic energy (green arrow in Figure 2.7). These regions have a negative baroclinic term, meaning that the perturbation kinetic energy is subsequently lost to perturbation potential energy (red arrow in Figure 2.7). Although a full

explanation of this phenomenon requires a potential energy budget equation (which we plan to address in future work), we postulate that this energy is lost to radiative effects ($Q_c$ and $Q_b$ in Figure 2.7) that tend to drive the model back toward its control state (T. McConnochie, NASA Goddard, personal communication). Regions of barotropic conversion are found along the edges of the polar vortex, in particular near the poles, as well as within the lowest 1 km of the surface. Here, kinetic energy of the control run is being converted to kinetic energy of the BV, enhancing the magnitude of the wind difference between the two runs. The final energy conversion region is located poleward of the SH jet core, and is barotropically stable and baroclinically unstable (opposite of region 2). Potential energy is converted to bred vector kinetic energy, which eventually adds to the kinetic energy of the control. This dipole structure of regions 2 and 3 is found symmetrically in the NH (not shown) as solstice season approaches.

Figure 2.9 shows the same quantities as Figure 2.8, but for a cross section through 345° degrees latitude for a 2-sol average at the NH autumnal equinox. According to Figure 2.4, this intersects BV with zonal wave 2 near the surface, and zonal wave 1 through the upper atmosphere. The cross section reveals that these waves are largely distinct (with opposite sign BVs), with the near surface BV growing from baroclinic conversion, and the upper atmosphere wave benefiting from barotropic conversion.

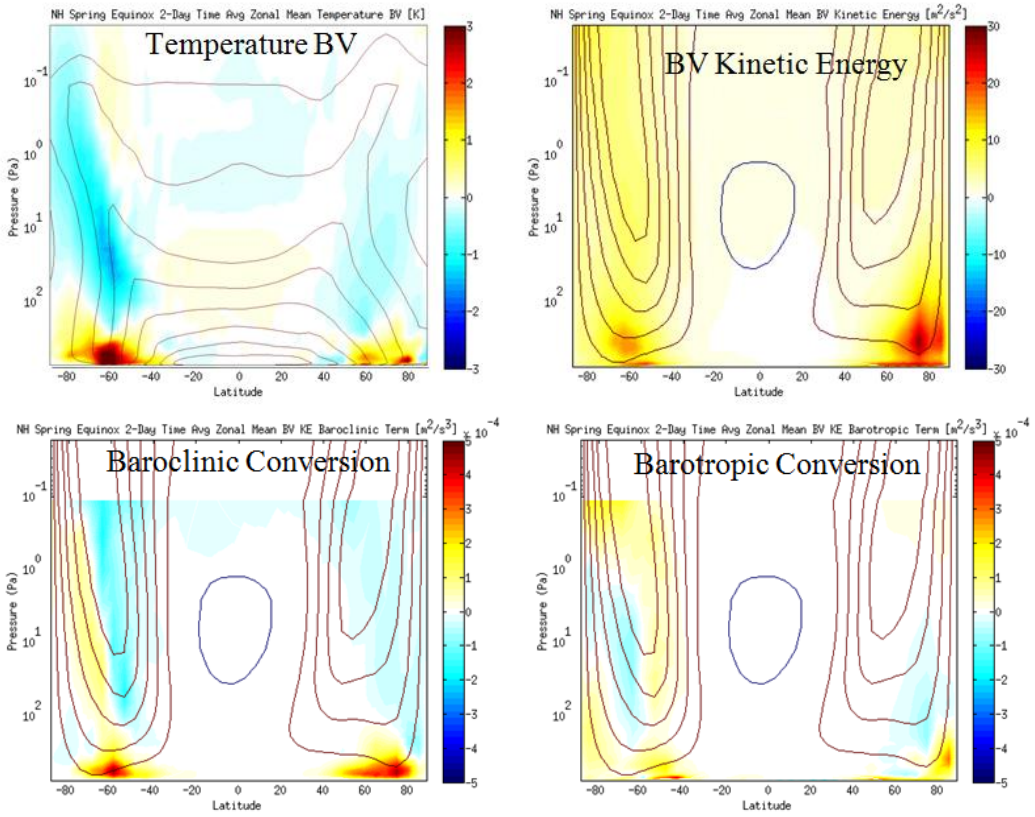


**Figure 2.9: Same quantities as Figure 2.8, but for a cross section through longitude 345°, plotted as a 2-sol mean starting at the NH autumnal equinox ($L_s$ 180°), Hour 00. Top left panel is time mean signed bred vector rather than magnitude.**

### *2.4.2 Role of Topography: Lee Cyclogenesis*

Figure 2.10 shows the time averaged magnitude of temperature BV activity at ~3 km (Level 20) for NH Late Autumn ($L_s$ 180°–240°). Because the field is time averaged over a period of ~100 sols, the impact of the arbitrary positions of peaks and troughs of traveling waves should be smoothed out, and one might expect a zonally symmetric plot. However, there is a strong asymmetry with latitude in the NH, with three peaks in BV activity at longitudes 80°, 180°, and 320°. There are similar peaks in the SH mid-latitudes, particularly around 65° S latitude and 290° longitude. Comparing these areas to topography, these correspond to locations downstream of

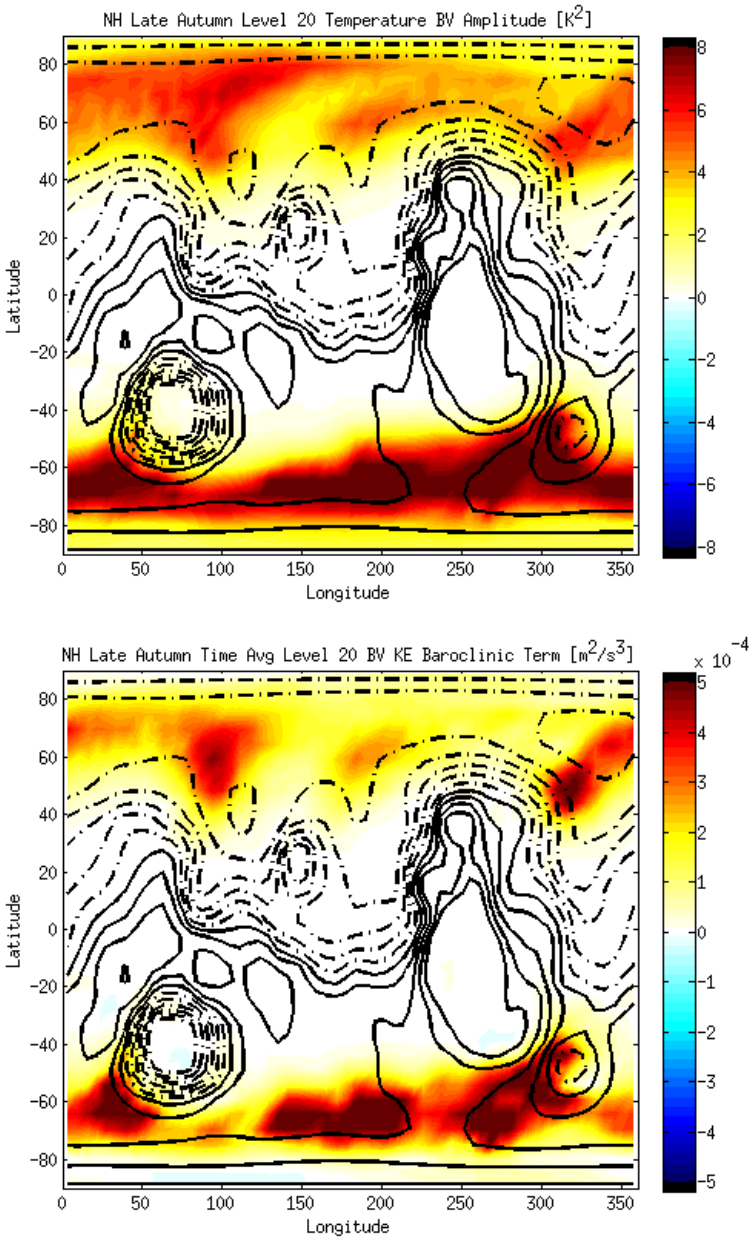


**Figure 2.10: NH late autumn ($L_s$ 180°–240°) time averaged temperature BV amplitude ($K^2$, shaded, top) and BV KE baroclinic term ($m^2\ s^{-3}$, shaded, bottom) near 3 km altitude (Level 20), plotted with surface topography (positive: solid contours, negative: dashed contours).**

elevated terrain. Differences in elevation, even on the coarse MGCM grid, are on the order of several kilometers. The increased instability in these regions is therefore likely the result of lee cyclogenesis. An eastward-propagating column of air must

conserve its angular momentum, and also its potential vorticity. As the column crosses elevated terrain, it is vertically compressed. It then expands vertically downstream of the mountains, which generates cyclonic relative vorticity. In a region of baroclinic instability, this vertical stretching can trigger low level cyclogenesis that focuses the instabilities into these regions downstream of topography.

### *2.5 Breeding in an Assimilation System*

Bred vectors are also related to data assimilation and forecasting, as they are analogous to unstable forecast errors arising from uncertain initial conditions that are amplified by the dynamics of the model. To illustrate this link, we perform a breeding experiment in the context of the analysis-forecast cycle, similar to that used by terrestrial operational numerical weather prediction centers worldwide. Observations are combined with a background, or first guess, field to create an analysis field. The analysis is then advanced in time (here, 0.25 of a sol) using an atmospheric model to create a new background state for the next assimilation stage. In ensemble data assimilation systems, uncertainty in the analysis and forecast is described by the diversity of ensemble members, while the ensemble mean is the best estimate of the true state. Here, Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES) temperature profiles (Smith et al., 2001) are assimilated by the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007). The LETKF configuration for Mars is described in detail in Hoffman et al. (2010) and Chapter 3.

For the control run of the forecast cycle breeding experiments, the nature run of the MGCM is replaced by the analyses and forecasts from data assimilation. Therefore, the initial control run state is the initial analysis from data assimilation, to

which random perturbations are added to create the perturbed run. After using the model to advance the control run and perturbed run forward in time 0.25 sol, the bred vector amplitude is calculated from the difference between the perturbed run and the 6-hour forecast (or background) from the assimilation. The rescaled bred vector is then added to the new data assimilation analysis (rather than the forecast) to create a new perturbed state, and the process is repeated to create a series of bred vectors throughout the assimilation experiment.

In Figure 2.11, bred vectors arising from the assimilation experiment are compared with forecast ensemble spread, which is a measure of uncertainty in the estimate of the state of the atmosphere. In the lower atmosphere, the largest ensemble spread, as well as the bred vectors, is primarily located in mid-latitude belts of strong temperature gradients. The top row of Figure 2.11 shows an example of good correspondence between bred vector amplitude and ensemble spread at model Level 25 (approximately 3 km altitude); local maxima are associated with traveling waves, and are found at 120°, 200°, and 330° longitude in the NH and 30°, 180°, and 270° in the SH. The bottom row of Figure 2.11 compares the zonal mean structure of bred vectors with temperature ensemble spread. The size of the bred vectors is determined largely by the choice of initial perturbation amplitude; by selecting a larger value, one could obtain a closer agreement between the magnitudes of the bred vectors and ensemble spread. The spatial patterns of the two plots have great similarity; large instabilities / ensemble spread are present near the surface in the mid-latitudes of each hemisphere, whereas the low level tropics experience a minimum in spread and BV amplitude. Chapter 3 discusses how this region of low spread is an important

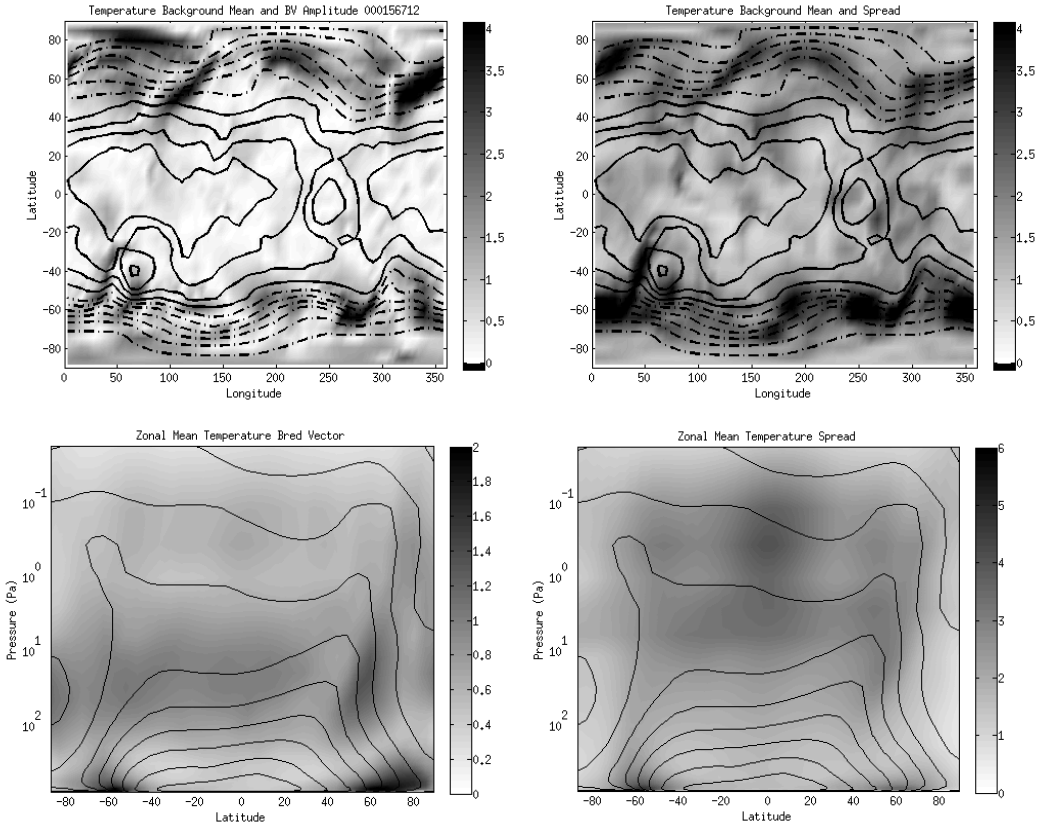


**Figure 2.11: Comparison of bred vectors (left panels) with ensemble spread from data assimilation (right panels) just after the NH autumnal equinox ($L_s$ 180°) of MY 24. The top row shows temperature bred vector [K] using assimilation analyses as a control run (left) and temperature analysis ensemble spread [K] (right) at model Level 20, or approximately 3 km altitude, at a snapshot in time ($L_s$ 186.7°). The bottom row shows zonal mean temperature BV amplitude (left) and ensemble spread (right) as a time mean over 30 sol ($L_s$ 180.5°–198.4°). Contours are analysis temperature [K].**

consideration in developing the data assimilation system. In this stable region, uncertainty in the background / forecast field can be underestimated, and varying the dust distribution among members of the ensemble is very beneficial in increasing ensemble spread and improving the quality of analyses. Local maxima in the two panels also extend vertically along the mid-latitude temperature gradient, as well as into the tropics above 10 Pa. Interestingly, the bred vectors from the assimilation contain a local maximum in the tropics between 100 and 10 Pa that is not present in

breeding based on the nature run of the model. It is important to note that TES observations extend vertically from near the surface to only around 10 Pa. Therefore, the attractor of the analyses / observations may be more unstable in this region than the attractor of the freely running model. Ensemble spread is also comparatively larger above 10 Pa, as observation information reduces uncertainty in the temperature field, and hence analysis and forecast spread, in the lower levels of the atmosphere.

## *2.6 Conclusions*

The bred vector technique has demonstrated its power in illuminating the instabilities of the Martian atmosphere. Bred vector activity (i.e., instability of the basic flow) is greatest during the transition seasons (early spring and late autumn) of each hemisphere, as well as during the NH winter. There is a solstitial pause, with weaker traveling wave activity and instabilities around the solstice in the winter hemisphere, which has also been demonstrated in the Oxford GCM (Read et al., 2011). During the NH summer there is an interval of time ($L_s$ 90°–115°) where BV activity is minimal throughout the planet. This period of time also experiences negative growth rates in Newman et al. (2002), although in that study the quiescent period lasts throughout much of the time between the NH spring and NH autumn equinoxes. There is also a strong link between transient eddies and waves of bred vectors, with strong instability growths preceding the seasonal initiation of traveling wave activity in the lower atmosphere.

The application of a kinetic energy equation to Mars bred vectors attributes strong near surface instabilities in the mid-latitudes to baroclinic energy conversions, whereas both baroclinic and barotropic conversion play a role in polar and upper

atmosphere instabilities. Locations downstream of elevated topography are particularly favorable for bred vector growth because of lee cyclogenesis, focusing baroclinic instability into regions of vertical stretching.

Finally, bred vectors generated from the analysis-forecast cycle of an ensemble data assimilation system show good agreement in spatial patterns with forecast ensemble spread. This similarity demonstrates the link between atmospheric instabilities as estimated with BVs and forecast errors and uncertainty.

In future experiments, the assumption of a fixed dust distribution will be relaxed (See Section 4.4). Newman et al. (2002) explored the impact of various dust scenarios (low dust, prescribed seasonally evolving dust, and high dust distributions) on bred vector amplitude, and showed that BV growth rates generally tend to increase with dust opacity, particularly in the most active seasons. In general, the bred vector growth rates for this study were somewhat smaller than those in Newman et al. (2002) and Rogberg et al. (2010), which both employ the UK MGCM. Neither study has fully explored breeding with a fully interactive dust response, including lifting; given the feedbacks between the dust and temperature field it is anticipated that instabilities would be even larger in such a scenario, and potentially could play an important role in the development and growth of dust storms. Such feedbacks would also motivate a closer examination of BV potential energy and its sources, including diabatic heating.

# Chapter 3: Ensemble Kalman Filter Data Assimilation of Thermal Emission Spectrometer (TES) Profiles into a Mars Global Circulation Model

Thermal Emission Spectrometer (TES) temperature profiles are assimilated into a Mars Global Circulation Model (MGCM) using the Local Ensemble Transform Kalman Filter (LETKF). Comparing short-term (0.25 sol) forecasts to independent observations reveals a clear improvement over the free running model. Additional enhancements, including adaptive inflation, varying the dust distributions among ensemble members, and an empirical bias correction technique, lead to further reductions in both RMSE and bias, and an improved estimate of the state of the atmosphere and its uncertainty. After bias correction, the largest errors are located in the mid-latitude baroclinic zone near the southern hemisphere ice cap edge, and are generally less than observation error elsewhere. Longer term forecasts show skill through 3–5 sols when compared to the freely running model. The greatest impediment to performance is bias between the model and observations, which we plan to ameliorate through the assimilation of dust opacities, as well as by incorporating improvements to the representation of dust and water ice clouds in the MGCM.

## *3.1 Introduction to Mars Data Assimilation*

Recent spacecraft missions have contributed a wealth of data to enhance our understanding of Martian weather and climate. In particular, retrievals from the Thermal Emission Spectrometer (TES; Smith et al., 2001), and now from the Mars

Climate Sounder (MCS; McCleese et al., 2007), provide a multi-year record of atmospheric temperatures and dust conditions. However, these observations are constrained by the orbit of the spacecraft, and do not provide full fields or information on all atmospheric variables, especially winds and surface pressure. Data assimilation provides the optimal framework for combining observations that are irregular in space and time, with a background state based on a forecast from a global circulation model, to create an analysis which best represents the state of the atmosphere at a given time.

Scientists at Oxford University performed the first complete and comprehensive reanalysis of the TES era (Lewis et al., 2007). This "UK Reanalysis" has been compared to independent radio science profiles (Hinson et al., 1999) by Montabone et al. (2006) and has led to improvements in the understanding of atmospheric tides (Lewis and Barker, 2005), the variability of dust storms (Montabone et al., 2005), tropical water ice clouds (Wilson et al., 2008), and predictability (Rogberg et al., 2010). The UK Reanalysis used the analysis correction scheme (Lorenc et al., 1991), where model fields are repeatedly nudged toward observations, a technique which was employed operationally for the Earth's atmosphere in the UK Meteorological Office in the 1990s. Since that time, there have been several key advances in data assimilation systems for meteorology that have led to marked improvements in Numerical Weather Prediction (NWP) skill, including the emergence and maturation of ensemble-based methods (e.g., Kalnay, 2010).

Our Mars data assimilation system employs the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007), a state of the art system that has been

applied successfully for terrestrial atmospheric data assimilation (Miyoshi and Yamane, 2007; Szunyogh et al., 2008). The advantages of ensemble-based methods include an estimate of analysis uncertainty at each synoptic time as depicted by the spread of the ensemble, thereby providing flow-dependent background error covariances, which determine the strength and impact pattern that an observation has on the analysis. The LETKF also has a multitude of properties that can aid assimilation studies, including the ability to estimate observation error (Li et al., 2009).

M. Hoffman et al. (2010) describe a test of our LETKF assimilation system with the GFDL Mars Global Circulation Model (MGCM; Wilson et al., 2002) and simulated observations. In an Observing System Simulation Experiment (OSSE), the "truth" is a known model trajectory, and synthetic observations are obtained by sampling the truth and adding a random error term at the locations, times, and vertical spacing of actual TES observations. The assimilation system, which has no knowledge of the truth except through the imperfect observations, produces analyses of temperature and wind, which are then compared to the true state. Using only temperature observations at TES locations with a 3 K observation error, analysis errors were less than 1 K for temperature and less than 5 m/s for wind speed. The system converged to its optimal performance within 5 sols.

In this chapter, we assimilate actual TES temperature profiles using the MGCM and the LETKF for a 30-sol period during the NH autumn of Mars Year (MY) 25. Section 3.2 provides background on the TES observations, MGCM, and LETKF. Section 3.3 outlines the specific configuration of our assimilation system

and experiments. Section 3.4 provides an evaluation of the system by comparing both short (0.25 sol) and longer range (1–5 sol) forecasts to independent (in time) TES observations, and Section 3.5 summarizes and indicates areas of future work. The eventual goal of this project is the creation of a new Mars atmosphere weather and climate reanalysis of atmospheric and aerosol fields spanning several Martian years.

## *3.2 Data and Methods*

### *3.2.1 Thermal Emission Spectrometer (TES) Profiles*

The Mars Global Surveyor (MGS) spacecraft arrived in Martian orbit in September 1997, and, following a period of aerobraking, began its primary science mission in February 1999, which continued until the loss of communication with the spacecraft in November 2006 (Figure 1.4). During the mapping phase, MGS completed 12 orbits each sol, passing over the surface in bands 30 degrees longitude apart at 0200 and 1400 local time. From 1997 to 2004, or Mars Year (MY) 24 to early MY 27, the Thermal Emission Spectrometer (TES; Christensen et al., 1992) provided hundreds of millions of radiance measurements of the Martian surface and atmosphere. Nadir, or downward-pointing, measurements have a surface footprint size of 3 × 3 km. Radiances in the 15 micron $CO_2$ absorption band are used to infer atmospheric temperature up to around 40 km altitude (0.1 hPa), with an effective vertical resolution of one scale height (10 km) (Conrath 2000).

Smith et al. (2001) describes the temperature profiles, column dust opacity, water ice opacity, and surface brightness temperatures retrieved from TES, which are publicly available in the Planetary Data System (PDS) online archive, and were used

for this study (Figure 3.1). Errors in the temperature profile are estimated to be around 2 K, with higher errors possible near the surface and in polar regions. There are also systematic errors that vary orbit-to-orbit (T. McConnochie, personal communication). Eluszkiewicz et al. (2008) note that the PDS TES profiles do not account for differing surface emissivities over the polar ice caps, which limits the accuracy of near surface temperatures in polar regions. Retrieval errors of both temperature and dust are larger during dust storms (high opacities) where the assumption that the aerosols are non-scattering is less accurate (Smith et al., 2001). Dust opacities are also not available when surface temperatures fall below 220 K, as there is insufficient thermal contrast between the surface and the atmosphere for an accurate retrieval.

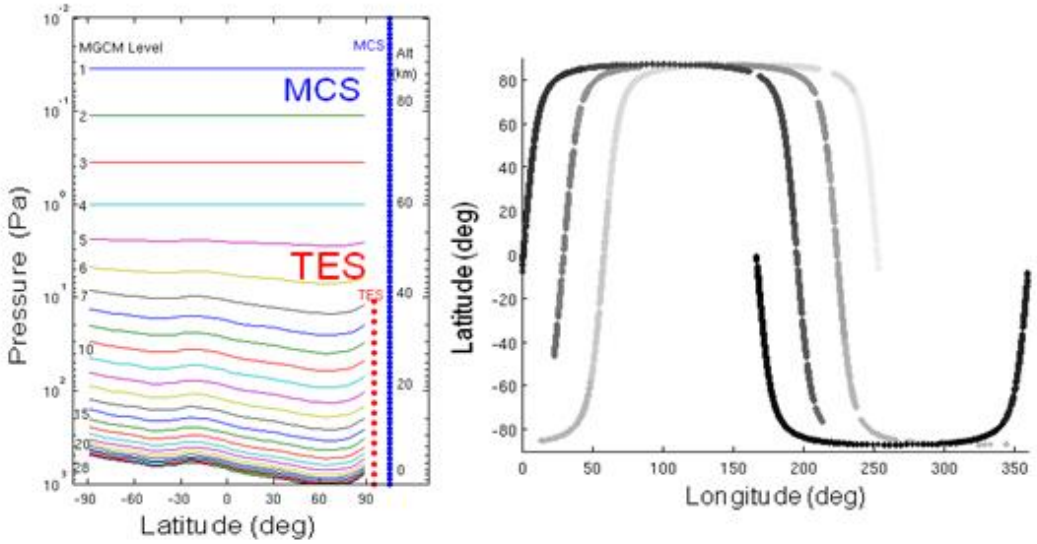


**Figure 3.1: Left panel: vertical coverage of TES and MCS retrievals relative to MGCM vertical levels. Right panel: sample horizontal coverage of TES observations in a 0.25 sol window. Shading represents the passage of time (successive orbits).**

### *3.2.2 Mars Global Circulation Model (MGCM)*

The GFDL Mars General Circulation Model (MGCM) has been used in a multitude of atmospheric studies, from tides and planetary waves (Wilson and Hamilton, 1996; Hinson and Wilson, 2002; Wilson et al., 2002; Hinson et al., 2003), the water cycle (Richardson and Wilson, 2002; Richardson et al., 2002), and the dust cycle (Basu et al., 2004), to cloud radiative effects (Hinson and Wilson, 2004; Wilson et al., 2008). The model, originally based on the GFDL SKYHI terrestrial GCM, is now adapted to the GFDL Flexible Modeling System (FMS) with a finite volume dynamical core (Lin, 2004). Model physics include a budget for gaseous and condensed $CO_2$, surface and subsurface physics including a soil model, mass-conserving inventories of water and dust aerosol, and options for dust lifting and transport.

The model configurations used in this chapter are essentially the same as employed for the assimilation experiments in M. Hoffman et al. (2010), as well as described in Section 1.4. We use a latitude-longitude horizontal grid with $5 \times 6$ degree resolution, and 28 vertical levels with a hybrid pressure-sigma vertical coordinate system. Our choice of dust parameters is elaborated upon in Section 3.3.

### *3.2.3 Local Ensemble Transform Kalman Filter (LETKF)*

The data assimilation cycle consists of two phases: the analysis stage and the forecast stage (Figure 3.2). During the analysis stage the data assimilation system creates an analysis, or the best estimate of the state of the atmosphere, by optimally combining information from both observations and the background. The background, or first guess, is usually a short-term model forecast from the previous analysis.

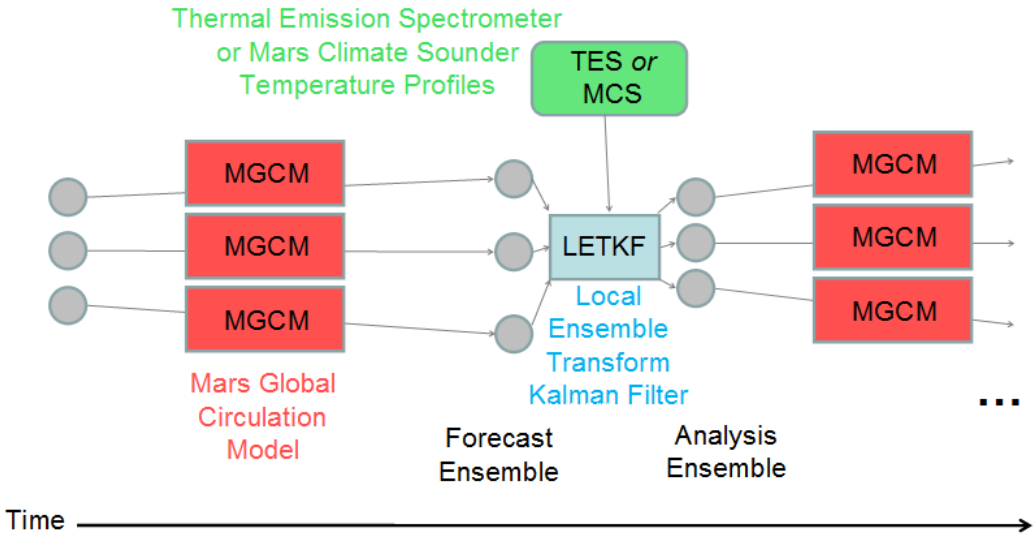


**Figure 3.2: Schematic illustrating the components of ensemble data assimilation for Mars: the Mars Global Circulation Model (MGCM), Thermal Emission Spectrometer (TES) observations, and the Local Ensemble Transform Kalman Filter (LETKF) assimilation system.**

During the forecast stage, the analysis is then advanced in time using the forecast model (MGCM) to provide a background state for the next analysis (0.25 sol in the future). The uncertainty estimates for the observations and background determine the optimal weighting of each source of information when producing the analysis. The background error covariance (denoted **B** here) also determines the spatial impact pattern of an observation on the analysis, as well as how observations of one variable type influence model fields of another variable type (e.g., how temperature observations update the wind field).

Ensemble Kalman Filters (EnKF; Evensen 1994) describe the state of the atmosphere and its uncertainty as an ensemble of plausible states, where the ensemble mean represents the best estimate of the true state. Therefore, during the forecast stage, each analysis ensemble member is advanced in time by the forecast model (MGCM) in parallel (Figure 3.2) to produce the forecast ensemble, or background

ensemble. These ensemble members, in addition to having different atmospheric states (temperature, wind, surface pressure), can have different model parameters, such as dust and water ice cloud properties, which span the range of our uncertainty about these processes. Larger ensemble spread (standard deviation from the ensemble mean) represents a greater uncertainty in the ensemble. A significant advantage of ensemble data assimilation techniques (e.g., Kalnay et al., 2007) is that the background error covariance is determined from the ensemble at each analysis time, and is therefore flow dependent. It is also straightforward to determine correlations between different model variable types, such as between atmospheric state variables and aerosol fields.

The Local Ensemble Transform Kalman Filter (LETKF) was developed by Hunt et al. (2007) as an efficient formulation of the Ensemble Kalman Filter (EnKF). The LETKF has been adapted to operational numerical weather prediction (NWP) models for the terrestrial atmosphere, including the Global Forecasting System (Szunyogh et al., 2008) and the Japan Meteorological Agency model (Miyoshi et al., 2010).

Mathematically, the analysis vector $\mathbf{x}_a$ is calculated through optimal combination of forecast $\mathbf{x}_b$ and observations $\mathbf{y}_o$:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\,(\mathbf{y}_o - h_{op}(\mathbf{x}_b))\,. \qquad (3.1)$$

The optimal weight matrix $\mathbf{K}$, or Kalman Gain, is given by

$$\mathbf{K} = \mathbf{B}\,\mathbf{H}^T\,(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \qquad (3.2)$$

where **B** is the background error covariance matrix, **R** the observation error covariance matrix, and **H** the linearization of the observation operator $h_{op}$ which interpolates and transforms model variables into observation variables.

The LETKF algorithm can be described as follows, following Kalnay (2010). We begin with a background ensemble of model states, where each of *K* members can be described as a vector $\mathbf{x_b}^{\mathbf{k}}$ (containing information over all grid points and variable types), and whose ensemble mean is $\overline{\mathbf{x}}_\mathbf{b}$. We can then describe a matrix of ensemble perturbations $\mathbf{X}_\mathrm{b}$, where column *k* represents the difference between model state $\mathbf{x_b}^{\mathbf{k}}$ and the ensemble mean $\overline{\mathbf{x}}_\mathbf{b}$. The background error covariance is then expressed as

$$\mathbf{B} = \frac{1}{K-1}\mathbf{X}_b\mathbf{X}_b{}^T. \qquad (3.3)$$

As **B** is expressed in terms of the ensemble perturbations, the background error covariance evolves with the atmospheric flow. The observation operator $h_{op}$ converts model variables into observation variables, including interpolation to observation locations. An ensemble member in observation space (that is, interpolated to observation locations and transformed to observation variables) can be written as $\mathbf{y}_\mathrm{b}{}^\mathrm{k}$ = $h_{op}$ ($\mathbf{x}_\mathrm{b}{}^\mathrm{k}$). The background error in observation space is thus $\mathbf{Y}_\mathrm{b}$, with column *k* as $\mathbf{y}_\mathrm{b}{}^\mathrm{k} - \overline{\mathbf{y}}_\mathbf{b}$, with $\overline{\mathbf{y}}_\mathbf{b}$ being the mean of the ensemble in observation space.

The analysis is performed locally at each grid point. Localization has several benefits, including reducing sampling error from spurious long-distance correlations arising from small ensemble size, taking advantage of the low dimensionality of atmospheric errors in local regions (Patil et al., 2001), and significantly reducing computational cost. At each grid point we select the local observations that will

impact that grid point, specifically those within 3.65 times the horizontal (and vertical) localization distance $L$ (Miyoshi et al., 2007). Rather than abruptly terminating the observation influence at a cutoff distance, it is decreased smoothly by increasing its error using the inverse of a Gaussian function

$$f_{Rloc} = \exp\left[\frac{+d(i,j)^2}{2L^2}\right] \qquad (3.4)$$

with localization length scale $L$, where $d$ is the distance between observation $i$ and model grid point $j$ (Hunt et al., 2007). This localization is accomplished by multiplying the observation standard deviation (in the observation error covariance matrix **R**) by $f_{\mathrm{Rloc}}$ in a procedure referred to as "R-localization" by contrast to localizations performed on the background covariance inflation **B** (Hamill et al., 2001; Houtekamer and Mitchell, 2001). Greybush et al. (2011) examined R-localization using the LETKF, and found that the geostrophic balance between the height and wind fields in LETKF analysis increments is comparable to that of other ensemble Kalman filters using B-localization, with similar performance in terms of analysis accuracy. Appendix B elaborates upon these results.

If $K$ is the ensemble size, typically of the order 15–100, the analysis error covariance in ensemble space ($K \times K$) is given as

$$\hat{\mathbf{P}}_{\mathbf{a}} = \left[(K-1)\mathbf{I} + \mathbf{Y_b}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{Y_b}\right]^{-1}, \qquad (3.5)$$

where $\mathbf{Y_b}$ refers to the local background error covariances in observation space, **R** to the local observation error covariance matrix, and **I** is the $K \times K$ identity matrix. Computing this quantity in ensemble space allows the LETKF to be computationally

efficient for large NWP applications. The local weight vector for the ensemble mean is

$$\overline{\mathbf{w}}_a = \hat{\mathbf{P}}_\mathbf{a}(\mathbf{Y_b})^\mathbf{T}\mathbf{R}^{-1}(\mathbf{y_0} - \overline{\mathbf{y}}_\mathbf{b}), \qquad (3.6)$$

and

$$\mathbf{W_{ap}} = \left[(K-1)\hat{\mathbf{P}}_\mathbf{a}\right]^{1/2} \qquad (3.7)$$

is the weight matrix for the ensemble perturbations. Adding $\overline{\mathbf{w}}_a$ to each column of $\mathbf{W_{ap}}$ yields the full weight matrix $\mathbf{W_a}$. We then use this matrix to transform the background ensemble to the analysis ensemble

$$\mathbf{X_a} = \mathbf{X_b}\mathbf{W_a} + \overline{\mathbf{x}}_\mathbf{b}. \qquad (3.8)$$

Equation (3.8) shows that the analysis increment, or update, is a linear combination of background ensemble perturbations so that the LETKF analysis weights describe the relative contribution of each ensemble member to the analysis. This information can be helpful in inferring which choice of physical parameters in the model best matches the observations.

Ensemble Kalman filters require increasing the ensemble spread (known as inflation) because the uncertainty of the ensemble tends to be underestimated in the presence of model deficiencies and nonlinearities. This procedure, known as inflation, is described in more detail in Section 3.3.4.

## *3.3 Experiment Design*

### *3.3.1 Experiment Duration*

Assimilation experiments of TES profiles are conducted for a 30-sol time period during the NH Martian autumn, approximately $L_s$ (solar longitude) 164.9°–

181.9°, or sols 530–560 past the perihelion. The data are from Mars Year (MY) 25 prior to the onset of the 2001 global dust storm (Cantor et al., 2001). During this season, traveling wave activity is present in both hemispheres in the mid-latitudes, and dust loading is light to moderate.

### *3.3.2 Observation and Analysis Variables*

For the experiments described in this chapter, we assimilate vertical TES temperature profiles, which we consider as 19 individual point observations. The temperature observations are used to update temperature *T*, zonal wind *u*, meridional wind *v*, and surface pressure *ps* model variables in the analysis.

Observations that are very close together geographically are combined into "super-observations." This helps to improve the representativeness of observations, as well as to reduce computation time and the random components of the observation error. Here, observations are assigned to the nearest model grid point, and the average latitude, longitude, and observation value for observations assigned to each grid point are used in the assimilation. With super-observations, order $10^5$ profiles are combined to result in approximately 500 profiles, or $10^5$ individual temperatures, that are assimilated each 6-hour window. We only assimilate observations with good quality control flags as indicated by the Planetary Data System (PDS). Quality control also occurs within the LETKF, where observations are rejected if they differ from the background by more than the QC threshold (in this case, more than 5 times the prescribed observation error, or 15 K).

The method of Li et al. (2009) permits the online estimation of observation errors within the LETKF. Here, we estimated the observation errors to be 2.7 K. As

there are also vertical error correlations not yet explicitly considered in the LETKF, we assign an observation error of 3 K to all TES temperature observations.

R. Hoffman (2010) developed a strategy for eliminating vertical observation error correlations, as well as the influence of the retrieval prior, but this requires the knowledge of the Averaging Kernels of the retrievals (Rodgers, 2000; R. Hoffman, 2010) that are not available in the PDS TES retrievals. We plan to use this method with an Optimal Spectral Sampling (OSS; Moncet et al., 2008) forward operator for TES (Eluszkiewicz et al., 2008) and compare it with the direct assimilation of radiances in future work.

### *3.3.3 LETKF Configuration*

Analyses are produced every 0.25 sols, or 6 hours. Here we use 4-dimensional LETKF, which compares observations to the background at 1-hour intervals (see Figure 3.3). For example, for a forecast started at time t=0 hours, observations are binned into intervals from t=3–3.5 hours, t=3.5–4.5 hours, t=4.5–5.5

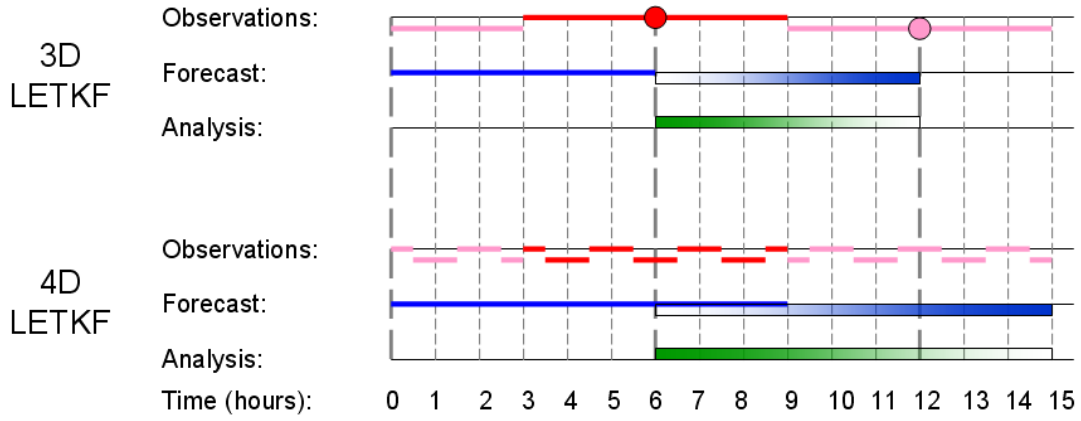


**Figure 3.3: Schematic illustrating the 4d-LETKF, where observations (binned together as red line segments for the first 6 hour assimilation cycle) are compared to forecasts (solid blue line) at the correct hour (3–9) rather than just to the 6 hour forecast as in 3d-LETKF.**

hours, and so forth through t=8.5–9 hours. These observation files are then compared to forecasts at t=3, 4, 5, 6, 7, 8, and 9 hours to produce an analysis valid at t=6 hours. Assimilating observations at their time of occurrence produces more realistic observation increments, particularly as there are large diurnal temperature changes on Mars.

Most experiments employ 600 km localization in the horizontal, 0.4 log P localization in the vertical, and 3 hour time localization. Section 3.4.1 explores the sensitivity of the results to the choice of horizontal localization distance.

### *3.3.4 Ensemble Selection and Inflation*

As ensemble perturbations represent the forecast error, and determine the local subspace of analysis corrections, it is important to develop and maintain a good ensemble for LETKF data assimilation. Initial ensemble members are taken from previous states of a nature run of the MGCM at 0.25 sol intervals. The initial perturbations project onto the unstable atmospheric modes, or bred vectors, which grow in time. Chapter 2 examines predictability using bred vectors, and shows a good correspondence between unstable regions of the atmosphere as revealed by bred vectors, and the ensemble spread obtained from data assimilation. Here, we use 16 ensemble members. An experiment with 32 members did not show a significant increase in performance, although a larger ensemble size may prove beneficial in future experiments to sample more sophisticated representations of dust and water ice clouds in the model. The baseline assimilation configuration uses ensemble members with the same dust field: spatially and temporally homogeneous 0.3 opacity when normalized to 6.1 hPa following the Conrath (1975) vertical profile. However,

Section 3.4 demonstrates that this leads to overconfidence in the model uncertainty estimates of temperature, particularly in tropical regions. Therefore, an ensemble with varying dust opacities (increasing linearly from 0.2 to 0.5) among the members is employed for most subsequent experiments. The next step will be using dust opacities that are constrained both by observations, and by aerosol transport in the MGCM.

The uncertainty of an analysis is less than that of the background, as it benefits from observation information. However, in EnKF systems, the analysis and forecast ensemble spread is typically underestimated, due to neglect of model error and other factors, and this overconfidence in the forecast can result in filter divergence whereby the observations are essentially ignored. Inflation increases the ensemble spread at each assimilation time. The initial multiplicative inflation value is set to 10%, meaning that the ensemble perturbations are multiplied by 1.1 during the assimilation step. For levels above ~0.1 hPa, there is no observation information available to constrain the ensemble spread, so inflation is not needed. There is a linear decrease in inflation values from 10% at model Level 7 (top of TES observations) to 0% at model Level 4.

Adaptive inflation allows the inflation value to be estimated online, rather than manually tuned. Li et al. (2009) provide a means for estimating a globally uniform value of multiplicative inflation, but on Mars there are great spatial variations in ensemble spread, and thus a need for spatially varying adaptive inflation. Miyoshi (2011) uses statistics of observation increments (observation minus forecast) to estimate multiplicative inflation values at each grid point via a Kalman filter,

which serves as a temporal smoother. We implemented Miyoshi (2011)'s adaptive inflation, and found the inflation value error variance of 0.08 to be effective in arriving at reasonable inflation values during the 30-sol assimilation, with significantly improved results compared with the use of the standard fixed inflation approach. Using temperature observation increments, we estimate an inflation value and apply it to temperature, u-wind, and v-wind model fields.

### *3.3.5 Adaptive Bias Correction*

As the results in Section 3.4 reveal, a significant portion of the temperature analysis error is due to bias rather than random error. We therefore employ an empirical bias correction method based on Danforth et al., 2007. External errors (as opposed to internal errors due to uncertain initial conditions and model instabilities) are due to the fact that the model is imperfect, for example in the subgrid scale parameterizations and radiation schemes. For Mars models, assumptions must be made about dust and water ice aerosol distributions, particle sizes, and optical properties. While insights from data assimilation can eventually lead to improvements in model parameterizations, a technique that can approximate the missing model forcing by treating the model as a black box and considering only forecast error statistics can be very helpful in improving assimilation performance. Here, we estimate model bias (in temperature, wind, and surface pressure) as the mean analysis increment from data assimilation over a period of 10–20 sols from the start of assimilation, or sols 540–550 past perihelion. For the purpose of reanalysis, we can then repeat the assimilation for that time period, adding the empirical bias correction term to each ensemble member prior to computing the analysis step. In a

forecasting setting, we can use the bias correction field computed from the previous *N* sols to correct the forecast at each future time step. (For example, we correct the bias for sols 550–560 based on bias fields from sols 540–550.) This technique can be further improved by considering separate bias correction fields for different times of day, or under different flow regimes or dust conditions. Ultimately, these bias correction fields also provide insights to improve model parameterizations.

## *3.4 TES Assimilation and Forecast Results*

### *3.4.1 Evaluation and Short Term Forecasts*

We now evaluate the Mars data assimilation system described in the previous sections. As a performance benchmark, we create a free run forecast of the MGCM that has no knowledge of the observations, and is therefore representative of the model climatology. This forecast was initialized from the initial background ensemble mean (average over the diurnal cycle), and is valid for the same time of year as the assimilation period. We compare the free run forecast, as well as 0.25-sol (ensemble mean) forecasts based on data assimilation analyses, with (independent, i.e., not yet used by the LETKF) TES observations, first showing the results and then discussing further the methodologies that led to significant improvements.

Figure 3.4 shows the global root mean square error (RMSE) of these 0.25 sol forecasts versus TES observations for the free run and assimilation systems of increasing complexity which successively augment a baseline run with 10% fixed multiplicative inflation and fixed dust opacity (0.3) for all ensemble members: spatially and temporally varying adaptive inflation, dust varying between 0.2 and 0.5 among ensemble members, and empirical bias correction. Note that the assimilation

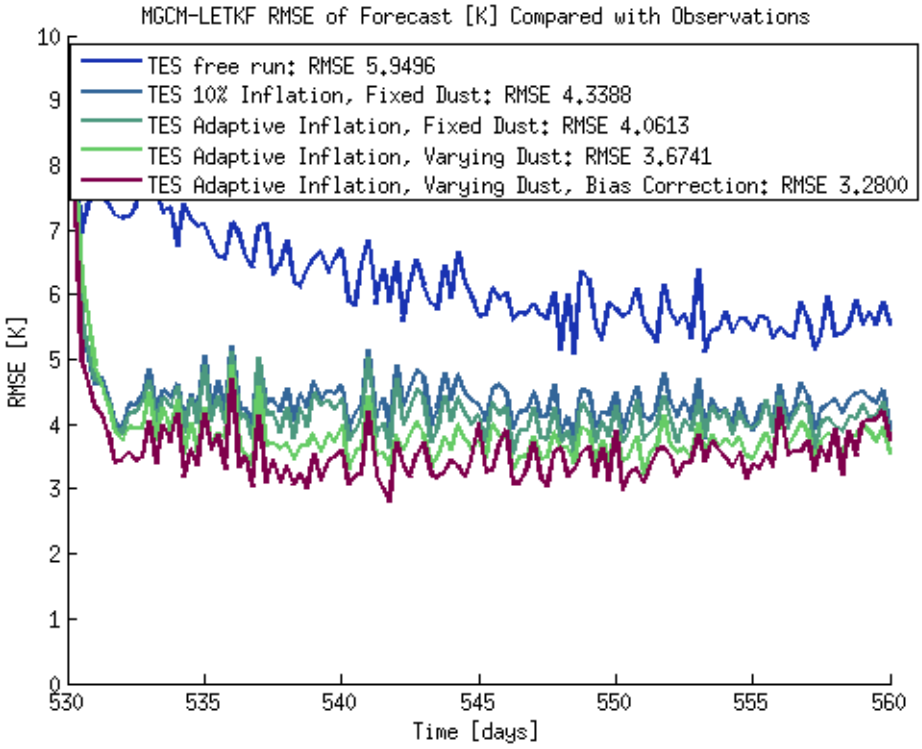


**Figure 3.4: Performance of the MGCM-LETKF, evaluated by comparing RMSE differences of 0.25-sol forecasts from assimilation analyses, as well as a free run forecast, with TES observations. RMSE values in the legend are from sols 10–20 of assimilation (after spin-up).**

systems rapidly converge from a climatological error of 6–7 K to a 0.25-sol forecast error of 3–4 K within 1–3 sols. There are short period (of a day or less) fluctuations in RMSE, which depend on the spatial coverage and error characteristics of the TES observations at that time, as well as flow-dependent model errors. Each of the improvements to the data assimilation system produces a decrease in RMSE, from 4.34 K for the baseline LETKF system to 3.28 K for the LETKF with varying dust, adaptive inflation, and empirical bias correction.

Figure 3.5 depicts latitude / altitude plots of zonal mean temperature bias and error standard deviation (defined here as $[\text{RMSE}^2 - \text{bias}^2]^{0.5}$) for the free run forecast, as well as short term forecasts from several LETKF configurations. The free run

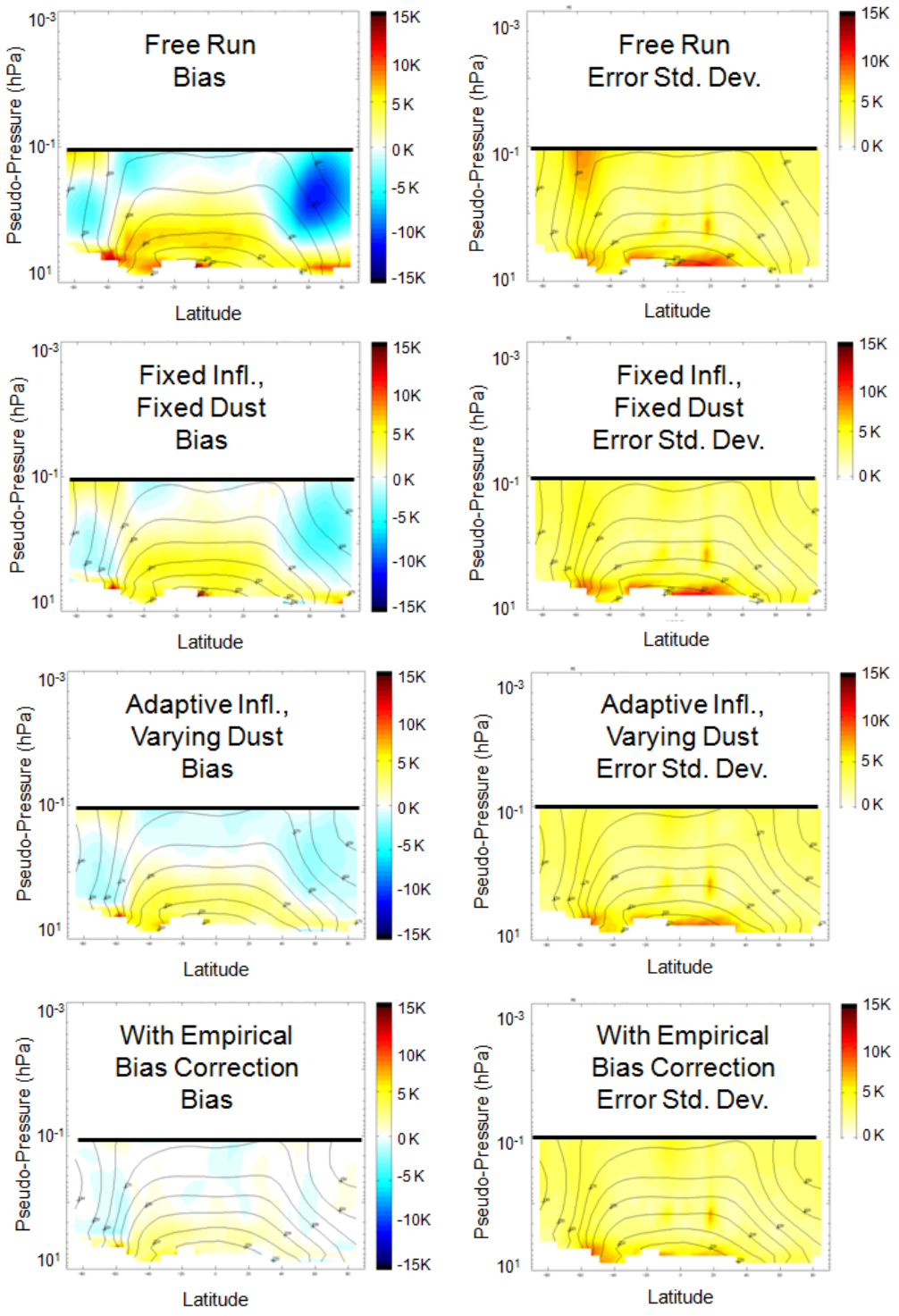


**Figure 3.5: Impact of improving data assimilation methods on zonal mean temperature bias (left, colors) and error standard deviation (right, colors) of 0.25-sol forecasts from data assimilation compared to TES observations for sols 10–20 of assimilation. Contours are forecast ensemble mean temperature. Row 1: free run forecast (no assimilation, just model climatology). Row 2: baseline data assimilation configuration, with fixed inflation and dust, no bias correction. Row 3: adaptive inflation and varying dust among the ensemble members, no bias correction. Row 4: Same as row 3, but with empirical bias correction.**

forecast (first row) RMSE is dominated by bias, particularly a warm bias (5–10 K) in the lower level tropics, and a cold bias (5–10 K) in the mid-level NH mid-latitudes. With the baseline LETKF configuration (second row), RMSE is generally < 10 K, and the regions of large bias are reduced to around 3–5 K for the tropics and the NH mid-latitudes. Error standard deviations are largest (> 5 K) near the SH cap edge instability belt, as well as near the surface in the tropics where diurnally evolving biases between model and observations are present.

Figure 3.6 helps to illuminate why the baseline LETKF struggles to correct the bias in the low level tropics. The upper left plot depicts background ensemble spread, which is the degree of uncertainty that the LETKF assigns to the background. Cool (warm) colors are depict a spread of less (more) than 3 K, where the LETKF trusts the background more (less) than the observations. The low level tropics are an area that is dominated by forcing from radiation (and hence dust distribution) rather than traveling waves, and is stable with respect to perturbations. In an ensemble with identical dust distributions, the temperatures among the different members in this region are similar, and the LETKF is therefore overconfident in the background estimate. This makes it difficult for the LETKF to update the temperatures in the tropics. Chapter 2 showed that this pattern of ensemble spread matches very well the atmospheric instabilities as inferred from bred vectors, particularly the mid-latitude baroclinic zones at around 60° N and S latitudes. Introducing varying dust distributions among the ensemble members (Figure 3.6, upper right) helps to increase the ensemble spread, particularly in the tropics, to more realistically represent our estimate of the uncertainty of the background ensemble. The ensemble spread is

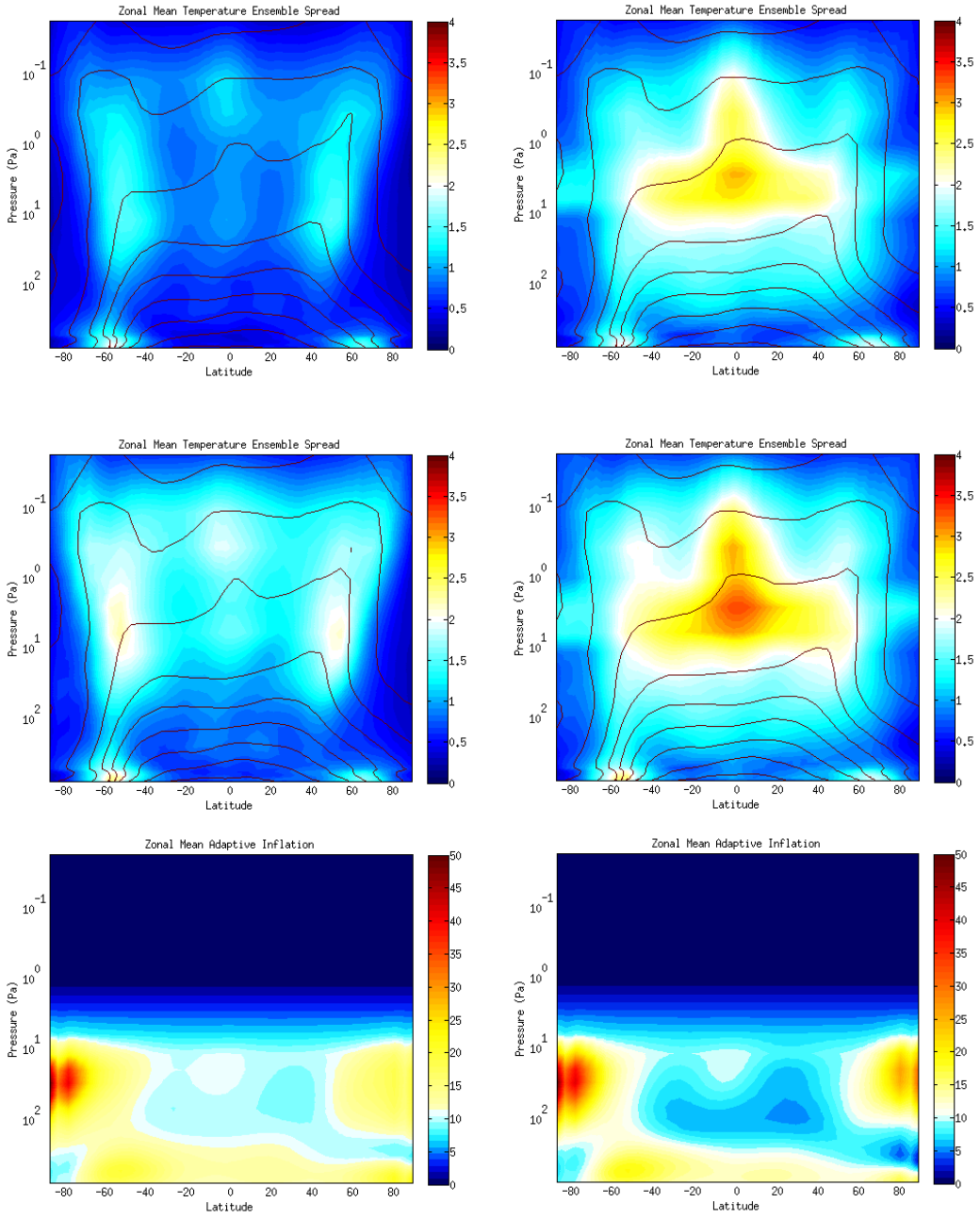


**Figure 3.6: Improvements to the ensemble spread through adaptive inflation and ensemble dust distributions. Top four panels are time averaged temperature [K] ensemble mean (contours) and spread (shaded). Upper left: fixed dust, fixed inflation. Upper right: varying dust, fixed inflation. Middle left: fixed dust, adaptive inflation. Middle right: varying dust, adaptive inflation. Lower row: covariance inflation (%) estimated adaptively for fixed dust (left) and varying dust (right) after 30 sols of assimilation.**

particularly large in the SH baroclinic zone, which has a tight temperature gradient near the edge of the polar ice cap, and in the upper levels of the atmosphere where there is no temperature data to constrain the uncertainty.

Adaptive inflation (Miyoshi 2011) uses error statistics to estimate the amount of inflation needed at each grid point so that forecast errors agree with the ensemble spread. The bottom panels of Figure 3.6 show the zonal mean adaptive inflation values (percentages) at the end of 30 sols of assimilation. White shading shows where the original value of 10% was preserved, and warmer (colder) colors show an increase (decrease) in inflation from this initial condition. Note the upper model levels have been assigned 0% inflation, as they have no observation information to constrain them. Less inflation is needed in the tropics when varying dust is used (right panel), as the differing dust distributions help to provide variation in atmospheric temperatures. Using both adaptive inflation and varying dust produces the best ensemble spread (Figure 3.6), as well as the best RMSE (Figure 3.5, row 3).

Even with adaptive inflation and varying the dust opacity, some bias (~3 K) remains between the short term model forecasts and observations. We therefore employ an empirical bias correction method (described in Section 3.3.5) to correct the MGCM each time step prior to assimilation to account for any mistuned parameterizations / model physics, using the mean analysis increment from the previous assimilation experiment as the bias correction field. We now have short term forecasts (Figure 3.5, bottom row) that are relatively free of bias error (order 1–3 K), and with a small RMSE (~3 K) except in the SH baroclinic zone near the surface where error standard deviations of >5 K remain. This region, with a maximum in

traveling wave activity, is unstable with respect to perturbations (Chapter 2), and is also occasionally subject to regional dust storms, which require assimilation of dust opacities to accurately address.

Figure 3.7 shows the sensitivity of the MGCM-LETKF performance (global RMSE) to our choice of horizontal localization distance. An original choice of 400 km was inspired by successful performance in OSSEs (M. Hoffman et al., 2010), as well as by the cutoff radius employed in the UK Reanalysis (Montabone et al., 2005; Lewis et al., 2007). The sensitivity curve shows a sharp decrease in performance for localization distances shorter than 400 km. Although 400 km produces a reasonable result, 600 km is slightly better with respect to RMSE. As longer localization distances tend to be more geostrophically balanced (Greybush et al., 2011), we selected 600 km horizontal localization as a better choice for our experiments.

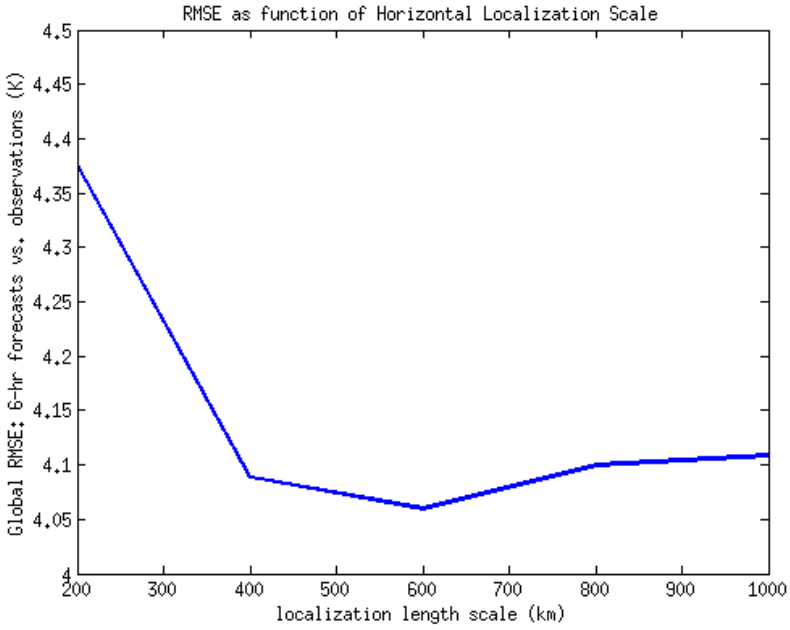


**Figure 3.7: Performance sensitivity of the MGCM-LETKF to localization length scale, evaluated by comparing RMSE differences of 0.25-sol forecasts from assimilation analyses with TES observations.**

### *3.4.2 Multi-Sol Forecasts and Predictability*

Another good approach for evaluating a data assimilation system is to produce longer-term forecasts and compare these to the latest analyses, as better initial conditions should lead to more accurate forecasts. To this end, we produce 10-sol ensemble forecasts, starting from the analysis ensemble at sol 20 of assimilation (sol 550 since perihelion). One forecast is produced from the analyses without bias correction (but with adaptive inflation and varying dust); the second is produced from analyses with empirical bias correction. We continue to apply the same bias correction term to the second forecast ensemble, although we do not assimilate any additional observations. For both the forecasts and assimilation, the bias was calculated from analysis increments during the 10 sols prior to this experiment (sols 540–550 since perihelion), so that the forecasts have no knowledge of future observations, and merely assume persistence of model bias.

Figure 3.8 shows forecast RMSE compared to TES observations, in the same style as Figure 3.5. The long-term forecasts begin with a similar RMSE to the analyses (3–4 K). After about 3 sols for no bias correction, and 5 or more sols with bias correction, the RMSE converges to that of the free run (6 K). Therefore, we conclude that forecasting with initial conditions from a data assimilation system is more skillful than just using model climatology out to a predictability horizon of about 3 sols, and that this can be extended to about 5–6 sols when accounting for the model bias as done within the data assimilation.

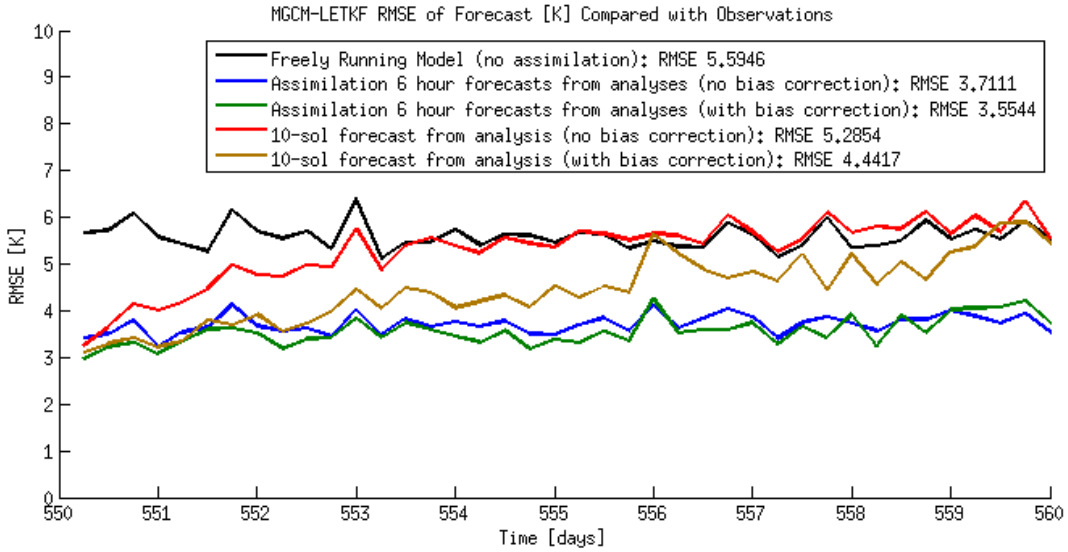


**Figure 3.8: Global RMSE of temperature forecasts compared to TES observations. The black curve is from a freely running model with no data assimilation. The blue (without bias correction) and green (with empirical bias correction) curves are from 0.25-sol forecasts based on data assimilation analyses. The red and gold curves are 10-sol forecasts that are initialized from a data assimilation analysis at the onset, but have no knowledge of observations afterwards. The red curve is without bias correction, and the gold curve applies a bias correction field, calculated from the mean analysis increment over the previous 10 days of assimilation, every 0.25 sols.**

Figure 3.9 shows plots of forecast errors (colors) for the surface pressure, zonal mean temperature, and u-wind versus analyses at 5 sols (a temporal average over sols 4.5 through 5.5). At 5 sols, without bias correction, the temperature errors are nearly as large as the free run forecast, although the surface pressure and wind errors are significantly smaller. With bias correction, the 5-sol temperature forecast shows a clear improvement over the free run, particularly in the lower level tropics and aloft in the NH mid-latitudes. The most challenging areas to forecast are in the SH mid-latitudes near the cap edge, where strong temperature gradients and baroclinic instabilities are present, and the NH polar regions, which also exhibit larger errors in surface pressure.

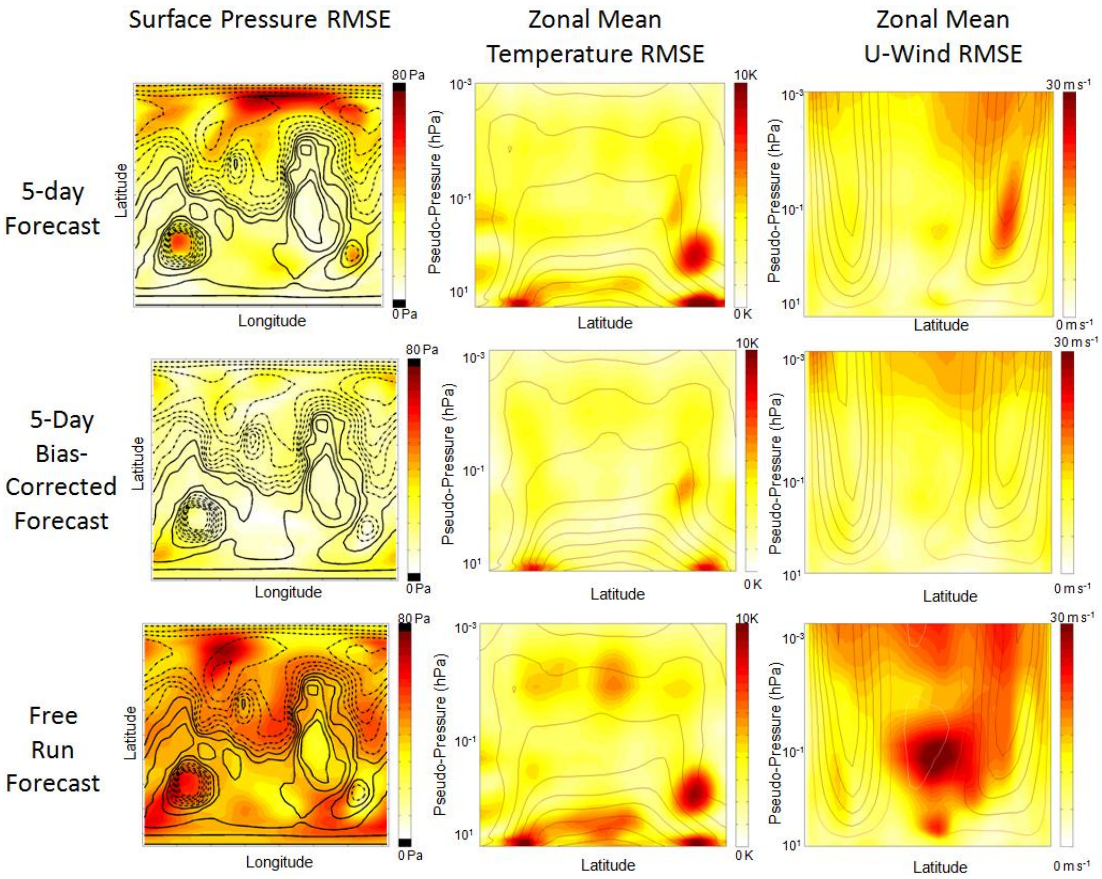


**Figure 3.9: 5-day forecast RMSE as compared to the most recent analysis from data assimilation for surface pressure (left column), zonal mean temperature (middle column), and zonal mean u-wind (right column). Colors are RMSE, whereas contours are time mean analysis quantity. The top row are forecasts without empirical bias correction, the middle row are forecasts with empirical bias correction, and the last row are a free run forecast without any observation data.**

As there are large variations in atmospheric stability with respect to season (Newman et al., 2002; Chapter 2), we intend to explore how this predictability horizon changes seasonally; we would also expect it to decrease significantly in the presence of dust storms. On Earth, large error growth rates tend to limit atmospheric predictability due to uncertain initial conditions to around two weeks (Lorenz, 1965). For Mars, growth rates tend to be smaller than for the Earth in most seasons (Chapter 2), and there are reproducible weather features each year (Cantor et al., 2002). However, model biases are larger, and, with a much faster radiative time scale on

Mars (on the order of 1–2 sol rather than 10 days), model forecasts tend to drift quickly back toward climatology in the absence of observations (Rogberg et al., 2010; P. Read, personal communication). While bias correction methods can help ameliorate the "climate drift" of Mars GCMs, improvements in the model physics that reduce temperature biases will lead to further increase in forecast skill. Dust storms should remain a significant challenge to Martian forecasters.

### *3.5 Conclusions and Future Work*

We have successfully assimilated TES temperature profiles into the GFDL Mars General Circulation Model (Wilson et al., 2002) during a 30-sol interval in NH autumn of MY 25, creating analyses of temperature, wind, and surface pressure as well as their uncertainty using an advanced data assimilation system, the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007). Short-term (0.25-sol) forecasts based on the analyses were compared to independent (in time) observations, and showed a significant improvement in both bias and RMSE compared to a free run forecast that did not use observations. Baseline performance of the LETKF data assimilation system was substantially improved through the use of adaptive inflation (Miyoshi, 2011). Varying dust distributions among the ensemble members helped to improve both the ensemble mean state and its spread (which depicts analysis uncertainty), particularly in the tropical lower levels where the atmosphere is more stable with respect to perturbations and is driven by radiation. An empirical bias correction scheme based on the time mean analysis increments produced further improvement, resulting in short-term forecasts with generally less than 3 K RMSE except for the near surface ice cap-edge baroclinic zone in the SH.

Finally, we conducted longer-term forecasts out to 10 sols. With the empirical bias correction technique, we estimate a predictability horizon of about 5 sols or more according to this single case forecast. That is, forecasts with initial conditions from our best configuration of LETKF data assimilation analyses and with correction of estimated bias during the forecast period remain superior to those produced by the freely running model out to 5 sols.

The techniques used in this chapter will form the basis for a new Mars weather and climate reanalysis spanning the entire TES period, which we will expand (Greybush et al., 2010) to include Mars Climate Sounder (MCS) data (Kleinboehl et al., 2009). The next major step for our research is the direct assimilation of dust aerosol from TES column opacities and MCS dust profiles. Future plans also include a new approach to retrieval assimilation that removes the influence of the prior and addresses vertical observation error correlations (R. Hoffman, 2010). We also will employ recent upgrades to the MGCM, which include parameterizations for topographic gravity wave drag and radiatively active water ice clouds, which we anticipate will reduce model biases. A comparison with the UK reanalysis is also underway (L. Montabone, personal communication), and preliminary results will be reported in the next Chapter.

# Chapter 4: Evaluation and Science Insights

## *4.1 Comparison to the UK Reanalysis and Radio Science*

### *4.1.1 Comparison between Reanalyses*

The "UK Reanalysis" (Lewis et al., 2007; Montabone et al., 2005), which comprehensively analyzed the TES era, provides a baseline for evaluating our LETKF reanalysis. This product used the Oxford-LMD Mars GCM (Forget et al., 1999) and contributed insights to atmospheric tides (Lewis and Barker, 2005), the variability of dust storms (Montabone et al., 2005), tropical water ice clouds (Wilson et al., 2008), and predictability (Rogberg et al., 2010). Table 4.1 provides a comparison of the methods and model used in each reanalysis. The UK reanalysis used the analysis correction scheme (Lorenc et al., 1991), where model fields are repeatedly nudged toward observations, a technique which was employed operationally for the Earth's atmosphere in the UK Meteorological Office in the 1990s. Our reanalysis using the Local Ensemble Transform Kalman Filter (LETKF) evolves an ensemble of model states in time, updating them with observations every 0.25 sol, thus providing an uncertainty estimate as well as a state estimate. The two reanalyses have roughly the same horizontal and vertical resolution. The UK reanalysis considers TES dust opacities in a manner similar to direct insertion; whenever a new opacity is available it directly replaces the previous value in an opacity forcing field. As there are large data voids in TES opacity (retrievals are not reliable for surface temperatures less than 220 K and during intense dust storms), this opacity field is not necessarily complete or timely. The LETKF reanalysis does not

**Table 4.1: Comparison of the UK Reanalysis and the LETKF Reanalysis.**

| Parameter | UK Reanalysis | LETKF Reanalysis |
|---|---|---|
| ***Assimilation Scheme*** | Analysis Correction Scheme (Lorenc et al., 1991), which is similar to nudging | Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007) with enhancements (adaptive inflation, empirical bias correction) |
| ***Update Frequency*** | Continuously (every model time step) | Assimilation cycle is 0.25 sol |
| ***Temporal Availability*** | Every 2 Mars hours | Every 6 Mars hours, or hourly as desired |
| ***Atmospheric Model*** | Oxford-LMD Mars model (Forget et al., 1999) | GFDL MGCM (Wilson et al., 2002; Hoffman et al., 2010) |
| ***Model Resolution*** | 72 × 36 × 25 levels | 60 × 36 × 28 levels |
| ***Vertical Coordinate*** | Sigma | Hybrid Sigma-Pressure |
| ***Temperature Data*** | PDS TES Profiles, with vertical averaging | PDS TES Profiles at TES levels |
| ***Dust Methodology*** | TES dust opacities directly inserted. | Initially, fixed dust opacity varied among ensemble members. Eventually, updated from observations. |
| ***Variables Updated*** | T, U, V, surface pressure, dust | T, U, V, surface pressure |
| ***Uncertainty Estimate*** | None | From Ensemble (16 members) |
| ***Localization Cutoff Radius*** | 1200 km | 1460 km (400 km Gaussian * 3.65) horizontal |
| ***Availability*** | Entire TES Period; some MCS | Intervals from MY 24, MY 25 for TES, MY 29 for MCS |

yet incorporate dust observations, although we plan to do so in the near future (Section 4.4).

A preliminary intercomparison between the UK reanalysis and LETKF reanalysis has been conducted for a 15-sol period in the N.H. autumn of MY 24 ($L_s$ 182.3°–191.1°). A more extensive comparison between the two reanalyses is planned to be underway in summer 2011 following the launch of a public web interface to the UK reanalysis dataset (Luca Montabone, personal communication). Both reanalyses have assimilated TES temperatures from the Planetary Data System (PDS) archive (Smith et al., 2001). To facilitate comparison, the UK reanalysis is interpolated to the

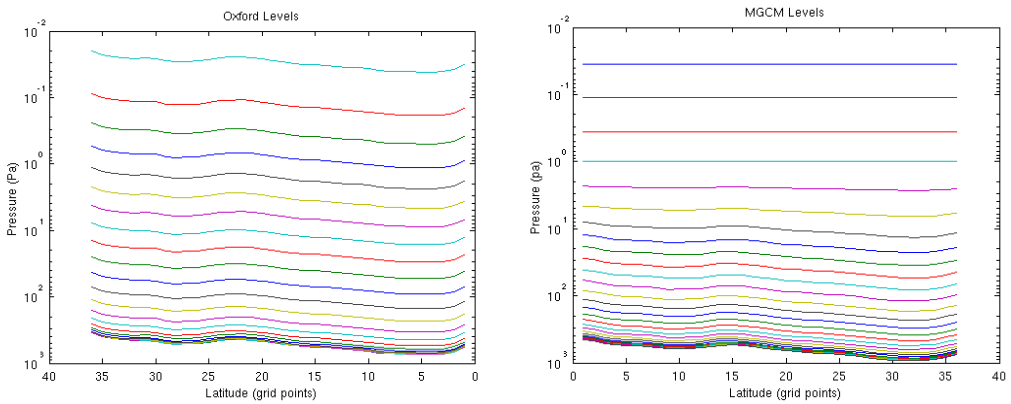


**Figure 4.1: Schematic comparing vertical resolution of the Oxford-LMD Mars model versus the GFDL MGCM. The Oxford-LMD model is purely sigma coordinate, whereas the MGCM transitions to pressure coordinates at higher altitudes.**

same horizontal grid as the MGCM. The UK reanalysis employs a sigma vertical coordinate throughout the domain, whereas the MGCM uses a hybrid system that reverts to pressure surfaces at upper levels (Figure 4.1), so the next step is to vertically interpolate the UK reanalysis to the sigma level that matched the LETKF reanalysis hybrid pressure level at each grid point. Finally, the reanalyses are compared every 0.25 sol corresponding to exactly the same time period; this required a 0.25-sol shift in the listed "model $L_s$" for the UK Reanalyses to match the real $L_s$ used by observations and the LETKF system (S. Lewis and R. J. Wilson, personal communication). Figure 4.2 shows an example of temperatures at the lowest model level as the UK reanalysis (upper left) is interpolated to the LETKF grid (upper right) and compared to the LETKF (lower left). The temperature maxima due to the thermal tide align in longitude, indicating that the times match. The lower right shows the UK column dust opacity (as an optical depth, where a value of 1 indicates intensity is reduced by 1/*e* times the radiation entering the top of the atmosphere). Values are between 0.2 and 0.5 (which we use as bounds for our ensemble dust

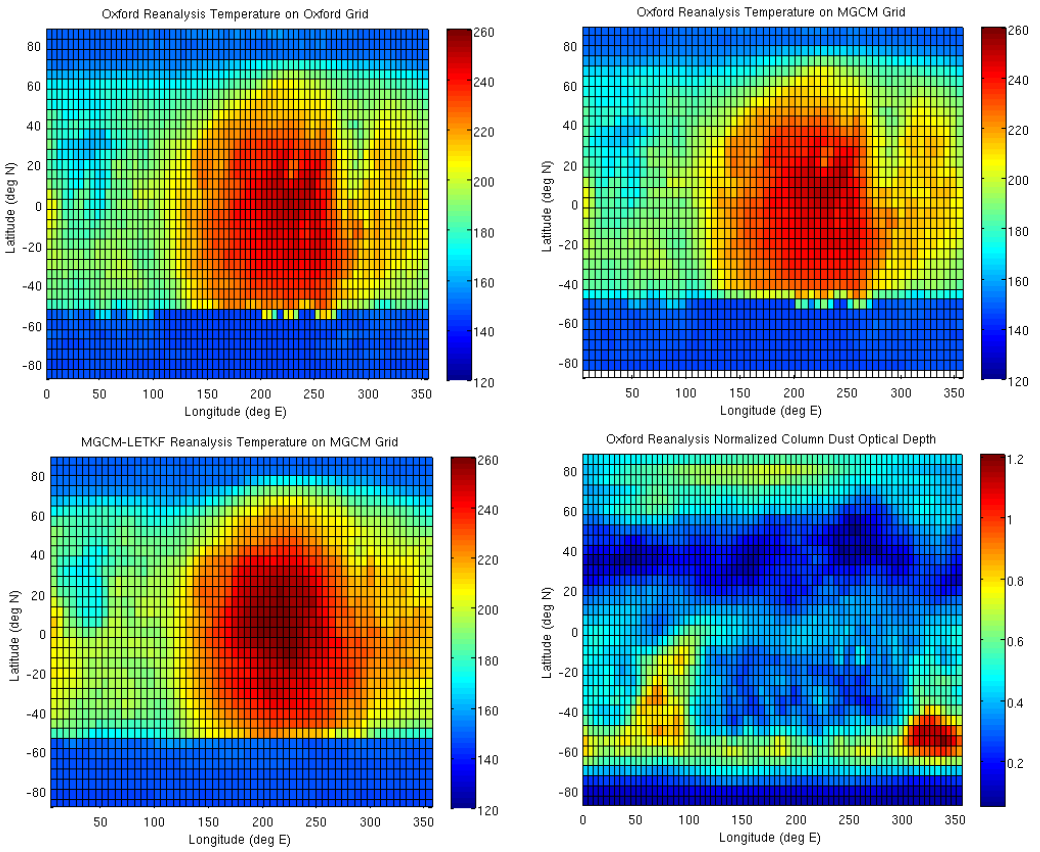


**Figure 4.2: Preliminary comparison of temperature fields in the lowest model level. This comparison takes place at a snapshot in time, on Mars Year 24, Mars Day 565, Mars Hour 00, or $L_s$ 182.3°. Upper left: UK Reanalysis temperature field on its own coordinates. Upper right: UK Reanalysis temperature field interpolated horizontally to the MGCM grid. Lower left: MGCM-LETKF temperature field on its own coordinates. Lower Right: UK Reanalysis column dust opacity, normalized to a surface pressure of 700 Pa.**

opacities) primarily in the tropics and NH mid-latitudes. Exceptions are the NH polar regions and SH sub-arctic latitudes where opacities are 0.6–0.9, and nearly 1.2 in a local dust storm at 310°–340° longitude.

Figure 4.3 shows the temperature bias and standard deviation difference between the reanalyses averaged over the 15-sol period. Generally the reanalyses agree in the zonal mean within 5 K throughout much of the domain. Larger differences in both bias and standard deviation exist at the cap edge baroclinic zones, which can either be due to differences in the strength and phase of traveling waves or

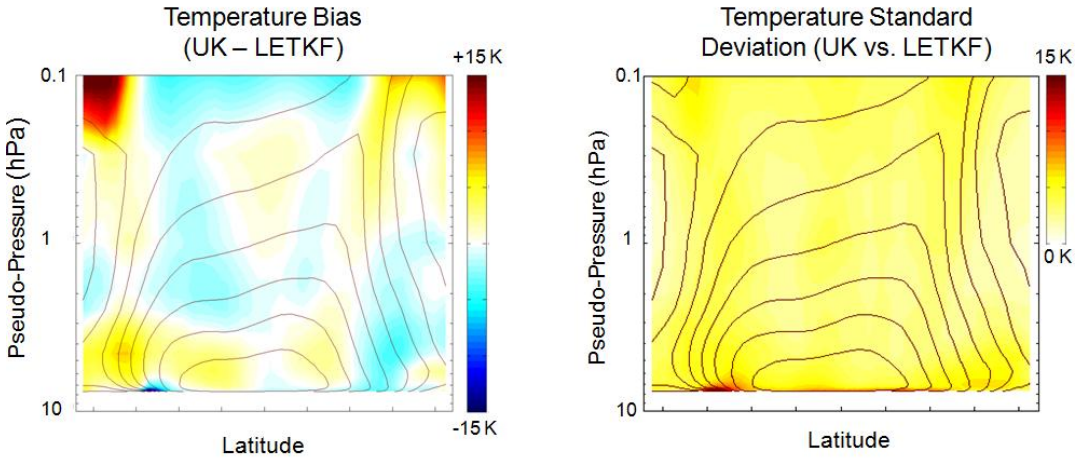


**Figure 4.3: Zonal mean temperature bias (UK – LETKF) and RMS difference between the UK reanalysis and the LETKF reanalysis.**

the presence of local dust storms. There is also a large bias in polar regions around 0.1 hPa which is due to differences in the model climatology, as this is the vertical limit of observation coverage.

Another way to compare the reanalyses is to examine the phase and amplitude of mid-latitude lower atmosphere traveling waves. Figure 4.4 shows Hovmoller diagrams of temperature perturbations for 15 sols at MGCM Level 20, or about 3 km in altitude, at 60 ° N and 60 ° S latitudes. The perturbations have been calculated as the difference from the time and zonal mean state. The diagrams reveal that the reanalyses generally agree on the phase of the traveling wave with zonal wavenumber 2, although the LETKF estimates a higher amplitude for waves in the SH.

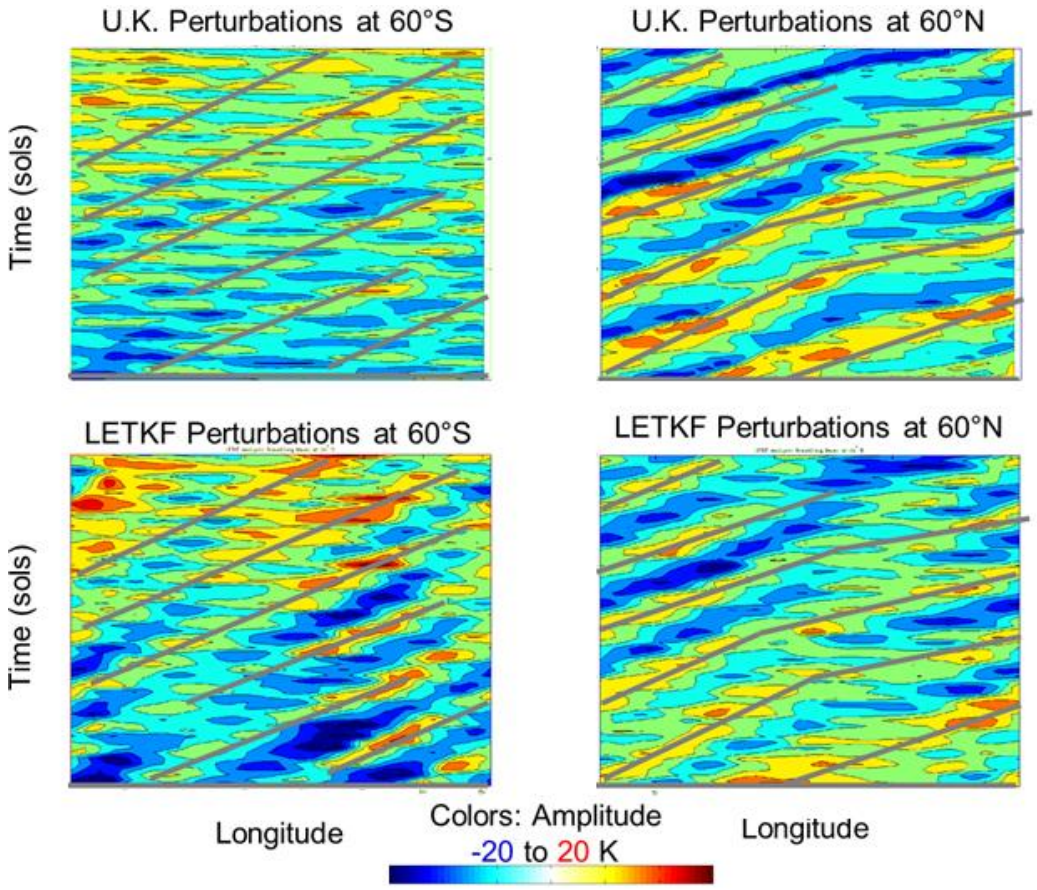


**Figure 4.4: Hovmoller diagram of traveling waves (temperature perturbations at ~3 km altitude) at 60°S (left panels) and 60°N (right panels) in the UK (top panels) and LETKF (bottom panels) reanalyses. Gray lines are the same for both reanalyses, and represent a subjective depiction of the peaks of the traveling waves by the author.**

### *4.1.2 Comparison with Independent Radio Science Profiles*

Radio Science (RS) temperature profiles (Hinson et al., 1999) provide an independent set of observations to evaluate the reanalyses, as they have not been assimilated in either product. These profiles are obtained by examining the effects of atmospheric refraction on a coherent microwave signal from the Mars Global Surveyor (MGS) spacecraft as the signal passes through the limb of the Mars atmosphere on its way to the NASA Deep Space Network (DSN) receivers on the Earth. From the profiles of refractivity, number density can be obtained with

knowledge of the principal atmospheric constituents. Finally, pressure and temperature profiles are derived by Hinson et al., assuming hydrostatic balance and the ideal gas law. The RS profiles are most accurate near the surface (uncertainty < 1 K), and grow more uncertain as the altitude increases, with uncertainties of up to 10 K near 40 km (Figure 4.6, green curve). Hinson et al. (2004) found a good agreement between RS and TES profiles when taking into account the limited vertical resolution of TES profiles.

RS profiles are available only at select latitudes, which vary from season to season and across Martian years (Figure 4.5, top panel). There are no RS profiles available during the MY 24 NH autumn assimilation period shown in Section 4.1a, but there are in the NH sub-arctic (Figure 4.5, bottom panel) during the MY 25 assimilation experiments just prior to the 2001 global dust storm. During a 20-sol intercomparison period ($L_s$ 170.5°–181.9°), we conduct a three-way intercomparison

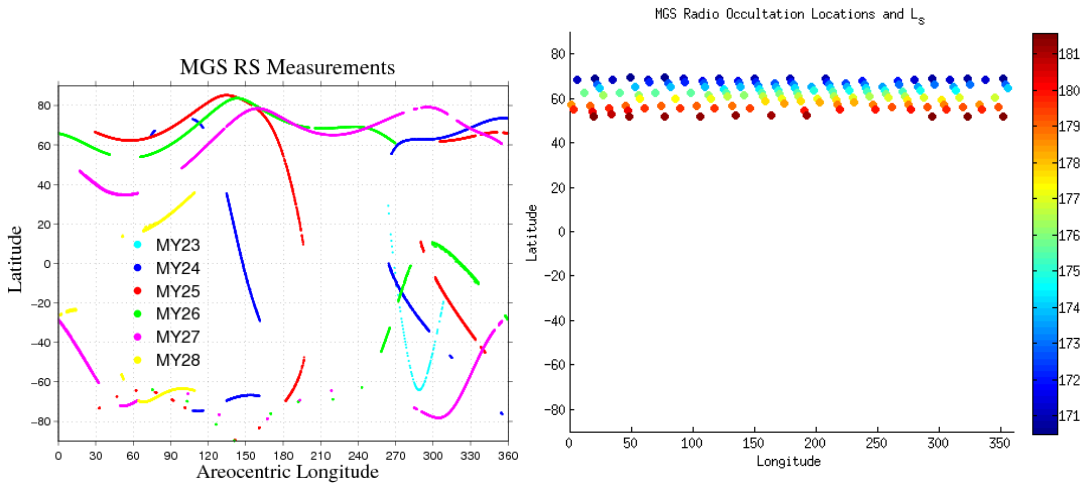


**Figure 4.5: Location of the Radio Science profiles over the entire MGS mission (top panel), and those used to evaluate the reanalyses during MY 25 NH autumn (bottom panel). The profiles are generally located in the NH sub-arctic, with the latitude shifting gradually southward over the course of 20 sols (colors are $L_s$ value).**

of both UK and LETKF Reanalyses interpolated separately to the locations and times of the RS profiles.

Figure 4.6 plots the mean temperature difference, both bias and RMSE, between the reanalyses and the RS profiles, taking the RS profiles to be "truth". The solid curves, depicting RMSE, shows the LETKF (red) to be closer to RS (black) than the UK (blue) at all levels, with a 4.56 K RMSE compared to 5.79 K. Differences are smallest at around 1–4 hPa for both reanalyses. At higher altitudes, uncertainty of the RS temperatures also grows, meaning that the higher RMSE of the reanalysis does not necessarily imply that it is further from the true atmosphere. Both reanalyses

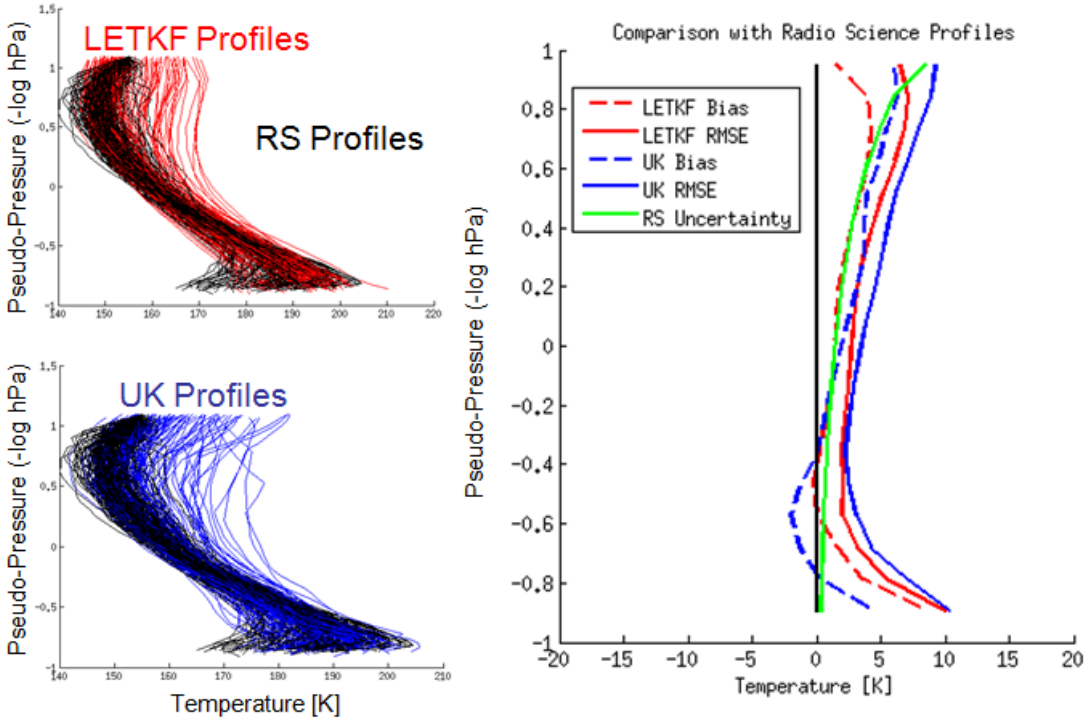


**Figure 4.6: Left panel: Spaghetti plot of temperature profiles from radio science (black), LETKF reanalysis (red), and UK reanalysis (blue). Right: bias and RMSE of the LETKF and UK reanalyses, taking the radio science data as truth. The uncertainty of the RS profiles is denoted by the green curve. Column mean values are 4.56 K RMSE and 2.03 K bias for the LETKF, and 5.79 K RMSE and 2.36 K bias for UK.**

exhibit a warm bias throughout much of the vertical column, particularly near the surface. With the coarse vertical resolution of TES observations, the near surface bias is likely due to a model bias in the reanalyses.

### *4.2 Mars Climate Sounder (MCS) Assimilation*

The Mars Climate Sounder instrument (MCS) is a limb sounder located aboard the Mars Reconnaissance Orbiter (MRO) spacecraft dedicated to mapping atmospheric temperature and aerosol vertical structure (McCleese et al., 2007). MCS offers improved vertical resolution (double that of TES) and vertical extent (twice that of TES) in the thermal structure; the MCS retrievals (Kleinboehl et al., 2009) have been used to provide improved temperature constraints for models to 80 km altitude at up to 105 vertical levels. In addition MCS offers the first-ever vertical distributions of dust and ice clouds (McCleese et al., 2010). Unfortunately, the TES instrument record ended due to a malfunction prior to the arrival of MRO, so a direct comparison between TES and MCS is not possible.

The MRO spacecraft makes approximately 13 orbits per sol, passing over the ground at 3 a.m. and 3 p.m. local time, and providing thousands of atmospheric observations each Mars day. As the MCS instrument is a limb sounder, observations are not vertically stacked, but can differ by up to 4 degrees latitude in a "profile;" measurements at each level are assigned unique latitude and longitude values when assimilated. Through much of the MCS mission, the limb view has been along track (parallel to the direction of orbit), although since September 2010 there have been some cross track measurements (A. Kleinboehl, JPL, personal communication). There are more temporal gaps in MCS than TES, due to a combination of stricter quality

control and disruptive spacecraft operations. Retrieval residuals (Figure 4.7, bottom right panel) are generally less than 1 K for temperature below 50 km; considering systematic errors as well, temperature uncertainties are 1–2 K in the lower atmosphere, and can be higher in the case of large aerosol opacities. Some caveats for aerosol retrievals (D. Kass, JPL, personal communication) include that $CO_2$ ice in polar regions might be mistaken as dust; and when both dust and ice aerosol are present, there is less confidence in measurements of the minor constituent.

Figure 4.7 displays a sample 5-sol MCS mean near the NH autumnal equinox of temperature, water ice, and dust aerosol vertical structure. Temperatures at $L_s$ 171° are nearly symmetric about the equator, slightly cooler in the SH high latitudes with its residual ice cap from the previous winter. There is an elevated temperature maximum aloft due to the polar warming phenomena, with the coolest temperatures in the tropics. Dust concentration decreases with height, with greater vertical extent in the tropics. There is a maximum in water ice at around 5 Pa in the tropics, a favored region for water ice clouds, as well as at low levels in the polar regions.

Figure 4.8 shows the result of assimilating MCS temperature profiles with the LETKF during a 20-sol period of NH autumn ($L_s$ 170.5°–181.9°) of MY 29. The LETKF and MGCM settings are similar to those used in Section 3.3. A RMSE of 5–15 K is greatest at upper levels where there were previously no TES observations to constrain the MGCM, and the RMSE has a large component of bias (compared to error standard deviation), although the bias is significantly improved from the freely running model. Bias, computed as observation minus 0.25 sol model forecast, is largest at low levels (observations warmer than model), the tropical mid-levels

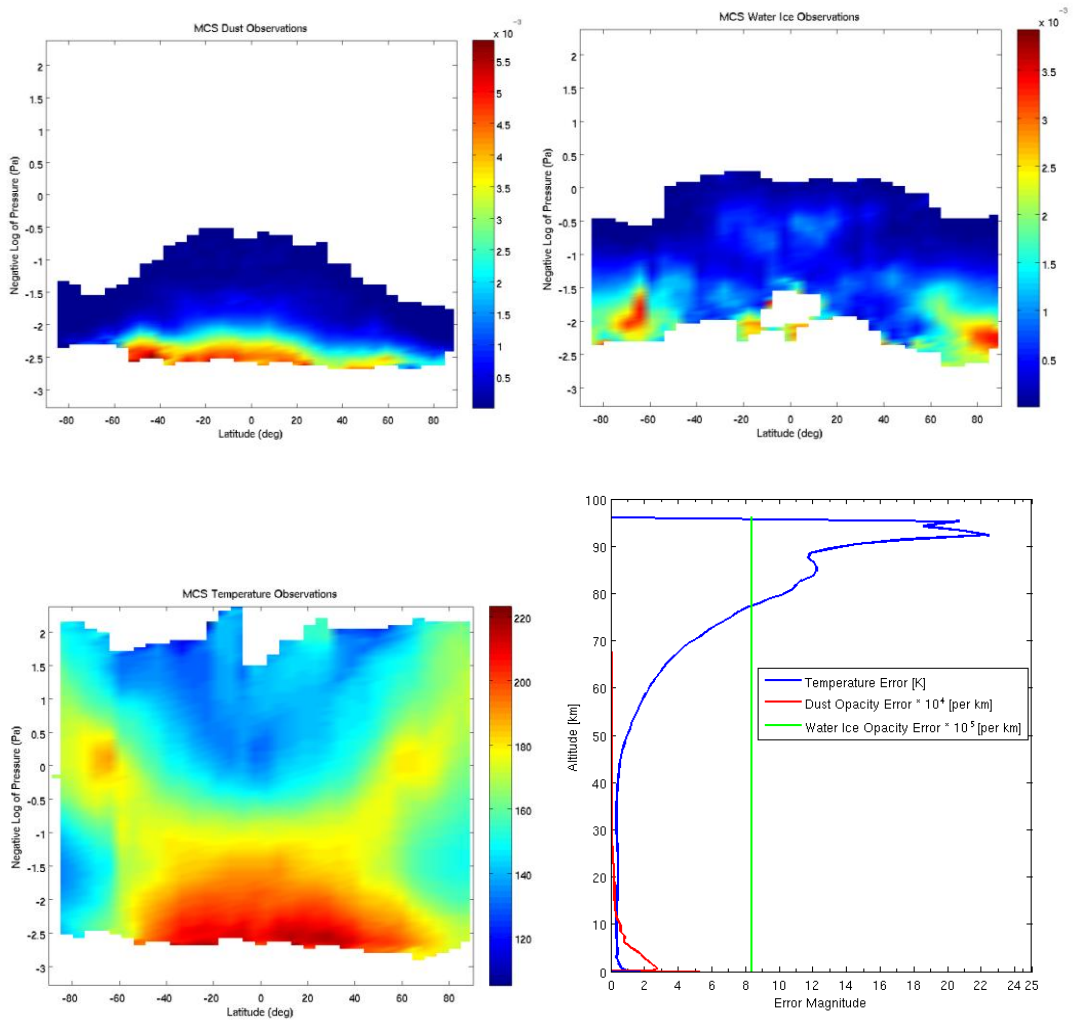


**Figure 4.7: Sample Mars Climate Sounder (MCS) climatology for Days 540–544 (Ls 170°–172°) of Mars Year 29. Upper left: dust aerosol concentration [$10^{-3}$ $km^{-1}$]. Upper right: water ice concentration [$10^{-3}$ $km^{-1}$]. Lower left: temperature [K]. Lower right: Uncertainty values for each observation type (temperature in K, dust opacity in $10^{-4}$ $km^{-1}$, water ice opacity $10^{-5}$ $km^{-1}$).**

(another warm bias), the tropical upper levels (cold bias), and the polar upper levels (warm bias). The empirical bias correction method described in Section 3.3 helps to reduce the forecast bias to less than 6 K through much of the domain, which consequently improves the RMSE (middle panels). The next section describes physical insights that may help explain these biases.

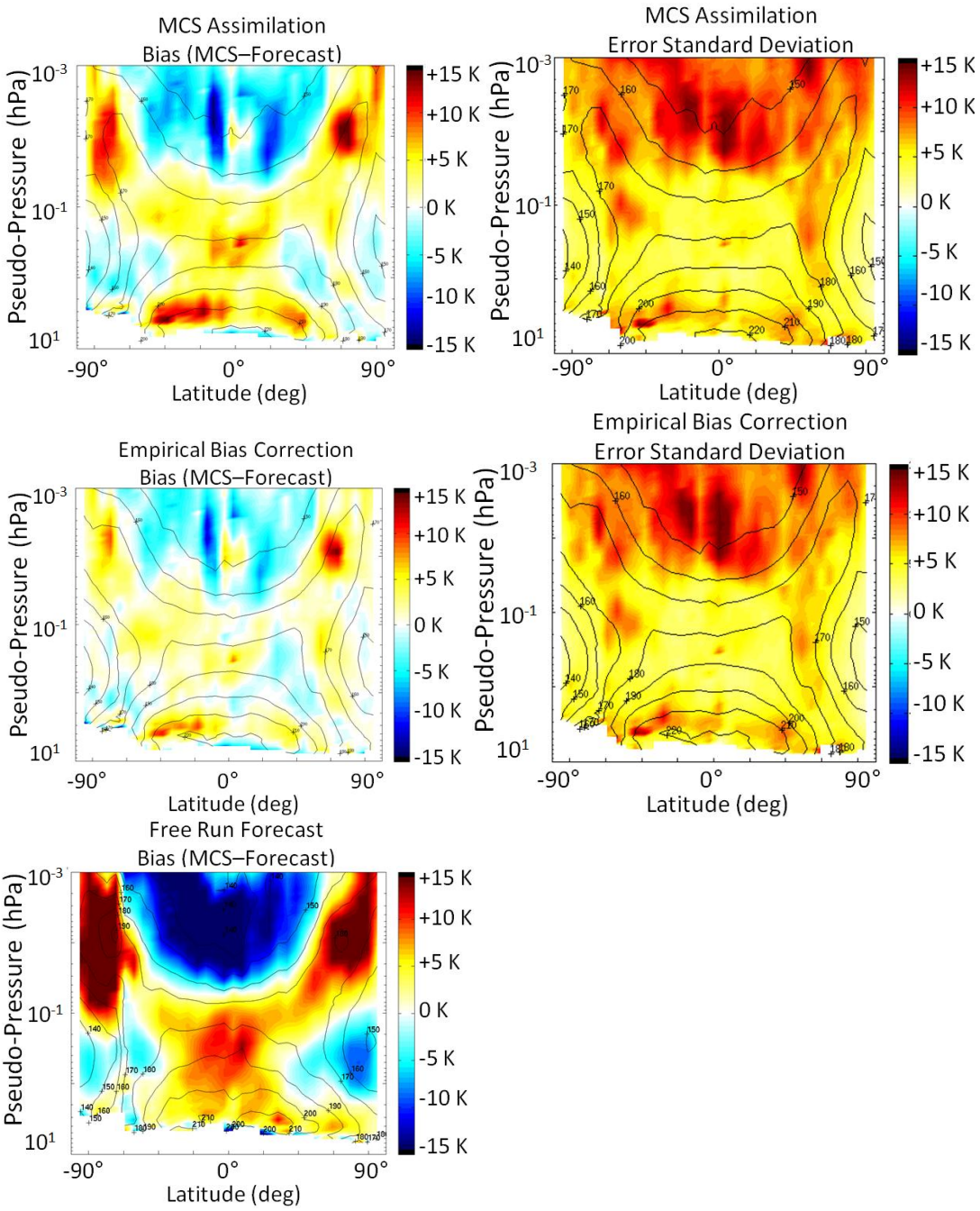


**Figure 4.8: MGCM-LETKF assimilation of MCS version 3 profiles using adaptive inflation, shown as an average over sols 10–20. Left: Bias (MCS minus forecast). Right: error standard deviation. Contours: analysis ensemble mean. Top row: no bias correction; middle row: with empirical bias correction; bottom row: freely running model.**

## *4.3 Insights for Model Development*

Evaluation of the Mars assimilation analyses can produce insights into potential improvements of the MGCM, particularly for biases between the model and observations.

Figure 4.9 explores the bias between observations and short-term model forecasts for several configurations of the MGCM. Panel A depicts the bias from the control assimilation without empirical bias correction (same as Figure 4.8, top left panel). In Panel B, the MGCM has been upgraded with new physics, including a Mellor-Yamada boundary layer scheme (Mellor and Yamada, 1982), a topographic wave drag parameterization, a more realistic radiation scheme from NASA Ames, and revised polar ice cap albedo and emissivity and dust properties. These changes help to reduce the 10 K warm bias to less than 5 K at low levels, particularly in the SH. However, biases remain large above 40 km. One reason for these biases is that before the recent arrival of MCS observations, model developers had little information available by which to tune upper level temperatures.

Dust aerosol and water ice clouds play a significant role in modifying atmospheric temperatures. Panels A and B of Figure 4.9 use an MGCM without radiatively active water ice clouds, which means the model is too cool at 30–40 km, and with a vertical dust distribution extending to high altitudes, which means the model is too warm in the tropics above 50 km. Panel C determines the vertical extent of the dust based upon model dust tracers which have been lifted using the convective (dust devil) lifting parameterization and updated using TES opacities from MY 24 (see the second dust option in Section 1.4). This reduces the upper level tropical bias,

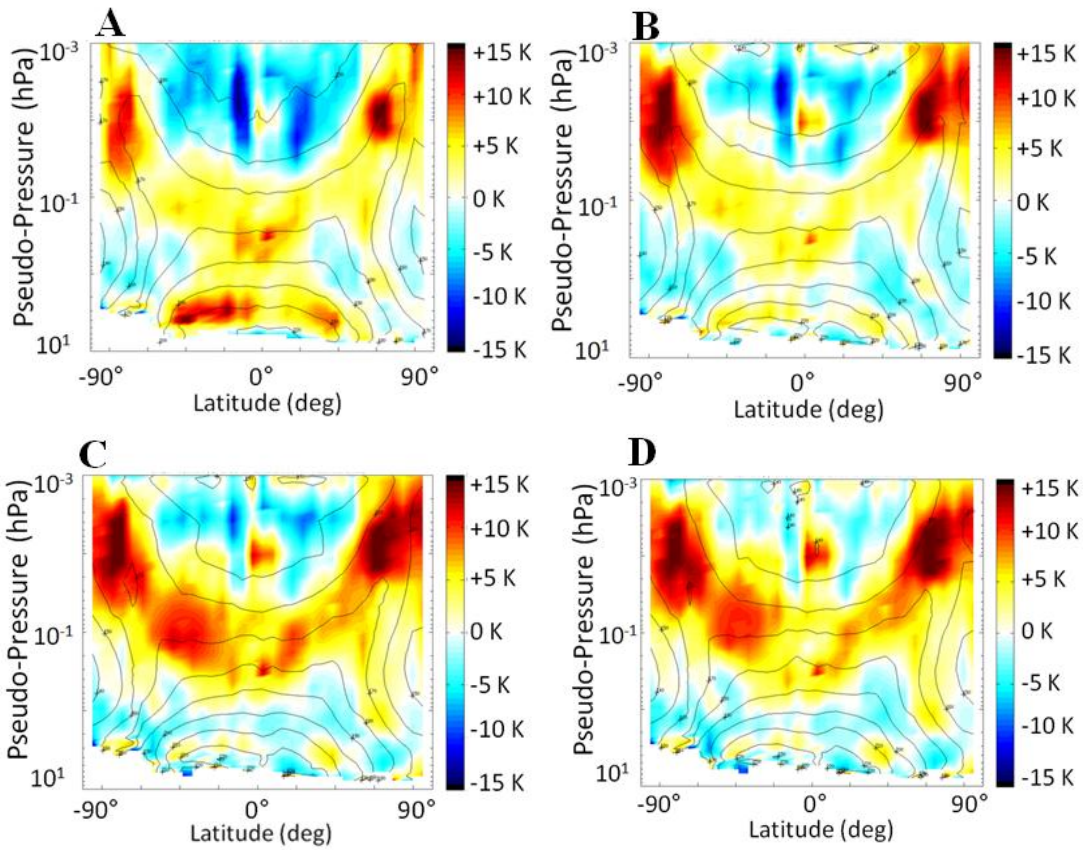


**Figure 4.9: Bias (observation minus forecast) of MGCM-LETKF 0.25 sol forecasts from assimilation compared with MCS temperature observations (contours and experimental period same as Figure 4.8) for various MGCM configurations. (A) Control assimilation from Figure 4.8 with spatially fixed dust opacity (0.2–0.5), original dust height, and original MGCM physics. (B) Same as (a), but with new MGCM physics (see text). (C) Same as (b), but with dust opacities determined based on TES climatology and dust height determined from model tracers. (D) Same as (c), but with radiatively active water ice clouds.**

but the model is now too cool in the mid-level tropics (red shading in Figure 4.9), as well as the upper level poles. Despite the added realism of the dust (vertical extent matches actual model tracers, and column opacities are based climatologically on observations), bias overall has increased somewhat. Experience has shown that simply tuning the dust in the MGCM is not enough to match observed temperatures at high levels (John Wilson, personal communication); water ice clouds need to be considered. Panel D adds radiatively active water ice clouds and varies their strength

among ensemble members, which reduces bias in the tropics at 30–40 km. Polar warming at 60 km continues to be underestimated by the MGCM; increasing the model top to more than 100 km may help address this problem (McDunn et al., 2011). Assimilation of TES and MCS aerosols, particularly the vertical aerosol distribution available with MCS, will help to improve forecast biases with observations; a strategy for dust assimilation is outlined in the next section. Insights toward improving the MGCM from MCS data can also improve the quality of the TES reanalysis.

### *4.4 Strategy for Improving the Representation of Dust*

Improvement of the representation of dust in these experiments should increase the realism of the simulations, as well as the performance of the assimilation system. Section 1.4 overviewed the ways in which the MGCM can specify dust distributions. Expanding from the fixed normalized opacity employed in Chapters 2 and 3, the next logical progression is a spatially and temporally evolving distribution based on observations. John Wilson has released MY 24 (a typical Mars year without a global dust storm) TES interpolated dust opacity fields that can drive the MGCM; dust aerosol in the model is represented by tracers that are transported by wind, and when there is a disagreement between the model column opacity and observations, dust is injected or removed from the boundary layer. Figure 4.10 shows dust visible opacities from a simulation with "TES dust," along with the extent of the polar ice cap. Note that TES dust opacities are unreliable for surface temperatures under 220 K (and hence over the polar caps). Black contours represent opacities in the range of 0.2 to 0.5, values which were used in the data assimilation ensemble. At $L_s$ 166°, these fixed opacities were reasonable for much of the domain with the exception of

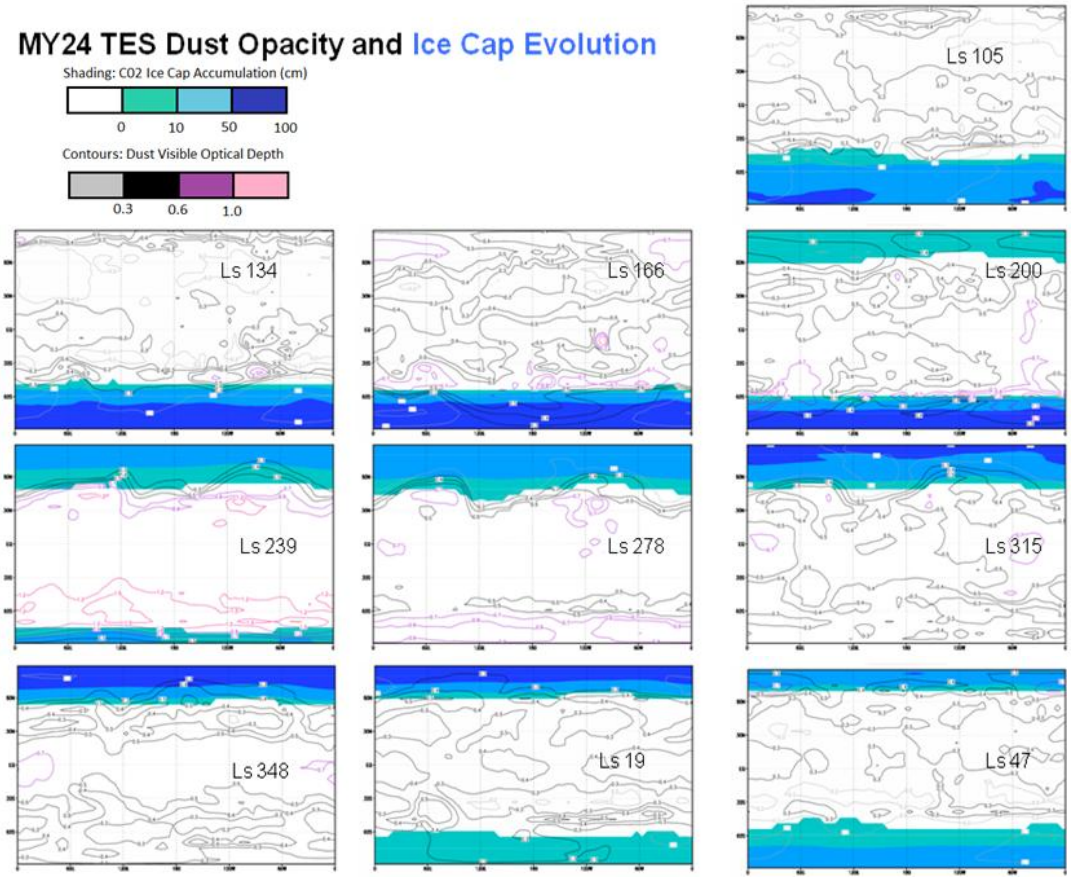


**Figure 4.10: Seasonal evolution of the $CO_2$ polar ice caps (shaded) and TES dust opacities (contours) through a MGCM simulated Mars Year 24. Each successive panel (latitude vs. longitude) represents the passage of 60 sols. Shading intervals represent accumulated $CO_2$; units approximately correspond to depth, in cm. Contours represent column dust opacity normalized to a surface pressure of 6.1 hPa; purple and pink contours correspond to high opacities.**

the SH mid-latitudes near the cap edge, regions of the NH arctic, as well as some local dust storms in the tropics where opacities exceeded 0.7. Throughout the year, dust levels cycle from a minimum in NH summer solstice season to a maximum prior to NH winter solstice.

Figure 4.11 shows zonal mean bred vectors for an annual simulation based upon a "TES dust" MGCM configuration. They can be compared with the original BVs in Figure 2.3 (repeated here as Figure 4.12). The transition seasons of early and

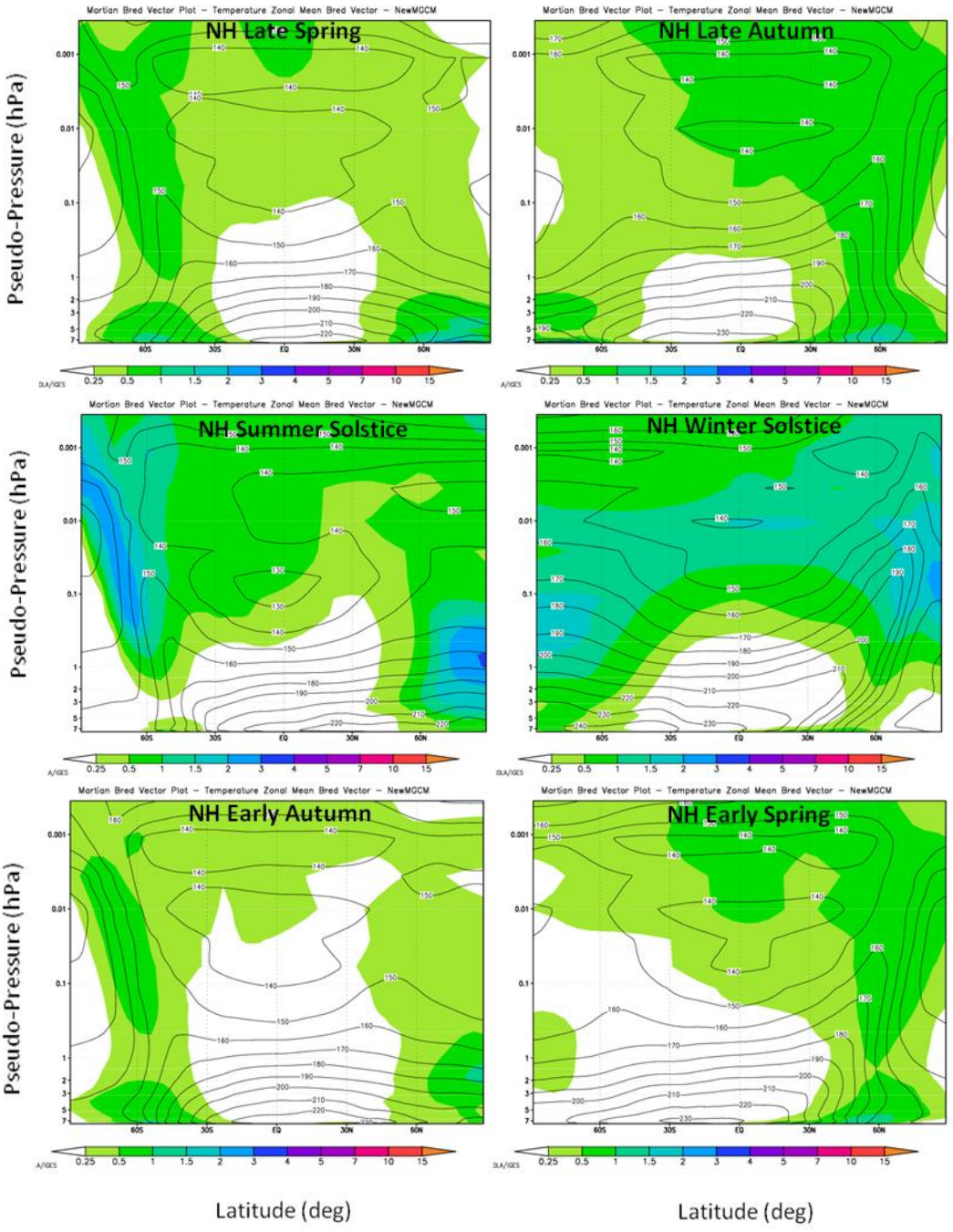


**Figure 4.11: Latitude-pseudopressure plots of bred vector structure as a temporal average for each Martian "season," using a "TES" dust distribution. Shading represents zonal mean BV amplitude; contours denote zonal mean temperatures.**

late autumn and spring are similar in both simulations, with bred vector activity concentrated along the polar temperature gradients, although near surface instabilities

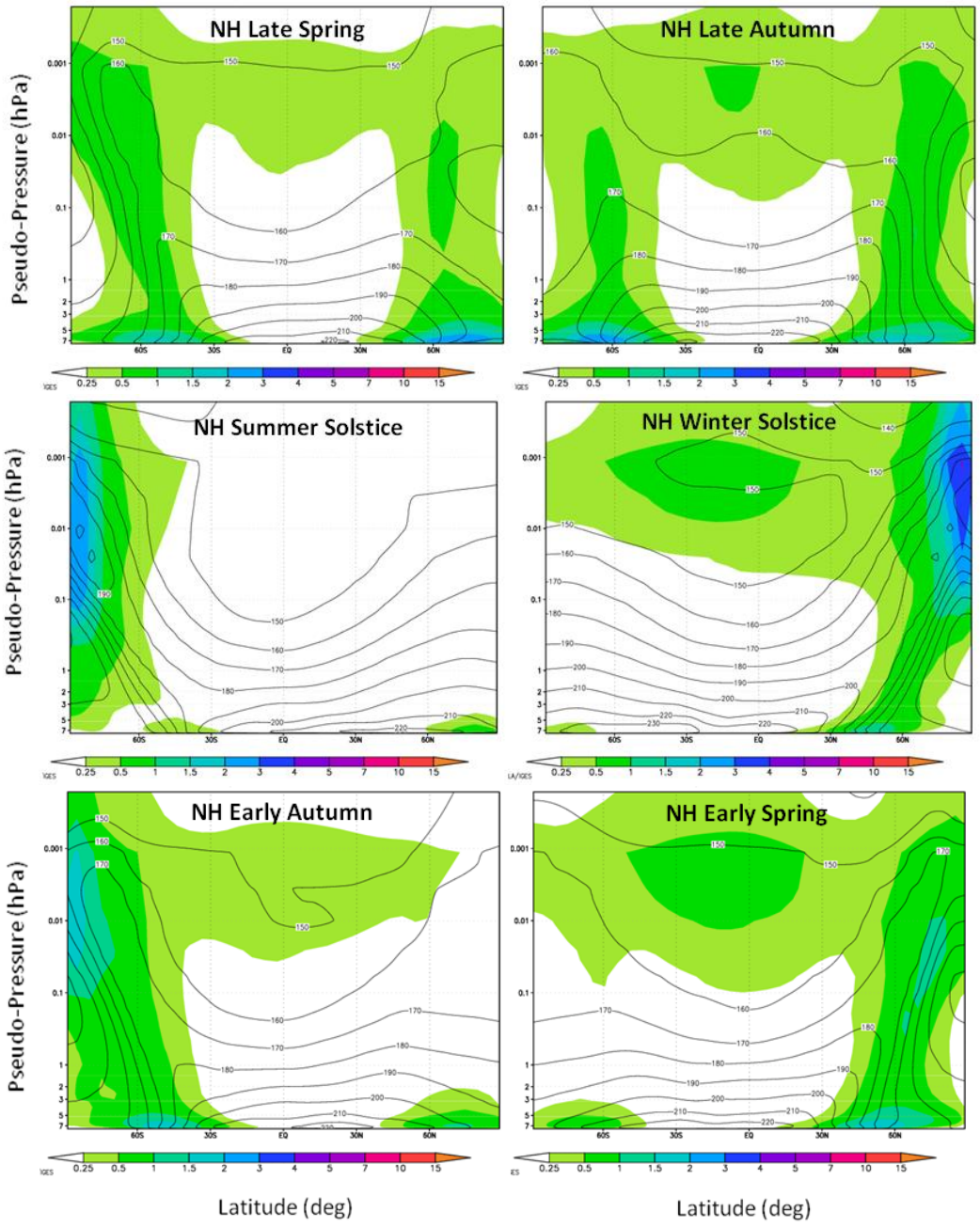


**Figure 4.12: Latitude-pseudopressure plots of bred vector structure as a temporal average for each Martian "season" using a fixed 0.3 dust opacity. Plots are the same as Figure 2.3, and have been repeated for easy comparison with Figure 4.11.**

are slightly weaker with TES dust. However, the solstice seasons feature expanded regions of instabilities extending from the winter instability maxima aloft, through the upper level tropics, to the mid levels of the summer hemisphere. Causes of these

differences are the subject of future research. BV growth rates (not shown) are largest during NH winter solstice, when dust opacity is at its global maximum.

The next major step toward creating a Mars weather and climate analysis is the assimilation of dust aerosol. TES provides column opacities, and MCS provides vertical profiles of aerosol concentration; we first focus on TES observations.

Wilson et al. (2011) showed that surface brightness temperatures are effective at determining opacity by interpolating from an ensemble of MGCM simulations with differing opacity values; this is particularly useful for regions of cold surface temperatures or high opacities where direct TES observations are less trustworthy. This technique naturally extends to the LETKF, which can readily create dust opacity analyses based upon observations of several variable types: either TES opacities, brightness temperatures, or both. The MGCM generates column opacities and brightness temperatures for each ensemble member, which are compared to observations. Variable localization (Kang et al., 2010) allows the user to only keep meaningful cross-variable ensemble correlations. For example, surface brightness temperature could update dust and lower atmosphere temperatures, but not winds. Variable localization parameters could be tuned over a series of assimilation experiments. Finally, the updated opacity field needs to be reintroduced to the MGCM. As in the "TES dust" configuration, the MGCM could inject or remove dust from the boundary layer so that three-dimensional tracer fields add up to the two-dimensional opacities. These tracers are then advected by the wind, and modified by settling and, potentially, surface dust lifting to produce the background opacity fields

for the next analysis. Maintaining a good ensemble spread of plausible dust distributions is critical to success.

As insights from MCS observations shed additional light on dust and water ice vertical distributions, the realism of the ensemble improves and biases are reduced. Eventually, MCS dust and water ice profiles could be directly assimilated as 3D fields. Maps of water ice clouds from the MARCI camera (Wolff et al., 2011) could also be used for assimilation or verification. Finally, convective and surface wind stress lifting parameterizations could update the dust field dynamically in the absence of recent observations, and the MGCM can be evaluated for its skill in initiating, evolving, and dissipating dust storms.

### *4.5 Summary and Future Work*

Using the Local Ensemble Transform Kalman Filter (LETKF), we have assimilated both Thermal Emission Spectrometer (TES) and Mars Climate Sounder (MCS) temperature profiles into the GFDL Mars Global Circulation Model (MGCM). Short term (0.25 sol) forecasts from the resulting analyses are compared to independent (i.e., not yet used by the system) observations, and these forecasts reveal a clear improvement over a freely running model. For TES, errors are generally less than 3 K, except for the SH cap edge baroclinic zone, which has been shown to be very unstable by a bred vector analysis. The low level tropics show the observations to be consistently warmer than the model; in this stable region that is primarily forced by radiation, ensemble spread is too small, and the assimilation system is overconfident in the background. The use of a varying dust distribution among ensemble members, as well as adaptive inflation, led to improvement in ensemble

spread and reduction in RMSE. Empirical bias correction based on time mean analysis increments led to further reduction in model biases.

Analyses from LETKF TES assimilation have been compared to the UK reanalysis. Lower atmosphere temperatures generally agree to within 5 K, and Hovmoller diagrams of traveling waves in the mid-latitudes agree on the phase of the waves, although there are some differences in amplitude. When evaluating the systems against independent radio science temperature profiles in the NH subarctic, the LETKF analyses have a smaller RMSE and bias than the UK analyses.

The bred vector (BV) technique exhibits the seasons and locations of instabilities in the Martian atmosphere. There are active (NH late autumn through early spring) and quiet (NH summer solstice) seasons. Instabilities tend to arise before the initiation of traveling wave activity in each hemisphere. A kinetic energy (KE) budget for the BVs gives insights to the physical origins of the instabilities. Near surface BVs on the mid-latitude polar front gain KE baroclinically from BV potential energy, whereas instabilities aloft along the westerly jets gain KE barotropically from the control state while losing energy baroclinically in some areas to the BV temperature field. Seasonal mean BV activity exhibits local maxima downstream of topography, consistent with a mechanism of lee cyclogenesis. Finally, BVs are linked to ensemble spread from assimilation and the growth of forecast errors.

Assimilation provides insights for improving the MGCM. For example, biases between short term forecasts and MCS observations, particularly at levels above 30 km, motivate the need for improving vertical distribution of dust and water

ice aerosols in the MGCM. Dust aerosol assimilation and representation in models will be a focus of future research. Assimilation can also help the refinement of retrieval algorithms. A parallel project led by Janusz Eluszkiewicz and Matt Hoffman (Hoffman et al., 2011) is using the Optimal Spectral Sampling method (OSS; Moncet et al., 2007) to create new TES retrievals that correct some shortcomings of the originals, and include averaging kernels. Comparison of these profiles to the MGCM through assimilation has helped iteratively refine the profiles; for example, early retrievals showed a larger diurnal cycle than the MGCM, which was corrected by separating the surface temperature retrieval from the lower atmosphere following a suggestion from David Kass (JPL). Eventually, these profiles will use the approach of R. Hoffman (2010) to remove vertical correlations and the influence of the prior from the observations.

Multi-sol forecasts initialized from LETKF analyses help to characterize the predictability horizon for Mars. In NH autumn in the absence of a major dust storm, and using bias correction, forecasts of temperature, wind, and surface pressure were superior to a freely running model when using analyses as a baseline. Given the promise for some useful predictability, a near real time Numerical Weather Prediction at higher resolution has been proposed for Mars using MCS observations in support of the 2012 Mars Science Laboratory mission. In addition to its purely scientific insights, a Mars weather and climate reanalysis and the accompanying improvement in general circulation models will hopefully prove valuable to future missions to Mars, both for planning and for operations.

# Appendix A: Bred Vector Kinetic Energy Equation

Here we derive the bred vector kinetic energy tendency equation in sigma coordinates. For notation, we use a subscript *c* to refer to the control (nature) run, and *b* for the bred vector (perturbation). Using shorthand, the perturbed state (sum of the control and bred vector) is written without subscript ($x_c + x_b = x$).

Begin from the momentum equation in sigma coordinates (Holton, 2003), applied to the control state,

$$\frac{\partial \boldsymbol{v}_c}{\partial t} + \boldsymbol{v}_c \bullet \nabla \boldsymbol{v}_c + \dot{\sigma}_c \frac{\partial \boldsymbol{v}_c}{\partial \sigma} + f\boldsymbol{k} \times \boldsymbol{v}_c = -\nabla \Phi_c + \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma},$$

(A1)

and to the perturbed state,

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v} \bullet \nabla \boldsymbol{v} + \dot{\sigma} \frac{\partial \boldsymbol{v}}{\partial \sigma} + f\boldsymbol{k} \times \boldsymbol{v} = -\nabla \Phi + \frac{\sigma}{p_s} \nabla p_s \frac{\partial \Phi}{\partial \sigma}.$$

(A2)

Note that a friction term could be added trivially. Taking the difference between A2 and A1 yields the BV momentum equation:

$$\frac{\partial \boldsymbol{v}_b}{\partial t} + \boldsymbol{v} \bullet \nabla \boldsymbol{v}_b + \boldsymbol{v}_b \bullet \nabla \boldsymbol{v}_c + \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} + \dot{\sigma} \frac{\partial \boldsymbol{v}_b}{\partial \sigma} + f\boldsymbol{k} \times \boldsymbol{v}_b = -\nabla \Phi_b + \frac{\sigma}{p_s} \nabla p_s \frac{\partial \Phi}{\partial \sigma} - \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma}.$$

(A3)

Bred vector kinetic energy is defined as $K_b = 1/2(\boldsymbol{v}_b \bullet \boldsymbol{v}_b)$. Taking the dot product of $\boldsymbol{v}_b$ with (A3) yields the kinetic energy equation for the bred vector

$$\frac{\partial K_b}{\partial t} + \boldsymbol{v} \bullet \nabla K_b + \boldsymbol{v}_b \bullet (\boldsymbol{v}_b \bullet \nabla)\boldsymbol{v}_c + \boldsymbol{v}_b \bullet \left( \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} \right) + \dot{\sigma} \frac{\partial K_b}{\partial \sigma} + \boldsymbol{v}_b \bullet (f\boldsymbol{k} \times \boldsymbol{v}_b) =$$
$$-\boldsymbol{v}_b \bullet \nabla \Phi_b + \boldsymbol{v}_b \bullet \left( \frac{\sigma \nabla p_s}{p_s} \frac{\partial \Phi}{\partial \sigma} - \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma} \right),$$

(A4)

Note that the dot product $\boldsymbol{v}_b \bullet (f\boldsymbol{k} \times \boldsymbol{v}_b)$ is between orthogonal vectors, and thus zero.

Grouping terms,

$$\frac{\partial K_b}{\partial t} + \mathbf{v} \bullet \nabla K_b + \dot{\sigma} \frac{\partial K_b}{\partial \sigma} =$$
$$-\mathbf{v}_b \bullet (\mathbf{v}_b \bullet \nabla)\mathbf{v}_c - \mathbf{v}_b \bullet \left( \dot{\sigma}_b \frac{\partial \mathbf{v}_c}{\partial \sigma} \right) - \mathbf{v}_b \bullet \nabla \Phi_b + \mathbf{v}_b \bullet \left( \frac{\sigma \nabla p_s}{p_s} \frac{\partial \Phi}{\partial \sigma} - \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma} \right)$$

(A5)

The LHS is the transport of BV KE by the total flow.

Noting that $\nabla \bullet (\boldsymbol{v}_b \Phi_b) = \boldsymbol{v}_b \bullet \nabla \Phi_b + \Phi_b \nabla \bullet \boldsymbol{v}_b$, then

$$RHS = -\mathbf{v}_b \bullet (\mathbf{v}_b \bullet \nabla)\mathbf{v}_c - \mathbf{v}_b \bullet \left( \dot{\sigma}_b \frac{\partial \mathbf{v}_c}{\partial \sigma} \right) - \nabla \bullet (\mathbf{v}_b \Phi_b) + \Phi_b \nabla \bullet \mathbf{v}_b + \mathbf{v}_b \bullet \left( \frac{\sigma \nabla p_s}{p_s} \frac{\partial \Phi}{\partial \sigma} - \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma} \right).$$

(A6)

Noting that $\nabla \bullet (\boldsymbol{v}_b p_{sb}) = \boldsymbol{v}_b \bullet \nabla p_{sb} + p_{sb} \nabla \bullet \boldsymbol{v}_b$, we obtain

$$RHS = -\boldsymbol{v}_b \bullet (\boldsymbol{v}_b \bullet \nabla)\boldsymbol{v}_c - \boldsymbol{v}_b \bullet \left( \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} \right) - \nabla \bullet (\boldsymbol{v}_b \Phi_b) + \frac{\Phi_b}{p_{sb}} \left( \nabla \bullet (\boldsymbol{v}_b p_{sb}) - \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \ldots$$

.

(A7)

We note the BV continuity equation in sigma coordinates (Holton, 2003):

$$\frac{\partial p_{sb}}{\partial t} + \nabla \bullet (\boldsymbol{v}_b p_{sb}) + p_{sb} \frac{\partial \dot{\sigma}_b}{\partial \sigma} = 0$$

.

(A8)

Substituting in the continuity equation yields

$$RHS = -\boldsymbol{v}_b \bullet (\boldsymbol{v}_b \bullet \nabla)\boldsymbol{v}_c - \boldsymbol{v}_b \bullet \left( \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} \right) - \nabla \bullet (\boldsymbol{v}_b \Phi_b) - \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + p_{sb} \frac{\partial \dot{\sigma}_b}{\partial \sigma} + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \ldots$$

(A9)

Applying the product rule $\frac{\partial \Phi_b \dot{\sigma}_b}{\partial \sigma} = \dot{\sigma}_b \frac{\partial \Phi_b}{\partial \sigma} + \Phi_b \frac{\partial \dot{\sigma}_b}{\partial \sigma}$,

$$RHS = -\boldsymbol{v}_b \bullet (\boldsymbol{v}_b \bullet \nabla)\boldsymbol{v}_c - \boldsymbol{v}_b \bullet \left( \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} \right) - \nabla \bullet (\boldsymbol{v}_b \Phi_b) - \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + \frac{p_{sb}}{\Phi_b} \left( \frac{\partial \Phi_b \dot{\sigma}_b}{\partial \sigma} - \dot{\sigma}_b \frac{\partial \Phi_b}{\partial \sigma} \right) + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \ldots$$

(A10)

Reordering terms,

$$RHS = -\left[ \boldsymbol{v}_b \bullet \left( (\boldsymbol{v}_b \bullet \nabla)\boldsymbol{v}_c + \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} \right) \right] - \left[ \nabla \bullet (\boldsymbol{v}_b \Phi_b) + \frac{\partial \dot{\sigma}_b \Phi_b}{\partial \sigma} \right] + \left( \dot{\sigma}_b \frac{\partial \Phi_b}{\partial \sigma} \right) - \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \ldots$$

$$\ldots + \boldsymbol{v}_b \bullet \left( \frac{\sigma \nabla p_s}{p_s} \frac{\partial \Phi}{\partial \sigma} - \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{\partial \Phi_c}{\partial \sigma} \right)$$

(A11)

Use the sigma-coordinate hydrostatic relation $\frac{\partial \Phi}{\partial \sigma} = -\frac{RT}{\sigma}$,

$$RHS = \ldots \left( -\dot{\sigma}_b \frac{RT_b}{\sigma} \right) - \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \boldsymbol{v}_b \bullet \left( -\frac{\sigma \nabla p_s}{p_s} \frac{RT}{\sigma} + \frac{\sigma}{p_{sc}} \nabla p_{sc} \frac{RT_c}{\sigma} \right),$$

(A12)

and the ideal gas law $p\alpha = RT$,

$$RHS = \ldots \left( -\dot{\sigma}_b \frac{p_b \alpha_b}{\sigma} \right) - \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \boldsymbol{v}_b \bullet \left( -\frac{p\alpha \nabla p_s}{p_s} + \frac{p_c \alpha_c}{p_{sc}} \nabla p_{sc} \right).$$

(A13)

Simplifying terms,

$$RHS = \ldots \left[ -\dot{\sigma}_b \alpha_b p_{sb} \right] - \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \boldsymbol{v}_b \bullet \left( -\sigma \alpha \nabla p_s + \sigma_c \alpha_c \nabla p_{sc} \right).$$

(A14)

The final BV kinetic energy tendency equation is given as:

$$\frac{\partial K_b}{\partial t} + \left[ \boldsymbol{v} \bullet \nabla K_b + \dot{\sigma} \frac{\partial K_b}{\partial \sigma} \right] + \left[ \nabla \bullet (\boldsymbol{v}_b \Phi_b) + \frac{\partial \dot{\sigma}_b \Phi_b}{\partial \sigma} \right] = -\left[ \dot{\sigma}_b \alpha_b p_{sb} \right] - \left[ \boldsymbol{v}_b \bullet \left( (\boldsymbol{v}_b \bullet \nabla) \boldsymbol{v}_c + \dot{\sigma}_b \frac{\partial \boldsymbol{v}_c}{\partial \sigma} \right) \right]$$
$$- \frac{\Phi_b}{p_{sb}} \left( \frac{\partial p_{sb}}{\partial t} + \boldsymbol{v}_b \bullet \nabla p_{sb} \right) + \boldsymbol{v}_b \bullet \left( -\sigma \alpha \nabla p_s + \sigma_c \alpha_c \nabla p_{sc} \right) .$$

(A15)

The first term on the LHS is transport of BV kinetic energy by the total flow, which includes a local time derivative and advection. The second bracketed term refers to pressure work. The third term is baroclinic conversion, and the fourth is barotropic conversion; both are similar to terms arising in a comparable derivation for pressure coordinates. The additional terms arise due to coordinate transformation to sigma coordinates.

# Appendix B: Balance and Ensemble Kalman Filter Localization Techniques

*Published in Monthly Weather Review, Vol. 19, p. 511-522 with coauthors E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt.*

In Ensemble Kalman Filter data assimilation, localization modifies the error covariance matrices to suppress the influence of distant observations, removing spurious long distance correlations. In addition to allowing efficient parallel implementation, this takes advantage of the atmosphere's lower dimensionality in local regions. There are two primary methods for localization. In B-localization, the background error covariance matrix elements are reduced by a Schur product so that correlations between grid points that are far apart are removed. In R-localization, the observation error covariance matrix is multiplied by a distance-dependent function, so that far away observations are considered to have infinite error. Successful numerical weather prediction depends upon well-balanced initial conditions to avoid spurious propagation of inertial-gravity waves. Previous studies note that B-localization can disrupt the relationship between the height gradient and the wind speed of the analysis increments, resulting in an analysis that can be significantly ageostrophic.

This study begins with a comparison of the accuracy and geostrophic balance of EnKF analyses using no localization, B-localization, and R-localization with simple one-dimensional balanced waves derived from the shallow water equations, indicating that the optimal length scale for R-localization is shorter than for B-localization, and that for the same length scale R-localization is more balanced. The comparison of localization techniques is then expanded to the SPEEDY global

atmospheric model. Here, natural imbalance of the slow manifold must be contrasted with undesired imbalance introduced by data assimilation. Performance of the two techniques is comparable, also with a shorter optimal localization distance for R-localization than for B-localization.

## *B.1 Introduction*

The Ensemble Kalman Filter (EnKF) (Evensen, 1994) is a Monte-Carlo approximation to the traditional filter of Kalman (1960) that is suitable for high-dimensional problems such as Numerical Weather Prediction (NWP). One of the strengths of Ensemble Kalman Filters is the ability to evolve in time estimates of forecast error covariance, using the flow-dependent information inherent in an ensemble of model runs.

Localization is a technique by which the impact of observations from distant regions upon an analysis is suppressed. There are two categories of localization techniques (discussed in detail in section 2b): those that operate on background error covariances $\boldsymbol{B}$, which we call B-localization, and those that operate on observation error covariances $\boldsymbol{R}$, which we call R-localization. Adaptive localization techniques, such as the hierarchical filter of Anderson (2007) and ECO-RAP of Bishop and Hodyss (2009a, 2009b), are beyond the scope of this work.

It is the error covariances between model variables, along with the observation error characteristics, that ultimately describe the impact pattern of an observation upon the analysis via the Kalman gain $\boldsymbol{K}$. In practice, the accuracy of the background error covariance estimate is limited by the size of the ensemble, which must be kept small for computational feasibility (typically of order 20–100 for NWP).

Empirically, at larger geographical distances background error covariance estimates tend to be dominated by noise rather than signal (Hamill et al., 2001); it is this "distance-dependent assumption" that motivates the technique of (non-adaptive) localization to eliminate correlations that are deemed to be spurious.

The background error covariance determined from an ensemble of *P* members has at most *P-1* degrees of freedom to express uncertainty. However, in local regions of large error growth the atmosphere has been shown to exhibit low dimensionality (Patil et al., 2001). When using localization, the ensemble needs to account for the instabilities in a local region. Additionally, if local analyses can choose different linear combinations of ensemble members in different regions, this allows the analysis to greatly reduce the previously noted dimensionality limitation (Hunt et al., 2007). Lorenc (2003) notes that the assimilation of a perfect observation removes a degree of freedom from the ensemble, but that localization with a Schur product allows for extra degrees of freedom in the analysis.

Localization can also lead to significant savings in computational resources. The analysis at each grid point only needs to consider local observations and the values at nearby model grid points that are linked to these observations by the observation operator. Analyses for local regions can thus be considered independently, allowing for more efficient parallelization of the code (Hunt et al., 2007; Szunyogh et al., 2008).

Successful NWP depends upon well-balanced initial conditions to avoid the generation of spurious inertial gravity waves such as those that ruined the 1922 Richardson forecast. By balanced, we mean an atmospheric state in the slow

manifold that approximately follows physical balance equations appropriate to the scale and location, such as the geostrophic relationship. In practice, there are initialization techniques for improving the balance of an analysis, such as nonlinear normal mode initialization and digital filters (Lynch and Huang, 1992). However, once an analysis is filtered the resulting atmospheric state cannot be guaranteed to be optimal. Daley (1991, chapter 6) notes that there is no unique balanced state corresponding to a given unbalanced state; a filter may merely ignore the increment and move the solution back toward the balanced background state! Thus an ideal data assimilation system should avoid or reduce the initialization by filtering and try to create a well-balanced analysis.

The impact of localization on the balance of an analysis is discussed in Cohn et al. (1998) who noted an unrealistically high ratio of divergence to vorticity as a consequence of local observation selection. Mitchell et al. (2002) show that the optimum localization distance (in terms of improving analysis error) grows with ensemble size, and that balance is improved with longer localization distances. Lorenc (2003) provides an example of how localization produces imbalance. Consider the assimilation of a single height observation located at the origin ($x$=0) of Figure B.1. The solid lines in Figure B.1 represent a perfect scenario where the height $h$ and meridional wind $v$ are in geostrophic balance in the context of the shallow water equations (see Section B.2 for details). The black line is proportional to the error covariances between $h$ at the location $x$ and $h$ at the origin, while the gray line is proportional to the error covariances between $v$ at $x$ and $h$ at the origin. In the assimilation of a single $h$ observation, these lines are also proportional to the

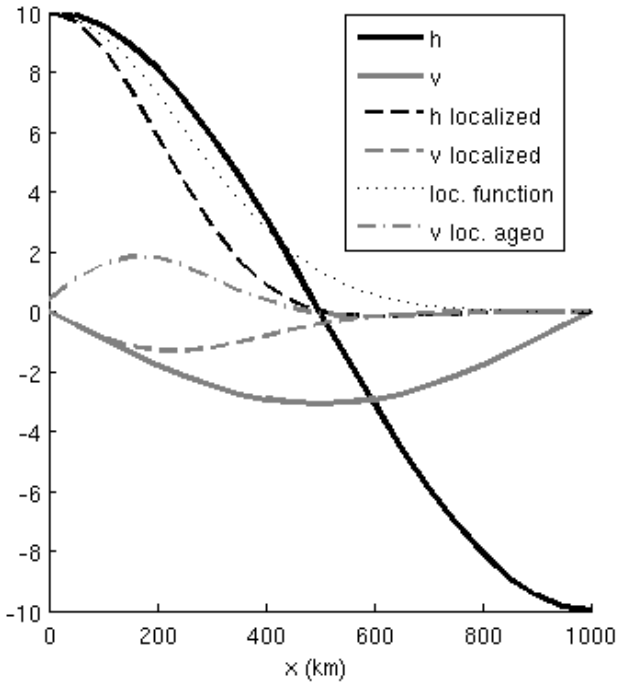


**Figure B.1: Example showing the introduction of imbalance by localization (after Lorenc, 2003). Waveforms of height (black) and meridional wind (gray) before (solid) and after (dashed) multiplication by a Gaussian localization function (dotted). Values on the y-axis denote the size of the analysis increment (m; m $s^{-1}$) from the assimilation of a height observation located at the origin. The ageostrophic portion of the wind increment after localization is dash-dotted.**

respective elements of the Kalman Gain matrix ***K***, and therefore the analysis increments. Localization is then applied to these error covariances by multiplying them by a Gaussian function with length scale 250 km based upon distance from the observation, so that the error covariances decay to zero for larger $x$ (dashed lines). In the region of $x$ = 250 km, the analysis increment of $v$ is reduced by localization. If geostrophic balance is to be maintained, then the magnitude of the height gradient with respect to $x$ should also be smaller. However, the height gradient is actually increased by localization and therefore the wind becomes significantly ageostrophic in this region (dash-dot line). In general, EnKF covariance localization modifies the elements of either the ***B*** matrix or the ***R*** matrix, which in turn reduces the elements of

$\boldsymbol{K}$ as one moves further from the observation. Thus, as in this example, the analysis increments asymptote to zero as the analysis converges to the background in the absence of observation information. During this transition the geostrophic balance of the analysis increment is disrupted.

Kepert (2009) demonstrates how assimilation of wind and height observations with localized covariances produce imbalanced analyses with excess divergence, and proposes assimilation in terms of streamfunction $\psi$ and velocity potential $\chi$ rather than *u* and *v* wind components. This technique results in a smaller (and more natural) ratio of divergence to rotation in the analysis, and hence balance is improved, but these improvements are less noticeable after initialization.

The purpose of this paper is to compare the B- and R-localizations and their impact on balance. Following a description of the EnKF and localization techniques (Section B.2), we first compare the localizations using a simple model (Section B.3), and then apply them to a global atmospheric model (Section B.4).

## *B.2 Methods*

### *B.2.1 Ensemble Kalman Filter Data Assimilation*

The data assimilation cycle consists of a forecast stage, where the estimate of the state is evolved in time using a model, and an analysis stage, where the estimate of the state $\boldsymbol{x}_a$ is improved through optimal combination of forecast $\boldsymbol{x}_b$ and observations $\boldsymbol{y}_o$.

$$\boldsymbol{x}_{\mathrm{a}} = \boldsymbol{x}_{\mathrm{b}} + \boldsymbol{K}\,(\boldsymbol{y}_{\mathrm{o}} - h_{op}(\boldsymbol{x}_{\mathrm{b}})) \qquad \text{(B.1)}$$

The optimal weight matrix $\boldsymbol{K}$, or Kalman Gain, is given by

$$\boldsymbol{K} = \boldsymbol{B}\,\boldsymbol{H}^{\mathrm{T}}\,(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1} \qquad \text{(B.2)}$$

where $\boldsymbol{B}$ is the background error covariance matrix, $\boldsymbol{R}$ the observation error covariance matrix, and $\boldsymbol{H}$ the linearization of the observation operator $h_{op}$. In ensemble data assimilation methods, the background error covariance matrix is estimated using an ensemble of $P$ forecasts:

$$\boldsymbol{B} = \frac{1}{P-1}\boldsymbol{X}_b \boldsymbol{X}_b^{\,T} \qquad \text{(B.3)}$$

where $\boldsymbol{X_b}$ is the matrix of background ensemble perturbations from the ensemble mean with each row referring to a model variable, and each column to an ensemble member. The exact technique for updating the analysis ensemble members depends on the version of EnKF.

### *B.2.2 Localization Techniques*

For B-localization, the $\boldsymbol{B}$ matrix is multiplied elementwise (i.e., through a Schur product) by another matrix $\boldsymbol{C}$ whose elements represent some localization function $f_{loc}$ of distance $d$ between grid points $i$ and $j$ (Houtekamer and Mitchell, 2001). Gaspari and Cohn (1999) describe a Gaussian localization function:

$$f_{Bloc} = \exp\left[\frac{-d(i,j)^2}{2L^{\,2}}\right] \qquad \text{(B.4)}$$

where $L$ is a localization distance used for scaling the width of the localization. Gaspari and Cohn (1999) also introduced a piecewise polynomial approximation of a Gaussian localization function with compact support (this means it becomes zero beyond some finite distance, in this case at about 3.65 times $L$). Physically, this

means that the background errors at model grid points that are far apart should have no statistical relationship.

With R-localization, modifications are made to the observation information. The simplest technique is through observation selection, by excluding observations that lie beyond a cutoff radius from the analysis (as in Houtekamer and Mitchell, 1998). However, abrupt localization cutoff can result in a noisy analysis. Hunt et al. (2007) proposed a gradual R localization by multiplying the elements of $\boldsymbol{R}$ by an increasing function of distance from the analysis grid point. Here we use the positive exponential function:

$$f_{Rloc} = \exp\left[\frac{+d(i,j)^2}{2L^2}\right] \qquad \text{(B.5)}$$

With uncorrelated observation error (which is a reasonable assumption for many instruments), $\boldsymbol{R}$ is diagonal. Then in (B.5), $d$ is the distance between observation $i$ and model grid point $j$. Since $d$ varies depending upon which grid point the analysis is being performed at, the rows of $\boldsymbol{K}$ (in Equation B.2) must be computed independently because the ($\boldsymbol{HBH}^T + \boldsymbol{R}$) term will be different at each grid point location. Physically, this means that far-away observations can be considered to have infinite error, and thus do not impact the analysis.

R-localization rather than B-localization is necessary for the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007), because as the calculations are done in ensemble space, the $\boldsymbol{B}$ matrix is not represented explicitly in physical space. The formulation of the Kalman gain (B.2) can be stated for the LETKF as

$$\mathbf{K} = \mathbf{X}_b\left[(P-1)\mathbf{I}_p + (\mathbf{H}\mathbf{X}_b)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}_b)\right]^{-1}(\mathbf{H}\mathbf{X}_b)^T\mathbf{R}^{-1} \quad \text{(B.6)}$$

where $\boldsymbol{I}_p$ is the $P$x$P$ identity matrix. For this study, we employ Gaussian localization (equations B.4 and B.5) with a cutoff distance of approximately 3.65 times $L$ beyond which there is no observation impact (the localization function is set to zero). The application of (B.5) to a diagonal $\boldsymbol{R}$ using an observation cutoff radius of 3.65 $L$ puts an upper bound on the conditioning number for $\boldsymbol{R}$ at $10^3$ for the case of uniform observation errors. Localization can also be applied by dividing the diagonal elements of $\boldsymbol{R}^{-1}$ in (B.6) by $f_{\text{Rloc}}$. This reduces the size of the rightmost term of the bracketed expression in (B.6); as this smaller term is then added to the identity matrix, the inversion of the bracketed expression remains a stable calculation. Note that some studies (i.e., Houtekamer and Mitchell, 2005) report localization values in terms of cutoff distance rather than $L$.

For NWP applications, $\boldsymbol{B}$ ($N$ x $N$, where $N$ is the dimension of $x$) is too large to be represented explicitly, therefore the $\boldsymbol{BH}^T$ and $\boldsymbol{HBH}^T$ terms of Equation B.2 are calculated directly from the ensemble, as in Houtekamer and Mitchell (2001). For the serial EnSRF (Whitaker and Hamill, 2002), localization by a distance-dependent function is performed upon $\boldsymbol{BH}^T$, where each element represents the covariance between a model grid point and observation. Because $\boldsymbol{HBH}^T$ is a scalar, it does not require localization. In the case of observations on grid points (which is the case used in this study), this form of localization (on $\boldsymbol{BH}^T$) is equivalent to B-localization. When observations are located off grid points, or relate to more than one grid point, this technique exhibits hybrid properties of B-localization and R-localization. The problem of defining distance for vertically integrated measurements, such as satellite

observations (Campbell et al., 2010), is equally challenging for $\boldsymbol{BH}^T$ and R localization techniques, as both require a distance between an observation and model grid point, and this issue is a motivation for adaptive localization (Anderson 2007; Bishop and Hodyss, 2009). This study focuses on horizontal localization with point observations; vertical localization in the LETKF is addressed in Miyoshi and Sato (2007).

## *B.3 Simple Model Experiments*

The goal of this section is to demonstrate the impact of EnKF localization on balance using a simple model consisting of one-dimensional balanced waveforms. These initially balanced wave solutions (which are not integrated forward in time) serve as truth and background ensemble states for identical twin data assimilation experiments; any disruption to the balance of the resulting analysis is thus easily detectable and attributable to the properties of the EnKF technique.

### *B.3.1 Simple Model Description*

Consider the shallow water momentum equation in the *x*-direction for a rotating (constant Coriolis parameter *f*), inviscid fluid:

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + fv - g\frac{\partial h}{\partial x} \qquad \text{(B.7)}$$

The geostrophic balance between the pressure gradient and Coriolis terms can thus be stated:

$$fv_g = g\frac{\partial h}{\partial x} \qquad \text{(B.8)}$$

Here $v_g$ is the geostrophic wind. Assuming that the wave structure is uniform in the *y*-direction, harmonic form is applied to the perturbation variables to achieve a wave solution for *h*, with $h_{depth}$ being the mean depth of the fluid, $h_{amp}$ the amplitude of the height perturbation, *k* the wavenumber, and $x_{ps}$ a wave phase shift:

$$h = h_{depth} + h_{amp}\cos(k(x - x_{ps})) \qquad \text{(B.9)}$$

Assuming geostrophically balanced wind field, we arrive at the wave solution for *v*:

$$v = -\frac{g}{f}kh_{amp}\sin(k(x - x_{ps})) \qquad \text{(B.10)}$$

For the simple model, consider a one-dimensional non-periodic domain of 5000 km along the *x*-axis, with model grid points spaced regularly at 50-km intervals. The Coriolis parameter *f* was selected to be $10^{-4}$ $s^{-1}$, a reasonable value for the mid-latitudes.

### *B.3.2 Experiment Design*

The truth state and 5 background ensemble members, plotted in Figure B.2, are defined for both height and *v*-component of the wind. Each ensemble member is generated by randomly selecting a height perturbation amplitude from a uniform distribution of [9, 11] m, a wavelength from [1950, 2050] km and phase shift from [-50, 50] km. The truth waveform (amplitude = 10m, wavelength = 2100 km, offset = -100 km) is fixed in order to avoid having a mean background state too close to the ensemble mean. This would be undesirable, as an analysis that moves further from the background toward an observation would be overly penalized, whereas one that remained close to the background would be falsely rewarded. The meridional wind waveform is then generated to be in geostrophic balance with the height waveform.

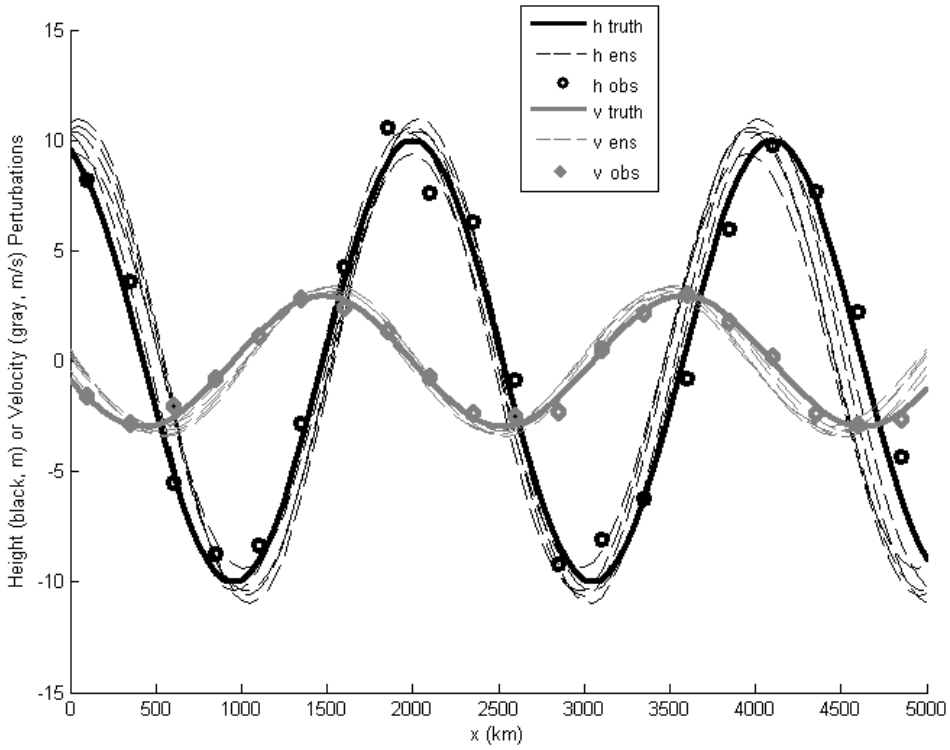


**Figure B.2: Sample experimental setup for the simple model experiment. The black curves represent the height waveform, while the gray represent the meridional wind. Thick solid lines depict the truth waveforms, whereas dashed lines are used for the ensemble members. Black circles are height observations, whereas gray diamonds are wind observations.**

These waves are represented discretely as height and meridional wind values at each of the 101 model grid points. Observations of both *h* and *v* at regularly spaced grid points 250 km apart are chosen based upon the truth value at the corresponding model grid point plus a random observation error equal to 10% of the wave amplitude.

Ensemble mean analyses resulting from assimilation using no localization, B-localization, and R-localization using various localization distances *L* are compared. As the wind can be partitioned into geostrophic and ageostrophic components ($\boldsymbol{v} = \boldsymbol{v_g} + \boldsymbol{v_a}$), the RMS value of $\boldsymbol{v_a}$ over all grid points is used as a summary metric of imbalance; accuracy is also assessed as the RMS difference from the truth. To obtain significant results not dependent upon the peculiarities of a specific random

configuration of ensemble members and observation errors, each configuration is repeated 100 times in a Monte Carlo experiment. Note that the model is not advanced in time, so boundary conditions are not needed.

### *B.3.3 Simple Model Results*

Figure B.3a shows the dependence of RMSE for each analysis as a function of localization distance $L$. LETKF rather than the generic EnKF formula is used for R-localization; the differences in accuracy and balance metrics between LETKF and EnSRF R-localization for this experiment (not shown) are on the order of 1%, so the comparison is fair. R-localization has an optimal scale of $L$ = 500 km, whereas B-localization is close to optimal at around $L$ = 1000 km and larger for 5 ensemble members. A scenario using 40 ensemble members and no localization is also plotted as a best-case performance scenario to which the localized 5-ensemble member analyses aspire. Note that results for $v$-wind error (not shown) are similar. An explanation for the disparity in optimal length scales is provided in Appendix C.

Figure B.3b shows the dependence of RMS imbalance (ageostrophic wind) for each analysis as a function of localization distance $L$. Analyses without localization show no ageostrophic wind, which is to be expected from the design of the experiment. For the localized cases, as the localization distance increases, the analysis becomes more balanced. R-localization is always more balanced than B-localization for the same localization distance $L$, although the levels of imbalance are comparable when considering the optimal configuration of each method.

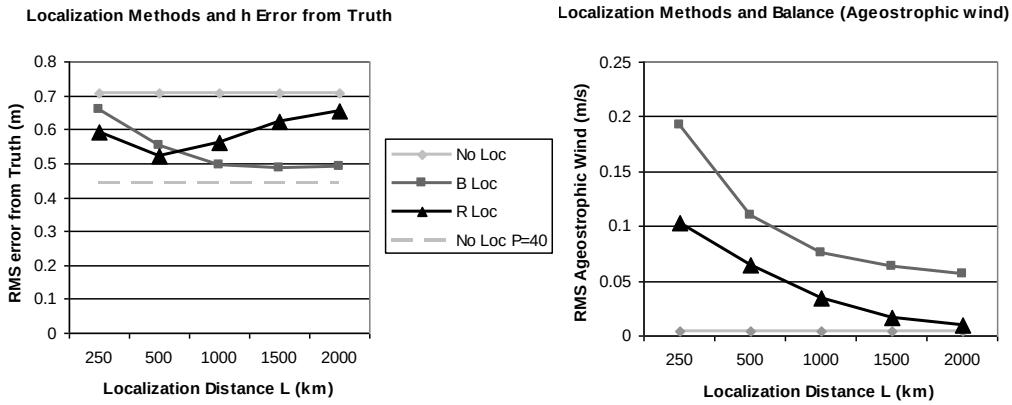


**Figure B.3: RMS error of the analysis from the truth for height (m; left panel) and RMS ageostrophic wind (m $s^{-1}$, right panel) using no localization, B localization, and R localization for 5 ensemble members and a variety of localization distances *L*. For comparison, an analysis with no localization and 40 ensemble members is also plotted. Arrows depict optimum values of *L*.**

## *B.4 SPEEDY Model Experiments*

### *B.4.1 Measuring Balance in a Realistic Model*

In a realistic atmospheric model we can no longer assume that the background state is initially balanced, since an atmosphere with purely geostrophic flow would not allow for interesting weather such as intense baroclinic development and the vertical motion associated with heavy precipitation. Therefore, although much of the energy in the atmosphere is associated with the slow mode (Daley 1991), there is a natural level of imbalance in the atmosphere. The challenge is to differentiate between this background amount of imbalance, and additional spurious amounts introduced as an artifact of data assimilation.

There are several metrics for evaluating atmospheric imbalance. Section B.3 (and Lorenc 2003) uses the magnitude of the ageostrophic wind. While this metric is straightforward to compute, it is not applicable at all latitudes; there are also more sophisticated balance equations, such as nonlinear balance (Raymond, 1992), to

consider. High frequency oscillations can be diagnosed directly by examining the second derivative of the surface pressure field in time (Houtekamer and Mitchell, 2005). Finally, the analysis can be compared to an initialized (filtered) version of itself using a Lynch and Huang (1992) Lanczos digital filter (as in Mitchell et al., 2002) that removes high frequency oscillations, and thus inertial-gravity waves, from the model time series (not included in this study). Similarly, Kepert (2009) used the magnitude of the NNMI increment as a measure of balance. The surface pressure and digital filter metrics require model output from several time steps at a relatively fine temporal resolution (smaller than one hour).

### *B.4.2 Experiment Design*

The Simplified Parametrizations, primitivE-Equation DYnamics, or SPEEDY, model (Molteni, 2003) is an atmospheric global circulation model of intermediate complexity designed for climate experiments. While containing many of the physics components found in larger models (including convection, condensation, cloud, radiation, and surface flux parameterizations), it is computationally inexpensive so it can be run on a single processor. There are seven vertical levels using the sigma coordinate system, with a horizontal spectral resolution of T30, which corresponds to a standard Gaussian grid of 96 by 48 points. The time scheme is leapfrog. There are five dynamical variables included in the output: zonal wind ($u$), meridional wind ($v$), temperature ($T$), specific humidity, and surface pressure ($p_s$). Miyoshi (2005) modified the SPEEDY model for weather forecasting by creating output every six hours, and implemented several data assimilation techniques on the SPEEDY model. Horizontal diffusion (of vorticity, divergence, temperature, specific humidity) in the

SPEEDY model is done with the fourth power of the Laplacian, and is applied on the sigma surfaces. Maximum damping time is 18 hours for temperature and vorticity, and 9 hours for divergence, with an additional 12 hours applied at the top level (representing the stratosphere). There is also vertical diffusion that simulates shallow convection in regions with conditional instability, as well as water vapor and static energy vertical diffusion (Molteni, 2003). Frequency damping with a Robert-Asselin filter (with filter parameter = 0.05) is included in the SPEEDY model to suppress the spurious computational mode. Amezcua et al. (2010) has examined the use of a Robert-Asselin-Williams (RAW) filter (which successfully dampens the computational mode without damping the physical solution; Williams, 2009) with the SPEEDY model, and found that there are very few changes to the model climatology that pass a field significance test, and the quality of the forecasts was slightly improved. This change in the high frequency damping did not seem to affect the model balance. Note that the RAW filter is not employed in the experiments presented in this paper.

The ultimate goal of using the SPEEDY model is a realistic comparison of B-localization and R-localization in terms of balance and accuracy. Here, B-localization is employed with the EnSRF algorithm (Whitaker and Hamill, 2002), whereas R-localization is used with LETKF (Hunt et al., 2007). In addition, a third configuration using the EnSRF with R-localization is employed to investigate whether any differences between the first two configurations are primarily due to variation in localization technique rather than assimilation algorithm (serial versus simultaneous, etc.); see Holland and Wang (2010) for an independent comparison of

EnSRF and LETKF. All systems use identical observations, which are generated as random perturbations from the nature run, or true state, in an identical twin experiment. The observation network used for this study approximately follows the rawinsonde locations (Figure B.7), with all observations located on model grid points. Observations are located at each of the seven model levels. Observation error is 1K for temperature, 1 m/s for *u* and *v* wind magnitudes, 1 g/kg for specific humidity, and 1 mb for surface pressure. Multiplicative inflation of 2 % is applied to the background ensemble spread. Vertical localization is by model level so that an observation corresponding to one of the model's seven levels does not impact any other level; previous experience with the SPEEDY model has shown that vertical correlations for wind and temperature errors are minimal. The ensembles are comprised of 20 members, with initial conditions taken from consecutive dates in January 1982.

The forecast-assimilation cycle is every 6 hours over a period of 48 days from Feb 1 to Mar 20, 1982. The assessment of accuracy is made by comparing the ensemble mean analysis of wind magnitude to the truth at each 6-hour period. Balance is assessed through the magnitude of the ageostrophic wind, as well as the second derivative of surface pressure. These metrics are applied during the month-long period of Feb 20 to Mar 20 following 20 days of spin-up. Wind metrics are obtained from model level 4 (~500 hPa). Results are reported as an areal mean, either globally or over mid-latitude bands (~30° to 60°) separately for the northern hemisphere (NH) and southern hemisphere (SH).

*B.4.3 SPEEDY Model Results*

Figure B.4 shows the accuracy of analyses (measured by mean absolute wind error at ~500 hPa) for the EnSRF B-localization and LETKF R-localization relative to the true state as a function of localization distance parameter *L* (see the discussion surrounding equations B.4 and B.5). The performance of the system is highly dependent upon the choice of localization parameter. Too long a localization distance and the system is dominated by spurious observation increments that prevent it from converging to the truth, whereas too short a localization distance and observations introduce imbalanced increments, as well as fail to adequately impact their neighborhood of grid points. An optimal localization distance parameter *L* with respect to accuracy is 500 km for R-localization, and 750 km for B-localization. Error is higher and the optimal length scale is slightly longer for the SH compared to the

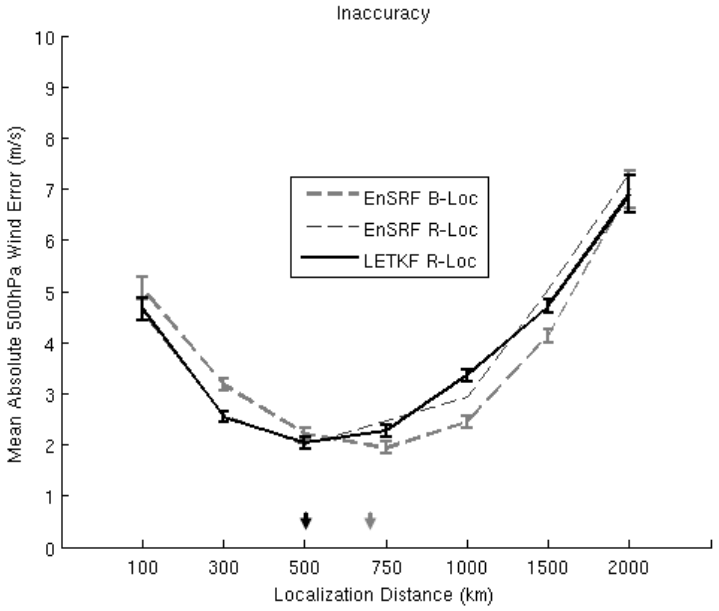


**Figure B.4: Summary of SPEEDY accuracy statistics for B-localization vs. R-localization. Error bars denote standard deviation over time. Arrows denote optimal values of localization distance *L*. For *L* < 500km, EnSRF R-localization and LETKF R-localization give essentially identical results.**

NH (not shown), as the former has a relative paucity of observations. The performance for R-localization in both LETKF and EnSRF is similar, particularly for $L$< 500 km where the results are essentially identical. The results for wind error at other vertical levels (not shown) reveal a similar dependence on localization, with slightly higher errors as altitude increases. Note that the areal mean ensemble spread (not shown) is also highly sensitive to $L$, with shorter $L$ corresponding to greater ensemble spread. Observation information reduces the uncertainty of an analysis; for shorter localization distances this reduction in analysis spread takes place over smaller regions (nearest to the observations), and thus the areal mean ensemble spread remains high.

Figure B.5 reveals the performance of the two systems with respect to balance, measured by the mean magnitude of the ageostrophic wind at ~500 hPa as a

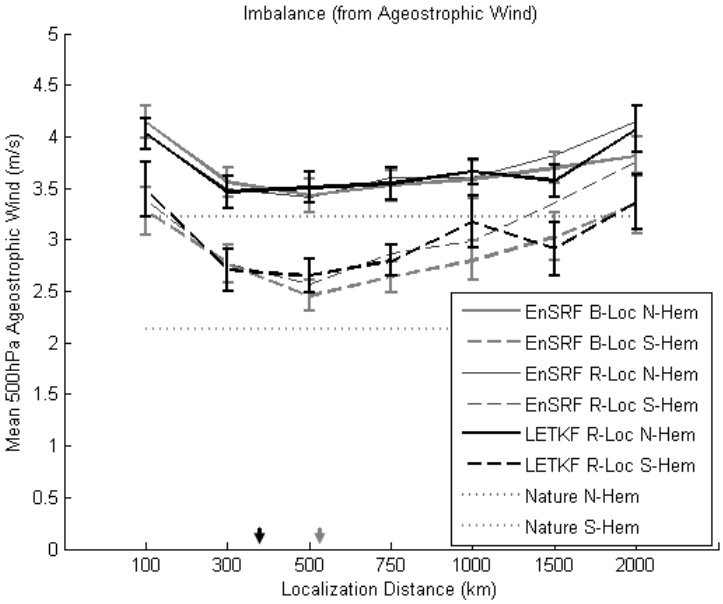


**Figure B.5: Summary of SPEEDY imbalance statistics for B-localization vs. R-localization as measured by the ageostrophic wind (m $s^{-1}$). Natural levels of imbalance are noted as horizontal lines. Error bars denote standard deviation over time. Arrows denote optimal values of localization distance *L*.**

function of the localization distance parameter $L$. There exists a larger natural state of geostrophic imbalance in the NH (~ 3 m/s) compared to the SH (~ 2 m/s) due to the presence of the Himalayan plateau protruding into the mid-latitude belt as well as the fact that the experiment occurred in the NH winter with its stronger wind speeds. In all cases, the imbalance of the analyses is larger than that of the true state, indicating that data assimilation has introduced artificial imbalance. Although the magnitudes of the mean ageostrophic winds are higher for the NH, the difference in imbalance between the nature run and assimilation runs (assimilation-induced imbalance) is greater for the SH. Short localization distances ($L$ < 300 km) are detrimental to balance, which agrees with the results of Section B.3 using a simple model. For very long localization distances ($L$=2000 km), presumed spurious correlations can lead to larger values of both error and imbalance. Examination of performance time series reveal that values of imbalance tend to stabilize, along with the error, after 20 days of spin-up, although there are day-to-day fluctuations on the order of 0.5 m/s that are reflected in both the nature run and assimilation analyses.

Figure B.6 also depicts imbalance, but measured by the second derivative of surface pressure at each model time step. As in Figure B.5, short localization distances (L<300 km) are very harmful to balance. Here, the NH is significantly more balanced than the SH, which agrees with the result for assimilation-induced imbalance in Figure B.5. Optimal values of $L$ are slightly larger using this metric compared to Figure B.5; averaging the optimal $L$ values for both metrics of imbalance results in an optimal $L$ that agrees with the results for accuracy in Figure B.4. The occasional lack of smoothness in the relationship curves between imbalance and $L$ in

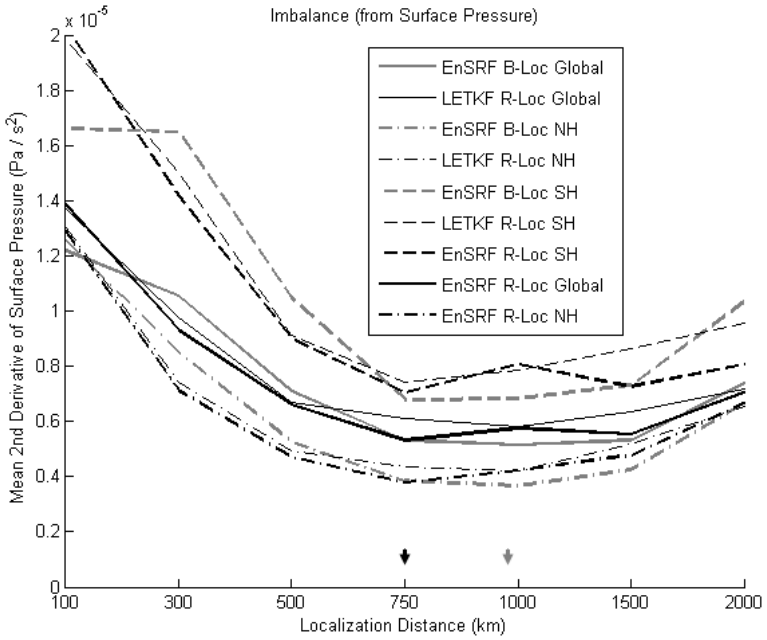


**Figure B.6: Summary of SPEEDY imbalance statistics for B-localization vs. R-localization as measured by the second derivative of surface pressure (Pa $s^{-2}$). Arrows denote optimal values of localization distance *L*.**

Figures B.5-6 reveal that an evaluation time period of at least one month is required to overcome sampling error for these techniques.

Figure B.7 reveals the spatial distribution of imbalance as a time mean over the period from Feb. 20 – Mar. 20. For short localization distances, imbalance is large in the immediate vicinity of observations. For long localization distances, imbalance is smaller and spread over broader areas. This finding agrees with the Lorenc (2003) explanation using Figure B.1 in that imbalance can be introduced in the region where the impact of an observation moves toward zero. The circular patterns of imbalance surrounding the Southern Ocean islands in the case of $L$=250 km demonstrate the detrimental impact of strong localization resulting from a sharp transition between a region with strong observation impact and a region with little observation impact. Imbalance is greatest along the Pacific coast of South America;

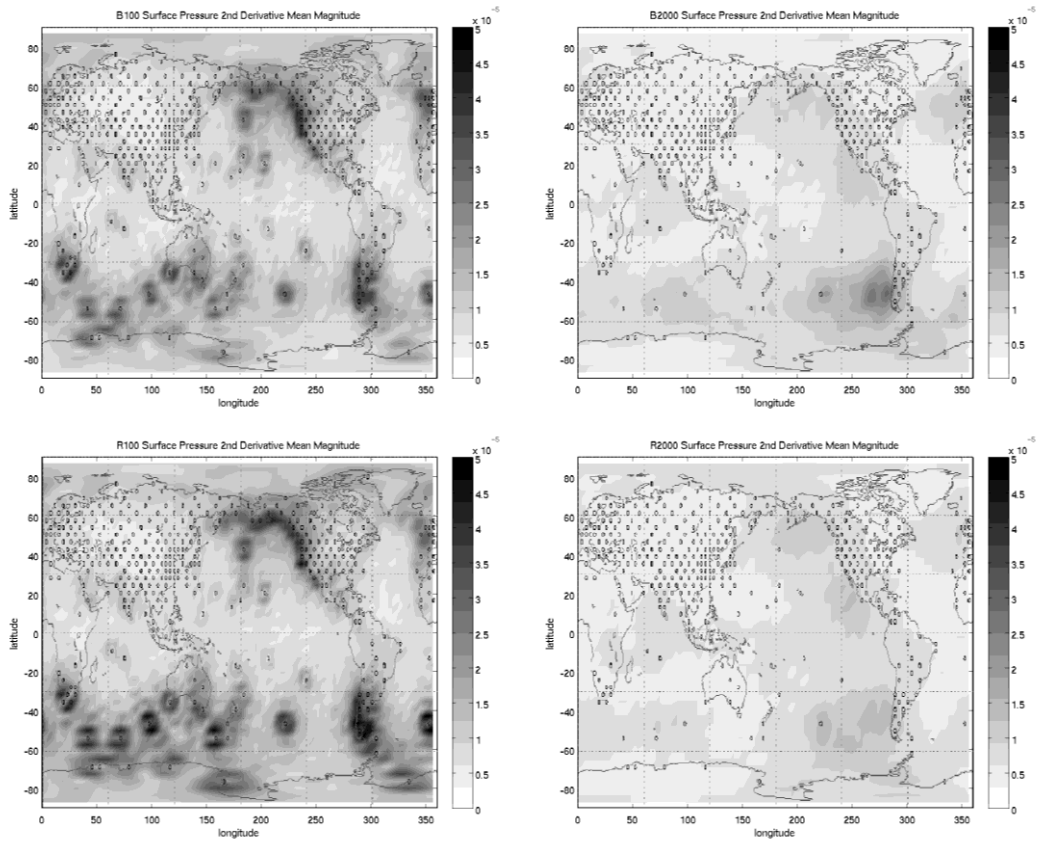


**Figure B.7: Time average spatial distribution of imbalance measured by the second derivative of surface pressure (Pa $s^{-2}$) for short (100 km, left panels) and long (2000 km, right panels) localization distances using EnSRF B-localization (top panels) and LETKF R-localization (bottom panels). Observation locations are depicted by black dots.**

the lack of observations in the South Pacific leads to large observation increments in the region. Inaccurate background fields, which require larger subsequent analysis increments resulting in greater potential for imbalance introduced by data assimilation, may explain the somewhat unexpected increase in imbalance for large *L* in Figures B.5 and B.6.

## *B.5 Conclusions*

This study has examined the impact of EnKF localization techniques upon the accuracy and balance of analyses. Localization is used to combat spurious correlations due to sampling error from finite ensemble size, to take advantage of low

dimensionality in local regions, and for efficient computation. Localization techniques can be classified into two methods: B-localization, where the background error covariance is modified by a distant-dependent localization function, and R-localization, where observation error variances are increased as distance from the analysis grid point increases. Variations of the B-localization technique are appropriate for EnSRF where the entire domain is updated with each observation, whereas R-localization is used for LETKF as the background error covariances are specified in ensemble space and each model grid point is updated independently. In addition to accurately depicting the state of the system, atmospheric data assimilation should produce a balanced analysis so that information is not lost through spurious inertial-gravity wave propagation.

We first described experiments with simple, one-dimensional waveforms based upon the shallow water equations. As the background ensemble is initially balanced, imbalance introduced by data assimilation is easy to measure as the magnitude of the ageostrophic wind. The two techniques have differing optimal localization distances $L$ with respect to analysis accuracy; approximately 500 km for R-localization, and 1000 km or larger for B-localization. For the same localization length R-localization is more balanced than B-localization but the balance of both techniques improves as $L$ grows larger.

We then made a more realistic comparison between EnSRF B-localization and LETKF R-localization involving the global SPEEDY model in identical twin experiments. Here, the background state can no longer be assumed to be in balance. Two methods for evaluating imbalance are used: the magnitude of the ageostrophic

wind, and the second derivative of surface pressure. The two localization techniques are roughly comparable in performance with respect to localization and balance when the optimal length scale of $L$ is selected: 500 km for R-localization, and 750 km for B-localization. This result is consistent with the discussion in the Appendix, which demonstrates that B-localization is more severe than R-localization for the same $L$. We conclude that the differences in data assimilation algorithm (LETKF vs. EnSRF) are smaller than differences in localization technique when identifying the optimal localization distance $L$.

Both types of localization introduce imbalance; as the solution reverts toward the background at long distances from observations, the damping of the height and wind increments results in a smaller wind increment but a larger height gradient, which does not satisfy the geostrophic relationship. Localization can also introduce excess divergence to an analysis (Kepert, 2009). The localization parameter $L$ should be tuned depending on the particular scale and application of data assimilation, as well as the size of the ensemble. Tuning inflation values for each localization parameter $L$ may result in improved performance. Future studies should consider balance in the context of the adaptive localization methods, as these techniques do not necessarily require a specification of $L$.

*Acknowledgments.*

We are grateful to the UMD Weather Chaos Group, Jeff Anderson, Craig Bishop, Dale Barker, two anonymous reviewers, and attendees of the WWRP/THORPEX Workshop on 4D-VAR and Ensemble Kalman Filter Inter-Comparisons in Buenos Aires (October 2008) for their helpful comments and

critiques of this project. This work was supported by NASA grants NNX07AM97G, NNX08AD40G, DOE grant DEFG0207ER64437, ONR grant N000140910418, and NOAA grant NA09OAR4310178.

# Appendix C: Mathematical Analysis of B- and R-Localizations

The relative strength of B-localization and R-localization techniques are verified mathematically using a simple example with model variable $x_1$ and $x_2$ at grid points 1 and 2, respectively. Consider a single observation of $x_1$, with $\boldsymbol{H}$=[1,0]. Using (B.2), the Kalman gain matrix (without localization) can be specified as follows:

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} B_{11} \\ B_{12} \end{bmatrix} \left( B_{11} + R_1 \right)^{-1} \qquad \text{(C.1)}$$

where $B_{ij}$ is the background covariance between $x_i$ and $x_j$, and $R_1$ is the observation covariance.

Consider the application of the B-localization function $f_{Bloc}$ (B.4) to (C.1). Using $f_{Bloc}$ ($d_{ii}$) = 1 where $d_{ij}$ is the distance between grid points $i$ and $j$, $K_1$ remains the same but $K_2$ becomes:

$$K_2 = f_{Bloc}(d_{12})B_{12}\,(f_{Bloc}(d_{11})B_{11} + R_1)^{-1} = f_{Bloc}(d_{12})B_{12}\,(B_{11} + R_1)^{-1} \qquad \text{(C.2)}$$

Note that since we are assimilating a single observation located on a grid point, (C.2) is identical for both B-localization and the $\mathbf{BH}^{\mathrm{T}}$ localization described at the end of Section B.2. Now we apply the R-localization function $f_{Rloc}$ (B.5). Again, $K_1$ remains the same as in (C.1). Using the fact that $f_{\mathrm{Bloc}} = f_{\mathrm{Rloc}}^{-1}$, $K_2$ becomes:

$$K_2 = B_{12}\,(B_{11} + f_{Rloc}(\mathrm{d}_{12})R_1)^{-1} = f_{Bloc}(\mathrm{d}_{12})B_{12}\,(f_{Bloc}(\mathrm{d}_{12})B_{11} + R_1)^{-1} \qquad \text{(C.3)}$$

Comparing (C.2) and (C.3), the R-localization (C.3) has an extra localization term in the denominator. The localization function $f_{\mathrm{Bloc}}$ ranges from 1 to 0. Therefore, the amplitude of $K_2$ (and hence the corresponding analysis increment) will be larger at grid point 2 for R-localization than for B-localization. This means that with B-localization, the analysis reverts to the background (ignores observation information) more quickly with distance compared to R-localization. In this respect, B-localization can be considered more "severe" than R-localization for the same localization distance parameter $L$; see discussion of (11) and (12) in Miyoshi and Yamane, (2007).

Now consider the same example but with two observations (one at each of the grid points) with uncorrelated errors, i.e., $\boldsymbol{H}$ is a 2-dimensional identity matrix. The Kalman gain would be written as

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \left( \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \tag{C.4}$$

where $R_1$ and $R_2$ represent the error variances of the two observations. Because the analysis process is the same for $x_1$ and $x_2$ by permuting the indices 1 and 2, we consider the impact of the localizations on $x_1$ (*i.e.*, $K_1$) only. The application of the B-localization function leads to

$$\begin{aligned} [K_1] &= \begin{bmatrix} f_{Bloc}(d_{11})B_{11} & f_{Bloc}(d_{12})B_{12} \end{bmatrix} \left( \begin{bmatrix} f_{Bloc}(d_{11})B_{11} & f_{Bloc}(d_{12})B_{12} \\ f_{Bloc}(d_{21})B_{21} & f_{Bloc}(d_{22})B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \\ &= \begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \end{bmatrix} \left( \begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \\ f_{Bloc}(d_{12})B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \end{aligned} \tag{C.5}$$

The application of the R-localization function with $f_{\mathrm{Bloc}} = f_{\mathrm{Rloc}}^{-1}$ gives

$$[K_1] = [B_{11} \quad B_{12}] \left( \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} f_{Rloc}(d_{11})R_1 & 0 \\ 0 & f_{Rloc}(d_{12})R_2 \end{bmatrix} \right)^{-1} \tag{C.6}$$

$$= [B_{11} \quad B_{12}] \left( \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \left( \begin{bmatrix} f_{Bloc}(d_{11}) & 0 \\ 0 & f_{Bloc}(d_{12}) \end{bmatrix} \right)^{-1} \right)^{-1}$$

$$= [B_{11} \quad B_{12}] \begin{bmatrix} f_{Bloc}(d_{11}) & 0 \\ 0 & f_{Bloc}(d_{12}) \end{bmatrix} \left( \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} f_{Bloc}(d_{11}) & 0 \\ 0 & f_{Bloc}(d_{12}) \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1}$$

$$= [B_{11} \quad f_{Bloc}(d_{12})B_{12}] \left( \begin{bmatrix} B_{11} & f_{Bloc}(d_{12})B_{12} \\ B_{21} & f_{Bloc}(d_{12})B_{22} \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \tag{C.7}$$

Comparing (C.5) and (C.7) in terms of the B-localization function $f_{Bloc}$, we note that the $\boldsymbol{BH^T}$ terms are identical. However, the $\boldsymbol{HBH^T}$ terms differ. Using this formulation, we arrive at an $\boldsymbol{HBH^T}$ matrix for R-localization in (C.7) that is no longer symmetric, although the original formulation of R-localization in terms of the R-localization function had symmetric covariance matrices (C.6). Consequently, it is not straightforward to compute a priori the quantitative difference in localization strength between the techniques in the case of multiple observations. With localized serial EnSRF, the resulting analysis depends upon the order in which the observations are assimilated; this is not true for the simultaneous assimilation of LETKF. For this study we focus on R-localization with the LETKF algorithm, performing EnSRF R-localization in order to confirm that differences in the results are primarily due to difference in localization technique rather than algorithm. Note that EnSRF R-localization requires a unique $\boldsymbol{R}$ (and hence a separate computation) for every gridpoint-observation pair.

# References

Allison, M., 1997: Accurate analytic representations of solar time and seasons on Mars with applications to the Pathfinder/Surveyor missions. *Geophys. Res. Lett.*, **24**, 1967-1970, doi:10.1029/97GL01950.

Allison, M., and M. McEwen, 2000: A post-Pathfinder evaluation of aerocentric solar coordinates with improved timing recipes for Mars seasonal/diurnal climate studies. *Planet. Space Sci.*, **48**, 215-235.

Amezcua, J., E. Kalnay, and P. Williams, 2010: The effects of the RAW filter on the climatology and forecast skill of the SPEEDY model. *Mon. Wea. Rev.*, **139**, 608-619.

Anderson, E. and C. Leovy, 1978: Mariner 9 television limb observations of dust and ice hazes on Mars. *J. Atmos. Sci.*, **35**, 723-734.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99-111.

Baker, V. R., R. G. Strom, V. C. Gulick, J. S. Kargel, G. Komatsu, and V. S. Kale, 1991: Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature*, **352**, 589-594.

Balme, M.R., P.L. Whelley, and R. Greeley, 2003: Mars: Dust devil track survey in Argyre Planitia and Hellas Basin. *J. Geophys. Res.*, **108**, 10.1029/2003JE002096.

Barnes, J. R., 1980: Time spectral analysis of mid-latitude disturbances in the Martian atmosphere. *J. Atmos. Sci.*, **37**, 2002-2015.

Barnes, J. R., 1984: Linear baroclinic instability in the Martian atmosphere. *J. Atmos. Sci.*, **41**, 1536-1550.

Basu, S., M.I. Richardson, and R.J. Wilson, 2004: Simulation of the Martian dust cycle with the GFDL Mars GCM. *J. Geophys. Res.*, **109**, E11906, doi:10.1029/2004JE002243.

Bishop, C. H., and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus*, **61A**, 84-96.

Bishop, C. H., and D. Hodyss, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus*, **61A**, 97-111.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical Covariance Localization for Satellite Radiances in Ensemble Kalman Filters. *Mon. Wea. Rev.*, **138**, 282-290.

Cantor, B., M. Malin, and K. S. Edgett, 2002: Multiyear Mars Orbiter Camera (MOC) observations of repeated Martian weather phenomena during the northern summer season. *J. Geophys. Res.*, **107**(E3), 5014.

Cantor, B. A., P. B. James, M. Caplinger, and M. J. Wolff, 2001: Martian dust storms: 1999 Mars Orbiter Camera observations. *J. Geophys. Res.*, **106**(E10), 23653-23687.

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. Gerald Brown, and S. Silverman, 1992: Thermal Emission Spectrometer Experiment: Mars Observer Mission. *J. Geophys. Res.*, **97**, 7719-7734.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the Effects of Data Selection with the DAO Physical-Space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913-2926.

Conrath, B. J., 1975: Thermal structure of the Martian atmosphere during the dissipation of the dust storm of 1971. *Icarus*, **24**, 36-46.

Conrath, B. J., J. C. Pearl, M. D. Smith, W. C. Maguire, P. R. Christensen, S. Dason, and M. S. Kaelberer, 2000: Mars Global Surveyor Thermal Emission Spectrometer (TES) observations: Atmospheric temperatures during aerobraking and science phasing. *J. Geophysical Res.*, **105**, 9509-9519.

Daley, R., 1991: Atmospheric Data Analysis. Cambridge University Press, Cambridge. 457 pp.

Danforth, C. M., E. Kalnay, and T. Miyoshi, 2007: Estimating and Correcting Global Weather Model Error. *Mon. Wea. Rev.*, **135**, 281-299.

Ellehoj, M. D., H. P. Gunnlaugsson, P. A. Taylor, H. Kahanpää, K. M. Bean, B. A. Cantor, B. T. Gheynani, L. Drube, D. Fisher, and A.-M. Harri, 2010: Convective vortices and dust devils at the Phoenix Mars mission landing site, *J. Geophys. Res.*, **115**, E00E16.

Eluszkiewicz, J., J.-L. Moncet, M. W. Shephard, K. Cady-Pereira, T. Connor, and G. Uymin, 2008: Atmospheric and surface retrievals in the Mars polar regions from the Thermal Emission Spectrometer measurements. *J. Geophysical. Res.*, **113**, 9 pp.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

Fisher, J., M. I. Richardson, C. E. Newman, M. A. Szwast, C. Graf, S. Basu, S. P. Ewald, A. D. Toigo, and R. J. Wilson, 2003: A survey of Martian dust devil activity using Mars Global Surveyor Mars Orbiter Camera images. *J. Geophys. Res.*, **110**, E03004, doi:10.1029/2003JE002165.

Forget, F., F. Hourdin, R. Fournier, C. Hourdin, O. Talagrand, M. Collins, S.R. Lewis, P.L. Read, and J.-P. Huot, 1999: Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *J. Geophys. Res.*, **104**, 24155-24175.

Formisano, V., S. Atreya, T. Encrenaz, N. Ignatiev, and M. Giuranna, 2004: Detection of methane in the atmosphere of Mars. *Science*, **306**, 5702, 1758-1761.

Garner, S. T., 2005: A topographic drag closure built on an analytical base flux. *J. Atmos. Sci.*, **62**, 2302-2315.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. Holben, O. Dubovik, and S.-J. Lin, 2001: Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geophys. Res.*, **106**, 20255-20273.

Greeley, R., P. L. Whelley, R. E. Arvidson, N. A. Cabrol, D. J. Foley, B. J. Franklin, P. G. Geissler, M. P. Golombek, R. O. Kuzmin, G. A. Landis, M. T. Lemmon,

L. D. V. Neakrase, S. W. Squyres, and S. D. Thompson, 2006: Active dust devils in Gusev crater, Mars: Observations from the Mars Exploration Rover Spirit, *J. Geophys. Res.*, **111**, E12S09.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Mon. Wea. Rev.*, **139**, 511-522. Doi: 10.1175/2010MWR3328.1.

Haberle, R.M., J.B. Pollack, J.R. Barnes, R.W. Zurek, C.B. Leovy, J.R. Murphy, J. Schaeffer, and H. Lee, 1993: Mars atmospheric dynamics as simulated by the NASA/Ames general circulation model I. The zonal mean circulation. *J. Geophys. Res.*, **98**, 3093-3124.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **129**, 2776-2790.

Heavens, N.G., M.I. Richardson, and A.D. Toigo, 2008: Two aerodynamic roughness maps derived from Mars Orbiter Laser Altimeter (MOLA) data and their effects on boundary layer properties in a Mars general circulation model. *J. Geophys. Res.*, **113**, E02014, doi:10.1029/2007JE002991.

Hess, S. L., R. M. Henry, C. B. Leovy, J. A. Ryan, and J. E. Tillman, 1977: Meteorological Results From the Surface of Mars: Viking 1 and 2. *J. Geophys. Res.*, **82**(28), 4559–4574.

Hinson, D. P., R. A. Simpson, J. D. Twicken, G.L. Tyler, and F. M. Flasar, 1999: Initial results from radio occultation measurements with Mars Global Surveyor. *J. Geophys. Res.*, **104**, 26997-27012.

Hinson, D. P., and R. J. Wilson, 2002: Transient waves in the southern hemisphere of Mars. *Geophys. Res. Lett.*, **29**, 10.1029/2001GL014103.

Hinson, D. P., R. J. Wilson, M.D. Smith, and B.J. Conrath, 2003: Stationary planetary waves in the atmosphere of Mars during southern winter. *J. Geophys. Res.*, **108**(E1) 5004, doi:10.1029/2002JE001949.

Hinson, D. P., and R. J. Wilson, 2004: Temperature inversions, thermal tides, and water ice clouds in the Martian tropics. *J. Geophys. Res.*, **109**, E01002, doi:10.1029/JE002129.

Hinson, D. P., M. D. Smith, and B. J. Conrath, 2004: Comparison of atmospheric temperatures obtained through infrared sounding and radio occultation by Mars Global Surveyor. *J. Geophys. Res.*, **109**, E12002.

Hinson D. P., H. Wang and M. Smith, 2011: Baroclinic wave Transitions, Autumn Weather Patterns, and Dust Storms in the Northern Hemisphere of Mars. The Fourth International Workshop on the Mars Atmosphere, Paris, France.

Hoffman, M. J., E. Kalnay, J.A. Carton, and S.-C. Yang, 2009: Use of Breeding to Detect and Explain Instabilities in the Global Ocean. *Geophys. Res. Lett.*, **36**, L12608, 5 pp.

Hoffman, M. J., S. J. Greybush, R. J. Wilson, G. Gyarmati, R. N. Hoffman, K. Ide, E. Kostelich, T. Miyoshi, I. Szunyogh, and E. Kalnay, 2010: An ensemble Kalman filter data assimilation system for the Martian atmosphere: Implementation and simulation experiments. *Icarus*, **209**(2), 470–481.

Hoffman M. J., J. Eluszkiewicz, R. Hoffman, S. J. Greybush, E. Kalnay and R. J. Wilson, 2011: Evaluation of an Optimal Spectral Sampling Retrieval

Algorithm For Thermal Emission Spectrometer Radiances. The Fourth International Workshop on the Mars Atmosphere, Paris, France.

Hoffman, R. N., 2010: A retrieval strategy for interactive ensemble data assimilation. *arXiv*, (1009.1561v1), 1–13, available at http://arxiv.org/abs/1009.1561.

Holland, B. W. and X. Wang, 2010: A comparison of the local ensemble transform Kalman filter and ensemble square root filter data assimilation schemes. *14th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Atlanta, GA, Amer. Meteor. Soc.

Holstein-Rathlou, C., H. P. Gunnlaugsson, J. P. Merrison, K. M. Bean, B. A. Cantor, J. A. Davis, R. Davy, N. B. Drake, M. D. Ellehoj, and W. Goetz, 2010: Winds at the Phoenix landing site. *J. Geophys. Res.*, **115**, E00E18.

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

Houtekamer, P. L. and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Wea. Rev.*, **126**, 796-811.

Houtekamer, P. L. and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Houtekamer, P. L. and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteorol. Soc.*, **131**, 3269-3289.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman Filter. *Physica D*, **230**, 112-126.

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng., Trans. ASME*, 35-45.

Kalnay, E., 2010: Ensemble Kalman Filter: Current Status and Potential, 24 pp. Data Assimilation: Making Sense of Observations, Springer, W. Lahoz, B. Khattatov, and Richard Menard, Ed.

Kalnay, E., H. Li, T. Miyoshi, S-C. Yang, and J. Ballabrera, 2007: 4-D-Var or Ensemble Kalman Filter?. *Tellus*, **59A**, 758-773. doi:10.1111/j.1600-0870.2007.00261.x

Kang, J.-S., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 2010: "Variable localization" to improve assimilation in an Ensemble Kalman Filter. J. Geophys., Res., submitted.

Kepert, J., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteorol. Soc.*, **135**, 1157-1176.

Kleinboehl, A., J. T. Schofield, D. M. Kass, W. A. Abdou, C. R. Backus, ZB. Sen, J. H. Shirley, W. G. Lawson, M. I. Richardson, F. W. Taylor, N. A. Teanby, and D. J. McCleese, 2009: Mars Climate Sounder limb profile retrieval of atmospheric temperature, pressure, and dust and water ice opacity. *J. Geophys. Res.*, **114**, 30 pp.

Kreslavsky, M. A., and J. W. Head III, 2000: Kilometer-scale roughness of Mars: Results from MOLA data analysis. *J. Geophys. Res.*, **105**(E11), 26,695–26,711.

Langevin, Y., F. Poulet, J.-P. Bibring, B. Schmitt, S. Douté and B. Gondet, 2005: Summer Evolution of the North Polar Cap of Mars as Observed by OMEGA/Mars Express. *Science*, **307**, 5715, 1581-1584.

Lemmon, M.T., M. J. Wolff, M. D. Smith, R. T. Clancy, D. Banfield, G. A. Landis, A. Ghosh, P. H. Smith, N. Spanovich, B. Whitney, P. Whelley, R. Greeley, S. Thompson, J. F. Bell III, and S. W. Squyres, 2004: Atmospheric Imaging Results from the Mars Exploration Rovers: Spirit and Opportunity. *Science*, **306**, 5702, 1753-1756.

Lewis, S. R., M. Collins, P. L. Read, F. Forget, F. Hourdin, R. Fournier, et al., 1999: A climate database for Mars. *J. Geophys. Res.*, **104**(E10), 24177-24194.

Lewis, S. R. and P. R. Barker, 2005: Atmospheric tides in a Mars general circulation model with data assimilation. *Advances in Space Res.*, **36**, 2162-2168.

Lewis, S. R., P. L. Read, B. J. Conrath, J. C. Pearl, and M. D. Smith, 2007: Assimilation of thermal emission spectrometer atmospheric data during the Mars Global Surveyor aerobraking period. *Icarus*, **192**, 327-347.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.*, **135**, 523-533.

Lin., S-J., 2004: A "vertically Lagrangian" finite volume dynamical core for global models. *Mon. Wea. Rev.*, **132**(10), 2293-2307.

Lorenc, A. C., R. S. Bell, and B. Macpherson, 1991: The Meteorological Office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **117**, 59-89.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quart. J. Roy. Meteorol. Soc.*, **129**, 3183 - 3203.

Lorenz, E. N., 1965: On the possible reasons for long-period fluctuations of the general circulation. WMO-IUGG Symp. on Research and Development Aspects of Long-range Forecasting, Technical Note No. 66, WMO-No. 162.TP.79, 203-211.

Lorenz, E. N., 1967: "The Nature and Theory of the General Circulation of the Atmosphere", World Meteorological Organization monograph.

Lynch, P. and X.-Y. Huang, 1992: Initialization of the HIRLAM Model Using a Digital Filter. *Mon. Wea. Rev.*, **120**, 1019-1034.

Malin, M. C., B. A. Cantor, T.N. Harrison, D.E. Shean and M.R. Kennedy (2011), MRO MARCI Weather Report for the week of 7 March 2011 – 13 March 2011, Malin Space Science Systems Captioned Image Release, MSSS-162, http://www.msss.com/msss_images/2011/03/16/.

McCleese, D. J., J. T. Schofield, F.W. Taylor, S.B. Calcutt, M.C. Foote, D. M. Kass, C. B. Leovy, D. A. Paige, P.L. Read, and R. W. Zurek, 2007: Mars Climate Sounder: An investigation of thermal and water vapor structure, dust and condensate distributions in the atmosphere, and energy balance of the polar regions. *J. Geophys. Res.*, **112**, 16 pp.

McCleese D. J. et al., 2010: Structure and dynamics of the Martian lower and middle atmosphere as observed by the Mars Climate Sounder: Seasonal variations in zonal mean temperature, dust and water ice aerosols. *J. Geophys. Res.*, **115**, E12016, doi:101029/2010JE003677.

McDunn, T., S. Bougher, M. Mischna, J. Murphy, and S. Nelli, 2011: Modeling Polar Warming at Mars: Preliminary Results of the Newly Vertically Extended Mars-WRF GCM and Comparisons with Constraints from Data. The Fourth International Workshop on the mars Atmosphere, Paris, France.

Mellor, G. L., and T. Yamada, 1982: Development of a Turbulence Closure Model for Geophysical Fluid Problems. *Reviews of Geophysics and Space Physics*, **20**, 851-875.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **130**, 2791-2808.

Miyoshi, T., 2005: Ensemble Kalman Filter Experiments with a Primitive-Equation Global Model. Ph.D. dissertation, University of Maryland, 226 pp.

Miyoshi, T. and S. Yamane, 2007: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. *Mon. Wea. Rev.*, **135**, 3841-3861.

Miyoshi, T., and Y. Sato, 2007: Assimilating Satellite Radiances with a Local Ensemble Transform Kalman Filter (LETKF) Applied to the JMA Global Model (GSM). *SOLA*, **3**, 37-40.

Miyoshi, T., 2011: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. *Mon. Wea. Rev.*, in press.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: model climatology and variability in multi-decadal experiments. *Climate Dynamics*, **20**, 175-191.

Moncet, J.-L., G. Uymin, A. E. Lipton, and H. E. Snell, 2008: Radiance modeling at high spectral resolution by optimal spectral sampling. *J. Atmos. Sci.*, **65**, 3917-3934.

Montabone, L., S. R. Lewis, and P. L. Read, 2005: Interannual variability of Martian dust storms in assimilation of several years of Mars global surveyor observations. *Advances in Space Research*, **36**, 2146-2155.

Montabone, L., S. R. Lewis, P.L. Read, and D. P. Hinson, 2006: Validation of martian meteorological data assimilation for MGS/TES using radio occultation measurements. *Icarus*, **185**, 113-132.

Montmessin, F., F. Forget, P. Rannou, M. Cabane, and R.M. Haberle, 2004: Origin and role of water ice clouds in the martian water cycle as inferred from a general circulation model. *J. Geophys. Res.*, **109**, E10004, doi:10.1029/2004JE002284.

Moores, John E., Léonce Komguem, James A. Whiteway, Mark T. Lemmon, Cameron Dickinson, and Frank Daerden, 2011: Observations of near-surface fog at the Phoenix Mars landing site. *Geophys. Res Lett.*, **38**, L04203.

Newman, C. E., P. L. Read, and S. R. Lewis, 2002: Investigating atmospheric predictability on Mars using breeding vectors in a general circulation model. *Q. J. R. Meteorol. Soc.*, **128**, 21 pp.

Ockert-Bell, M.E., J.F. Bell III, J.B. Pollack, C.P. McKay, and F. Forget, 1997: Absorption and scattering properties of the Martian dust in the solar wavelengths. *J. Geophys. Res.*, **102**, 9039-9050.

Orlanski, I. and J. Katzfey, 1991: The Life Cycle of a Cyclone Wave in the Southern Hemisphere. Part I: Eddy Energy Budget. *J. Atmos. Sci.*, **48**, 1972-1998.

Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878-5881.

Rafkin, S. C. R., 2011: Non-Local, Deep Transport in the atmosphere of Mars. The Fourth International Workshop on the mars Atmosphere, Paris, France.

Raymond, D. J., 1992: Nonlinear balance and potential-vorticity thinking at large Rossby number. *Quart. J. Roy. Meteor. Soc.*, **118**, 987–1015.

Read, P. L., and S. R. Lewis, 2004: The Martian Climate Revisited: Atmosphere and Environment of a Desert Planet. Springer / Praxis Publishing, Chichester, UK.

Read P. L., L. Montabone, D. P. Mulholland, S. R. Lewis, B. Cantor and R. J. Wilson, 2011: Midwinter Suppression of Baroclinic Storm Activity on Mars: Observations and Models. The Fourth International Workshop on the mars Atmosphere, Paris, France.

Richardson, M.I., and R.J. Wilson, 2002: Investigation of the nature and stability of the martian seasonal water cycle with a general circulation model. *J. Geophys. Res.*, **107**(E5), 10.1029/2001JE001536.

Richardson, M.I., R.J. Wilson, and A.V. Rodin, 2002: Water ice clouds in the martian atmosphere: General circulation model experiments with a simple cloud scheme. *J. Geophys. Res.*, **107**(E9), 10.1029/2001JE001804.

Richardson, M. I., A. D. Toigo, and C. E. Newman, 2007: PlanetWRF: A general purpose, local to global numerical model for planetary atmospheric and climate dynamics. *J. Geophys. Res.*, **112**, E09001.

Rodgers, C.D., 2000: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing Co. Ltd.

Rogberg, P., P. L. Read, S. R. Lewis, and L. Montabone, 2010: Assessing atmospheric predictability on Mars using numerical weather prediction and data assimilation. *Q. J. R. Meteorol. Soc.*, **136**, 1614-1635.

Schofield, J. T., J. R. Barnes, D. Crisp, R. M. Haberle, S. Larsen, J. A. Magalhães, J. R. Murphy, A. Seiff and G. Wilson, 1997: The Mars Pathfinder Atmospheric Structure Investigation/Meteorology (ASI/MET) Experiment. *Science*, **278**, 5344, 1752-1758.

Smith, M.D., J. C. Pearl, B. J. Conrath, and P. R. Christensen, 2001: Thermal Emission Spectrometer results: Mars atmospheric thermal structure and aerosol distribution. *J. Geophysical Res.*, **106**, 23929-23945.

Smith, M. D., J. L. Bandfield, P. R. Christensen, and M. I. Richardson, 2003: Thermal Emission Imaging System (THEMIS) infrared observations of atmospheric dust and water ice cloud optical depth. *J. Geophys. Res.*, **108**, E11, 5115.

Smith, P. H., L. K. Tamppari, R. E. Arvidson, and coauthors, 2009: $H_2O$ at the Phoenix Landing Site. *Science*, **325** (5936), p. 58-61.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus A*, **60**, 113-130.

Tanaka, H. L. and M. Arai, 1999: Linear baroclinic instability in the Martian atmosphere: Primitive equation calculations. *Earth Planets & Space*, **51**, 225-232.

Taylor, P. A., D. C. Catling, M. Daly, C. S. Dickinson, H. P. Gunnlaugsson, A.-M. Harri, and C. F. Lange, 2008: Temperature, pressure, and wind instrumentation in the Phoenix meteorological package. *J. Geophys. Res.*, **113**, E00A10.

Toth, Z. and E. Kalnay, 1993: Ensemble Forecasting at NMC: The Generation of Perturbations. *Bull. Amer. Met. Soc.*, **74**, 2317-2330.

Vaughan, A. F., J. R. Johnson, K. E. Herkenhoff, R. Sullivan, G. A. Landis, W. Goetz, and M. B. Madsen, 2010: Pancam and Microscopic Imager observations of dust on the Spirit Rover: Cleaning events, spectral properties, and aggregates. *Mars*, **5**, 129-145.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913-1924.

Whiteway, J. A., L. Komguem, C. Dickinson, C. Cook, M. Illnicki, J. Seabrook, V. Popovici, T. J. Duck, R. Davy, P. A. Taylor, J. Pathak, D. Fisher, A. I. Carswell, M. Daly, V. Hipkin, A. P. Zent, M. H. Hecht, S. E. Wood, L. K. Tamppari, N. Renno, J. E. Moores, M. T. Lemmon, F. Daerden, and P. H. Smith, 2009: Mars Water-Ice Clouds and Precipitation. *Science*, **325** (5936), 68-70.

Williams, P. D., 2009: A Proposed Modification to the Robert-Asselin Time Filter. *Mon. Wea. Rev.*, **137**, 2538-2546.

Wilson, R. J. and K. Hamilton, 1996: Comprehensive Model Simulation of Thermal Tides in the Martian Atmosphere. *J. Atmospheric Sci.*, **53**, 1290-1326.

Wilson, R.J., 1997: A general circulation model simulation of the martian polar warming. *Geophys. Res. Lett.*, **24**, 123-127.

Wilson, R.J., D. Banfield, B.J. Conrath, and M.D. Smith, 2002: Traveling waves in the northern hemisphere of Mars. *Geophys. Res. Lett.*, **29**(14), 10.1029/2002GL014866.

Wilson, R.J., G. Neumann, and M.D. Smith, 2007: The diurnal variation and radiative influence of martian water ice clouds. *Geophys. Res. Lett.*, **34**, L02710, doi:10.1029/2006GL027976.

Wilson, R.J., S.R. Lewis, and L. Montabone, 2008: Influence of water ice clouds on Martian tropical atmospheric temperatures. *Geophys. Res. Lett.*, **35**, L07202, doi:10.1029/2007GL032405.

Wilson, R. J. and M. A. Kahre, 2009: The role of spatially variable surface dust in GCM simulations of the martian dust cycle. Mars Dust Cycle workshop, NASA/Ames Research Center, CA.

Wilson, R. J., J. Noble, and S. J. Greybush, 2011: The Derivation of Atmospheric Opacity from Surface Temperature Observations. The Fourth International Workshop on the mars Atmosphere, Paris, France.

Wolff, M. J., R. T. Clancy, B. Cantor, and J.-B. Madeleine, 2011: Mapping Water Ice Clouds (and Ozone) with MRO/MARCI. The Fourth International Workshop on the mars Atmosphere, Paris, France.

Wolff, M.J., and R.T. Clancy, 2003: Constraints on the size of Martian aerosols from Thermal Emission Spectrometer observations. *J. Geophys. Res.*, **108**(E9), 5098, doi:10.1029/2003JE002058.