Assimilation of temperatures and column dust opacities measured by ExoMars TGO-ACS-TIRVIM during the MY34 Global Dust Storm

Roland M. B. Young, Ehouarn Millour, Sandrine Guerlet, François Forget, Nikolay Ignatiev, Alexey Grigoriev, Alexey Shakun, Alexander Trokhimovskiy, Franck Montmessin, Oleg Korablev

Assimilation of temperatures and column dust opacities measured by ExoMars TGO-ACS-TIRVIM during the MY34 Global Dust Storm

Roland M. B. Young[1,2], Ehouarn Millour[2], Sandrine Guerlet[2,3], François Forget[2], Nikolay Ignatiev[4], Alexey V. Grigoriev[4], Alexey V. Shakun[4], Alexander Trokhimovskiy[4], Franck Montmessin[5], Oleg Korablev[4]

[1]Department of Physics & National Space Science and Technology Center, UAE University, Al Ain, United Arab Emirates

[2]Laboratoire de Météorologie Dynamique (LMD/IPSL), Sorbonne Université, ENS, PSL Research University, École Polytechnique, Institut Polytechnique de Paris, CNRS, Paris, France

[3]LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, 92195 Meudon, France

[4]Space Research Institute (IKI), 84/32 Profsoyuznaya, 117997 Moscow, Russia

[5]LATMOS/IPSL, UVSQ Université Paris-Saclay, UPMC Univ. Paris 06, CNRS, Guyancourt, France

Key Points:

- We assimilate temperature and dust observations from ExoMars TGO-ACS-TIRVIM during the MY34 global dust storm.
- The analysis verifies well against independent Mars Climate Sounder temperature and dust profiles, and Curiosity pressure measurements.
- At the peak of the global dust storm the winds strengthen, the diurnal tide migrates poleward, and the semi-diurnal tide strengthens.

Corresponding author: Roland M. B. Young, roland.young@uaeu.ac.ae

Abstract

We assimilate atmospheric temperature profiles and column dust optical depth observations from the ExoMars Trace Gas Orbiter Atmospheric Chemistry Suite thermal infrared channel (TIRVIM) into the LMD Mars Global Climate Model. The assimilation period is Mars Year 34 $L_s = 182.3 - 211.4°$, covering the onset and peak of the 2018 global dust storm. We assimilated observations using the Local Ensemble Transform Kalman Filter with 36 ensemble members and adaptive inflation; our nominal configuration assimilated TIRVIM temperature profiles to update temperature and dust profiles, followed by dust column optical depths to update the total column dust abundance. The observation operator for temperature used the averaging kernels and prior profile from the TIRVIM retrievals.

We verified our analyses against in-sample TIRVIM observations and independent Mars Climate Sounder (MCS) temperature and dust density-scaled opacity profiles. When dust observations were assimilated, the root-mean-square temperature error verified against MCS fell by 50% during the onset period of the storm, compared with assimilating temperature alone. At the peak of the storm the analysis reproduced the location and magnitude of the peak in the nighttime MCS dust distribution, along with the surface pressure diurnal cycle measured by Curiosity with a bias of less than 10 Pa. The analysis winds showed that, at the peak of the storm, the meridional circulation strengthened, a 125 m s^{-1} asymmetry developed in the midlatitude zonal jets, the diurnal tide weakened near the equator and strengthened to 10–15 K at midlatitudes, and the semi-diurnal tide strengthened almost everywhere, particularly in the equatorial lower atmosphere.

Plain Language Summary

The ExoMars Trace Gas Orbiter (TGO) has been in a low orbit around Mars since early 2018. Halfway through 2018 (Mars Year 34) a large amount of dust was lifted from Mars' surface into the lower atmosphere. This dust eventually obscured most of the planet, called a global dust storm. These storms have a large effect on the temperature and weather in Mars' atmosphere.

In this study we have used observations from TGO's thermal infrared instrument TIRVIM. This measures how much light comes from the Martian atmosphere and surface at wavelengths where heat is emitted; these measurements can be used to work out

the atmospheric temperature and how much dust it contains. We combined these observations with a numerical model of Mars' climate in a way that takes into account how uncertain we are about the atmospheric properties measured by TIRVIM and the climate predicted by the model.

We compared our results with separate temperature and dust observations made by the Mars Climate Sounder instrument on board another Mars orbiting satellite, Mars Reconnaissance Orbiter. We also studied how the atmospheric wind, the day-night cycle, and the surface pressure all changed while the dust storm was raging.

1 Introduction

The ExoMars Trace Gas Orbiter (TGO) reached its final orbit around Mars in April 2018. It carries the Atmospheric Chemistry Suite (ACS), which comprises three infrared spectrometers covering different wavelength ranges (Korablev et al., 2018). The thermal infrared spectrometer TIRVIM continuously monitors the atmospheric column in nadir, while the near- and mid-infrared channels NIR and MIR (and occasionally TIRVIM) perform solar occultation measurements when the spacecraft-Mars-Sun geometry permits. TGO's orbit made ACS-TIRVIM's (TIRVIM hereafter) observations of Mars unique at the time among existing Mars spacecraft, because it systematically measures the atmospheric structure over all local times of day. Nearly all spacecraft that have orbited Mars since Mars Global Surveyor at the start of the modern era have taken observations at two fixed local times of day separated by 12 hours (Mars Express has an orbit that means it does take observations at varying local times of day (Giuranna et al., 2021), but not systematically, and since May 2021 the Emirates Mars Mission has observed Mars from high orbit with full local time coverage (Amiri et al., 2022)). TGO takes about 55 sols to cycle through the full 24 hours of sub-spacecraft local times, which is close to the optimal repeat period to sample the diurnal cycle (Capderou & Forget, 2004).

TIRVIM took observations at Mars from March 2018, shortly before TGO reached its final orbit, until the end of 2019, when its cryocooler failed, rendering it impossible to cool the instrument to the required temperature to make thermal infrared measurements. At the start of this ∼20-month period, about halfway through Mars Year (MY) 34, a Global Dust Storm (GDS) took place. This started around $L_s = 186°$ in Acidalia, reached a peak in visible dust optical depth between $L_s = 200{-}205°$, and faded away

by about $L_s = 300°$ (Kass et al., 2019; Montabone et al., 2020; Kleinböhl et al., 2020). The storm was severe enough to end the solar-powered Opportunity rover's 15-year mission. The start of the TIRVIM dataset covers the onset and peak of this GDS; fortuitously, as these events happen in only about one Mars year out of three (Kahre et al., 2017).

Over the last 20–25 years, scientists have studied Mars using data assimilation, which combines observations with our best scientific understanding encoded within numerical simulations (Kalnay, 2003). A variety of methods have been used, with most work using Analysis Correction (AC) (Lorenc et al., 1991; Lewis & Read, 1995; Lewis et al., 2007; Steele, Lewis, & Patel, 2014) or the Ensemble Kalman Filter (Evensen, 2003; M. J. Hoffman et al., 2010; Greybush et al., 2012; Navarro et al., 2014). These have assimilated various observations, primarily atmospheric temperature profiles, but also dust column optical depths (Lewis & Barker, 2005; Montabone et al., 2014; Ruan et al., 2021), dust profiles (Navarro et al., 2017; Ruan et al., 2021), and column abundances of water ice (Steele, Lewis, & Patel, 2014), water vapour (Steele, Lewis, Patel, Montmessin, et al., 2014), ozone (Holmes et al., 2018), and carbon monoxide (Holmes et al., 2019). The primary assimilated quantity is temperature, because it is readily available from thermal infrared measurements from multiple spacecraft, changes in other atmospheric quantities such as pressure and winds are related to changes in temperature via known physical laws, and while wind observations are a vital part of Earth data assimilation (Hersbach et al., 2019), they are not available for Mars.

Most relevant to the current work, Navarro et al. (2014, 2017) built a data assimilation scheme for Mars' atmosphere using the Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al., 2007) and the Laboratoire de Météorologie Dynamique Mars Global Climate Model (LMD Mars GCM) (Forget et al., 1999; Pottier et al., 2017), assimilating observations from the Mars Climate Sounder (MCS) instrument on board NASA's Mars Reconnaissance Orbiter (MRO). MCS retrieval products include temperature, dust, and water ice profiles taken from a Sun-synchronous polar orbit at ∼3 AM / 3 PM local mean solar time (Kleinböhl et al., 2009).

In this paper we report on assimilation of observations from the early part of the ExoMars TGO ACS-TIRVIM observation period, covering the onset and peak of the MY34 GDS. We make two methodological improvements to the existing scheme documented by Navarro et al. (2017). First, we improve the way temperature forecasts are compared

with observations by using the averaging kernel matrix and prior from the TIRVIM temperature retrievals. This is the correct way to compare forecast with observations when assimilating retrieval products (R. N. Hoffman, 2011). Second, the scheme can now assimilate column dust optical depths (CDOD); previous work with this particular scheme did not assimilate CDOD because such data are not available from MCS (while dust profiles are available, they do not reach the ground, and published MCS CDODs interpolate the dust profiles to the ground in the lowest part of the atmosphere, typically the lowest 10 km). We describe both of these improvements in detail in Sect. 3.

The aims of this paper are as follows. First, to demonstrate assimilation of observations from TIRVIM on board ExoMars TGO. Second, to validate the assimilation against independent datasets such as MCS temperatures and dust opacities, and surface pressure measurements from the Curiosity rover. Finally, to investigate the dynamics of Mars' atmosphere during the MY34 GDS, particularly those properties that cannot be measured directly but which can be retrieved by the assimilation process, such as wind. Section 2 describes the TIRVIM observations, and in Section 3 we describe the model, assimilation scheme, and observation operator. Section 4 describes sensitivity tests done to optimise the assimilation parameters. Section 5 describes the main results of the paper, and in Section 6 we conclude.

2 Observations

TIRVIM is a thermal infrared spectrometer that observes Mars at wavelengths between 1.7–17 μm (590–5900 cm^{-1}) (Korablev et al., 2018). It was designed to be continuously operating during TGO's mission (although in practice this was not possible due to the limited lifetime of the instrument's cryocooler). TIRVIM measures radiance spectra in nadir, from which have been retrieved atmospheric temperature profiles, surface temperatures, dust column optical depths (at 1090 cm^{-1}) and water ice column optical depths (at 820 cm^{-1}). These retrievals are fully described by Guerlet et al. (2022). In this paper we focus on atmospheric temperature and column dust optical depth observations only. TIRVIM operated from MY34 $L_s = 142.79°$ (13 March 2018) until MY35 $L_s = 115.16°$ (2 December 2019), when its cryocooler failed.

TIRVIM takes observations at all local times over a 55-sol period. TGO's inclination is 74°, so there are no nadir observations poleward of this latitude in either hemi-

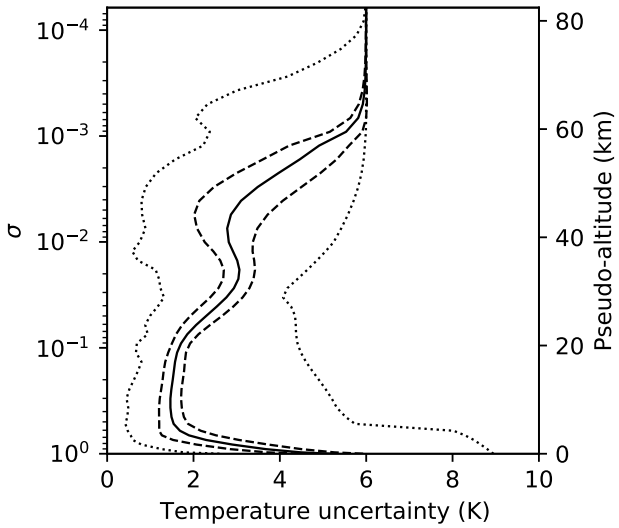


Figure 1. Temperature uncertainty in the TIRVIM dataset between MY34 $L_s = 182.288 - 211.388°$ as a function of $\sigma = p/p_{\mathrm{surf}}$. The right y-axis is pseudo-altitude $z = -H \ln \sigma$, assuming a scale height of $H = 8.5$ km, as per the Mars standard atmosphere from the surface to 85 km altitude in Table 4.7 of M. D. Smith et al. (2017) (an underestimate in the lower atmosphere and an overestimate in the middle atmosphere, but it is sufficiently accurate for plotting). We use $H = 8.5$ km for all pseudo-altitudes in this paper. The solid line is the mean, dashed lines are mean +/- 1 standard deviation, and the dotted lines show the minimum and maximum.

sphere. The temperature profiles have a maximum sensitivity between 5 and 50 km above the ground, and a vertical resolution of about 10 km (Guerlet et al., 2022). The instrument is only sensitive to the true atmospheric temperatures at altitudes of 3–55 km, based on retrieval averaging kernel statistics (see Section 3.3). From the trace of the averaging kernel matrix, the number of degrees of freedom in each temperature retrieval is 3.3 ± 0.2. The measurement uncertainty, shown in Fig. 1, is a function of altitude. It is typically 2–4 K over most of the altitude range, increasing at the top and the bottom of the profile. Column dust optical depths typically have an uncertainty of 0.06–0.20 (the 25–75% range over the assimilated observations, with mean 0.16).

In this paper we assimilate observations from just before the onset of the MY34 global dust storm to just after its peak, from $L_s = 182.288 - 211.388°$. Figure 2 shows the data availability and distribution in latitude and local time for this period. These specific start and end points were chosen because there are long gaps in the data imme-

diately before and afterwards. There are 312,741 temperature profiles and 214,806 column dust retrievals during this period. There is a gap in the data between $L_s = 191.567-194.420°$), with significantly higher observation density afterwards. The local time of day for the majority of observations moves backwards over time; almost all 24 hours of local time are sampled during this period. Exceptions are the northern hemisphere around 4 AM, and the southern hemisphere around 4 PM. A significant fraction of TIRVIM observations have missing values for CDOD, as it cannot be retrieved in conditions when the temperature contrast between surface and lower atmosphere is low, which typically occurs near dawn and dusk in non-stormy conditions (see Guerlet et al. (2022)). Most of the missing dust retrievals are poleward of 60°, where almost all dust retrievals are removed.

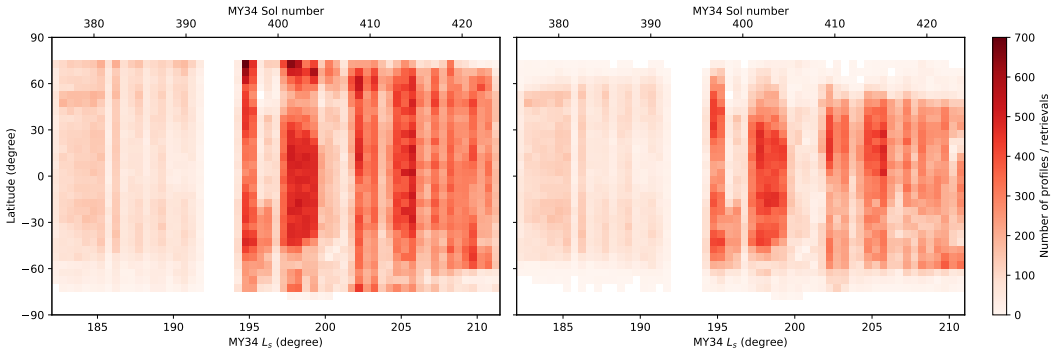
3 Methods

3.1 Model

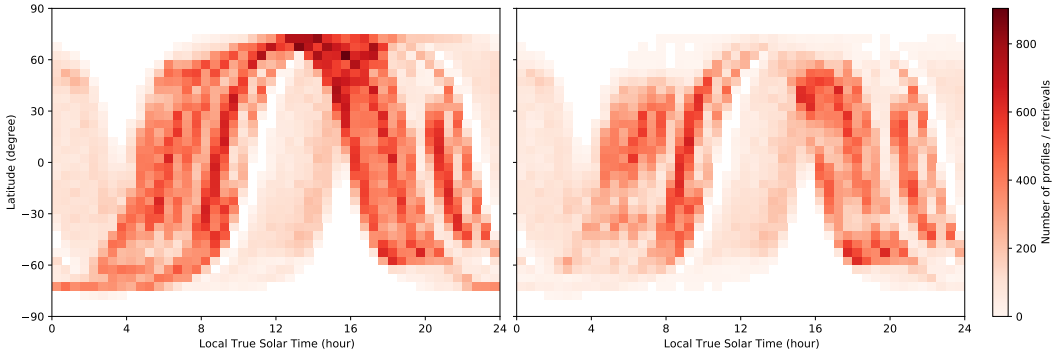
The LMD Mars GCM solves the hydrostatic primitive equations on a rotating sphere alongside parametrizations of many physical processes relevant to Mars (Forget et al., 1999). Most pertinent to the current work is the treatment of dust, which is transported by the model in a two-moment scheme which separately transports a mass mixing ratio and the number of dust particles within each grid cell (Madeleine et al., 2011). By transporting these two quantities separately one can deduce the dust particle size distribution by assuming a log-normal particle size distribution. The LMD Mars GCM has been validated against many observational data sets since the Viking landers in the 1970s (Forget et al., 1999).

We run the model at the standard climate model resolution of 64×48 points in longitude and latitude, corresponding to $5.625° \times 3.75°$ horizontal resolution. The vertical grid is stretched so there are more points near the ground; we use 32 hybrid (sigma-pressure) levels with the lowest level about 4 m above the ground, and the highest level at about 100 km altitude, depending on local conditions. The model's dynamical time step δt is 1/960 of a sol (∼94.5 s), and the physical parametrisations are called every 10 dynamical timesteps (i.e. every physics timestep, 1/96 of a sol, or four times per Martian hour). Supporting Text S1 describes the time axis used in the GCM and how it relates to observation times.

(a) vs MY34 L_s and latitude.



(b) vs local true solar time and latitude.



(c) vs MY34 L_s and local true solar time.

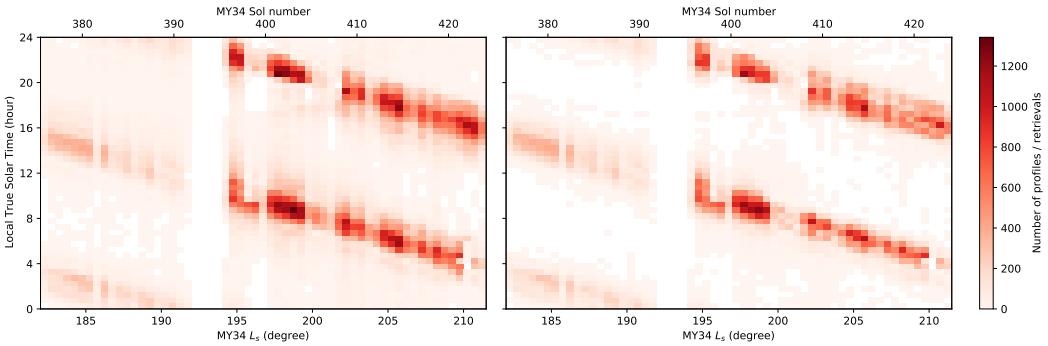


Figure 2. Number of TIRVIM temperature profiles (left) and column dust optical depth retrievals (right) as a function of MY34 L_s, latitude, and local true solar time during the period assimilated in this work. (a) vs MY34 L_s and latitude; (b) vs local true solar time and latitude; (c) vs MY34 L_s and local true solar time. Observations are counted in bins with width 5.0° in latitude, 0.5 hours in local time, and 0.5° in L_s; bins with no observations are white. The upper x-axis shows the equivalent model sol numbers; see Supporting Text S1 for details.

Dust sediments and can fall out of the bottom of the model onto the surface. When run without assimilation, dust column abundances in the GCM are forced by rescaling at each physics timestep to a global "dust scenario" map based on observed dust opacities, with gaps filled in by kriging (Montabone et al., 2015, 2020). The GCM also lifts dust from the surface at a constant rate, to preserve the dust particle size distribution by seeding the atmosphere with larger particles, which preferentially sediment out. Water is transported by the model (Pottier et al., 2017), and can condense into water ice clouds, whose particles have an effective radius that varies with local conditions, and which are treated as radiatively active. Dust acts as a source of cloud condensation nuclei, and precipitating water ice can remove dust from the atmosphere by scavenging.

3.2 Assimilation Scheme

Assimilation is based on the LETKF, which is a standard method for ensemble data assimilation (Evensen, 2003; Hunt et al., 2007). Full details of the LETKF itself as applied in this context are described in Navarro et al. (2014) and Navarro et al. (2017); here we give a summary and focus on the changes to the scheme since Navarro et al. (2017).

Before starting the assimilation we spin up an ensemble of N GCM simulations. This ensemble is used as the forecast for the first assimilation step. Its ensemble mean $\overline{\mathbf{x}}^{\mathrm{b}}$ is used as the *background* atmospheric state in the assimilation, and the purpose of the ensemble is to estimate the uncertainty in the background state. We use the spread of values to set the background error covariance matrix $\mathbf{B}$ in the assimilation. For the spin up stage in this case we ran an ensemble of GCM simulations starting at MY34 sol 318 ($L_s = 150.2°$), i.e. about 30 L_s before the start of the assimilation period. This gave the ensemble members enough time to spin up to an equilibrium state individually, and to spread out over the model state space such that the standard deviation of the ensemble at each point quantified the uncertainty in the background state. As dust is the major unknown in Mars' atmosphere, and the atmospheric state is so sensitive to it, to span the range of possibilities we ran each spin-up simulation with a different globally-fixed visible CDOD, log-uniformly distributed between 0.05 and 1.5. This spans the typical range of realistic dust column optical depths in non-GDS years (Montabone et al., 2015).

The assimilation step computes an *analysis* ensemble $\mathbf{x}^{\mathrm{a},i}$, $i = 1 \dots N$, at the analysis time t_a. Most of the results are presented using the mean of this ensemble, which we shall simply refer to as the *analysis*. The assimilation uses the forecast ensemble as the background state $\mathbf{x}^{\mathrm{b},i}$, and observations $\mathbf{y}_\mathrm{o}$ within an assimilation window stretching from $t_a - t_b$ to $t_a + t_f$, where t_b is the window length backwards in time and t_f is the window length forward in time. To ensure each observation is used in exactly one assimilation step, $t_b = t_f = \Delta t/2$, where Δt is the assimilation cycle length, the time between consecutive assimilation steps. Adaptive covariance inflation is applied to the background ensemble. The observation operator $H(\cdots)$ is applied to the background ensemble, to interpolate the forecasts in space and time to the observation points, and to ensure that the forecast and observations are subsequently compared as if TIRVIM were observing the background state; Sect. 3.3 provides full details.

To perform the assimilation, at each model grid point we identify observations within $L_\mathrm{eh} = 900$ km horizontally (along a great circle), and within $L_\mathrm{ev} = 0.2\sqrt{6}$ vertically in $\log(p)$, a process called localisation (Hamill et al., 2001), which significantly speeds up the calculation and also reduces the effects of random noise in the observations acting over large distances. The measurement uncertainty for each observation is modified according to its horizontal (d_h) and vertical distance (d_v) from the grid point, where d_v is measured in $\log p$. For observation l, this modified observational uncertainty is

$$\sigma_l^\mathrm{o} = \sigma^\mathrm{o} \exp\left(\frac{d_h^2}{2L_\mathrm{ih}^2}\right) \exp\left(\frac{d_v^2}{2L_\mathrm{iv}^2}\right) \tag{1}$$

where σ^o is the unmodified observational uncertainty, $L_\mathrm{ih} = 600$ km is the internal horizontal localisation length, and $L_\mathrm{iv} = 0.2$ is the internal vertical localisation length.

The Local Ensemble Transform Kalman Filter equation for the analysis mean $\bar{\mathbf{x}}^\mathrm{a}$ is (Hunt et al., 2007; Navarro et al., 2017)

$$\bar{\mathbf{x}}^\mathrm{a} = \bar{\mathbf{x}}^\mathrm{b} + \mathbf{X}^\mathrm{b}\tilde{\mathbf{P}}^\mathrm{a}(\mathbf{Y}^\mathrm{b})^\top \mathbf{R}^{-1}(\mathbf{y}^\mathrm{o} - H(\bar{\mathbf{x}}^\mathrm{b})) \tag{2}$$

where $\bar{\mathbf{x}}^\mathrm{b}$ is the background mean, $\mathbf{X}^\mathrm{b}$ is the matrix whose columns are individual forecast ensemble members minus the ensemble mean, in model space, $\mathbf{Y}^\mathrm{b}$ is the matrix whose columns are individual forecast ensemble members mapped to the observation points mi-

nus the ensemble mean, in observation space, $\mathbf{R}$ is the matrix of observation uncertainties (diagonal) modified by the localisation weights above, $\mathbf{y}^{\mathrm{o}} - H(\bar{\mathbf{x}}^{\mathrm{b}})$ is the difference between the observations, in observation space, called the "innovation", and

$$\tilde{\mathbf{P}}^{\mathrm{a}} = \left[(N-1)\mathbf{I} + (\mathbf{Y}^{\mathrm{b}})^{\top}\mathbf{R}^{-1}\mathbf{Y}^{\mathrm{b}}\right]^{-1} \tag{3}$$

is a normalization factor. The terms before the innovation are essentially the ratio of the background error covariance to the sum of the background and observational error covariances, but for vectors of background forecasts and observations.

Equation 2 can be written as

$$\bar{\mathbf{x}}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\mathbf{w}^{\mathrm{a}} \tag{4}$$

where $\mathbf{w}^{\mathrm{a}}$ is a matrix of weights assigned to each ensemble member based on the differences between the forecast and observations, the observational error covariance, and the background error covariance. We are not restricted to using $\mathbf{w}^{\mathrm{a}}$ to update solely the quantity that is observed; $\mathbf{w}^{\mathrm{a}}$ can be used to update any prognostic variable. This is only physically sensible where there are correlations between changes in the observed quantity and changes in the other prognostic variable. This can be justified by noting that some quantities are related by known physical laws; for example the atmospheric temperature and wind structure are related, as are the atmospheric temperature and surface pressure. This assimilation of such *indirectly observed* variables is discussed by Navarro et al. (2017) in detail.

In this work we use atmospheric temperature observations to update the GCM atmospheric temperature field (temperature updates temperature, TuT). We then use $\mathbf{w}^{\mathrm{a}}$ computed for temperature to update the zonal and meridional velocities (TuW) and the surface pressure (TuPs, using $\mathbf{w}^{\mathrm{a}}$ at the lowest model grid level).

In most configurations (see below) we also use temperature to update the dust mass mixing ratio (TuD). Navarro et al. (2014) showed that Mars' observed detached dust layers (Heavens et al., 2011) are reproduced by the assimilation when TuD is employed, as dust modifies the temperature field. TuD is applied only where the correspondence between changes in dust and changes in temperature are strong, i.e. where the shortwave

heating rate is above 0.2 K hr^{-1} and the total insolation is above 100 W m^{-2} (this is essentially a condition on time of day and latitude as a function of time of year). It is also constrained to occur only above pseudo-altitude ∼11 km, because the correspondence between dust and heating is based on MCS observations, which have systematic biases close to the ground (Navarro et al., 2014). Whenever the dust mass mixing ratio (MMR) is updated by the assimilation the dust number is also updated, so that any new dust has a particle size distribution with fixed dust effective radius; this radius is a free parameter of the assimilation scheme.

In the forecast step we step forward the GCM in time, starting from the analysis ensemble, from analysis time t_a up to $t_a+\Delta t+t_f$, which is the next analysis time plus the window length forward in time. The cycle then begins anew at the new analysis time $t_a+\Delta t$. During the forecast step the GCM runs freely with no constraints on the dust field. In the assimilation experiments the only processes that can add or remove dust from the atmosphere are the assimilation step and sedimentation onto the surface.

3.3 Observation Operator Using Averaging Kernels

The comparison between forecast and observations $\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{b}})$ is a crucial part of the data assimilation process. This comparison must compare like-with-like to minimise representation errors and ensure that forecast errors are calculated accurately. The observation operator must not only interpolate the forecast to the observation locations, but also reproduce what TIRVIM would observe if it were to observe the forecast atmosphere. When retrievals are used, the correct way to do this is to use the same averaging kernels and prior profile that were used in the retrieval (Rodgers & Connor, 2003; R. N. Hoffman, 2011).

We have improved on earlier versions of this assimilation scheme by basing the observation operator for atmospheric temperature on the TIRVIM retrieval averaging kernels. Earlier work using this scheme by Navarro et al. (2017) had no vertical smoothing in the observation operator. This was (reasonably well) justified as they assimilated limb observations from MCS, which have a vertical resolution of about 5 km (Kleinböhl et al., 2009), similar to the model's vertical resolution throughout most of the atmosphere. Nadir measurements have a coarser vertical resolution, so some smoothing of the fore-

cast profile is required. We believe this is the first time averaging kernels have been applied directly in Mars data assimilation.

Before the averaging kernels are applied, we interpolate from the forecast grid to each observation location linearly in longitude, latitude, and time, and then vertically to the observed σ coordinates, linearly in $\log \sigma$.

The averaging kernels quantify the relative contributions to each point in a retrieved profile from the true and prior atmospheric states at each altitude. The vertical width of the averaging kernel sets the true vertical resolution of the instrument. The retrieval $\mathbf{y}^{\mathrm{o}}$ is related to the true atmospheric state $\hat{\mathbf{x}}$ and the prior profile $\mathbf{x}^{\mathrm{p}}$ by the averaging kernels $\mathbf{A}$ (Rodgers & Connor, 2003) such that

$$\mathbf{y}^{\mathrm{o}} = \mathbf{x}^{\mathrm{p}} + \mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}^{\mathrm{p}}) \tag{5}$$

A perfect retrieval has $\mathbf{A} = \mathbf{I}$, in which case $\mathbf{y}^{\mathrm{o}} = \hat{\mathbf{x}}$, with no prior information retained. A retrieval that contains no information from the true atmospheric state has $\mathbf{A} = \mathbf{0}$, in which case $\mathbf{y}^{\mathrm{o}} = \mathbf{x}^{\mathrm{p}}$. Real instruments are somewhere in between, and $\mathbf{A}$ is a strong function of altitude.

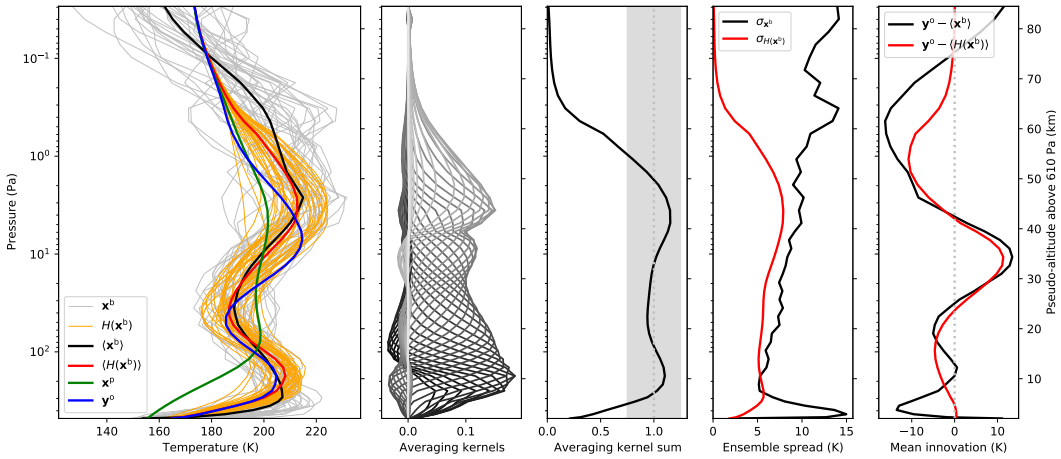
To assimilate a retrieval, the observation operator $H(\mathbf{x}^{\mathrm{b}})$ is a slightly modified form of Eq. 5:

$$H(\mathbf{x}^{\mathrm{b}}) = \mathbf{x}^{\mathrm{p}} + \mathbf{A}(\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{p}}) \tag{6}$$

where $\mathbf{x}^{\mathrm{b}}$ is the background profile interpolated to the retrieved profile levels. The result is a "retrieved forecast", i.e. what a retrieval of a hypothetical TIRVIM spectrum would look like if it observed the forecast atmosphere.

Figure 3 shows two examples of this observation operator applied to a real forecast of TIRVIM retrievals. Figure 3a is a "good" case near the top of the range of retrieval degrees of freedom (3.97), and Fig. 3b is a "bad" case at the low end of the range (2.62). As TIRVIM is a nadir sounder, the averaging kernels (second column) are quite broad, and the "retrieved forecast" ensemble and ensemble mean are smoother than the raw forecast ensemble and ensemble mean.

(a) $L_s = 208.86°$, longitude 34.53°E, latitude 64.94°S, LTST 2.24.



(b) $L_s = 208.84°$, longitude 108.94°W, latitude 40.82°N, LTST 15.87.

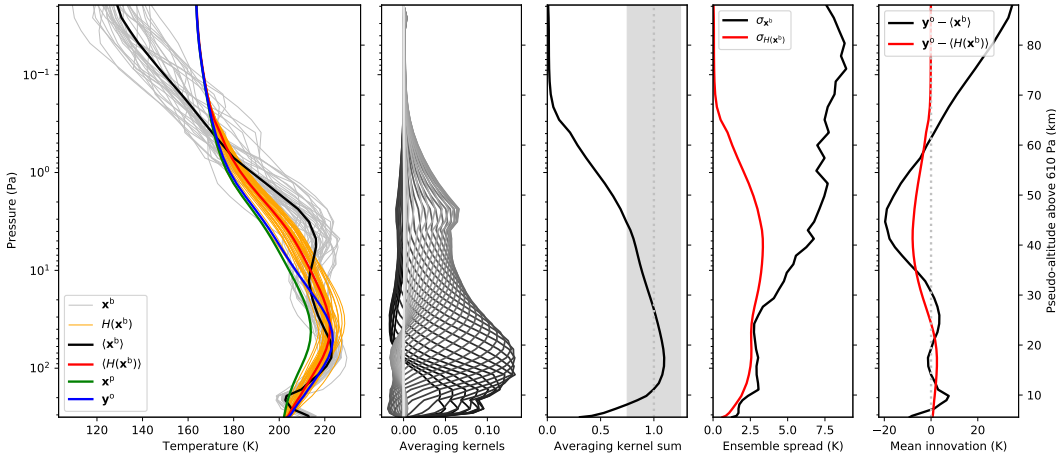


Figure 3. Demonstration of the observation operator acting on the ensemble forecast for two temperature profiles retrieved from TIRVIM radiance spectra. (a) $L_s = 208.86°$, longitude 34.53°E, latitude 64.94°S, local true solar time 2.24, retrieval degrees of freedom 3.97; (b) $L_s = 208.84°$, longitude 108.94°W, latitude 40.82°N, local true solar time 15.87, retrieval degrees of freedom 2.62. The left panel shows the raw forecast ensemble members and ensemble mean (thin grey lines and thick black line, respectively), forecast ensemble members and ensemble mean after the observation operator is applied (thin orange lines and thick red line, respectively), retrieval prior (thick green line), and the retrieved temperature profile, i.e. the "observations" (thick blue line). From left to right, the other panels show: (1) full set of averaging kernel functions, indicated by different greyscale shades; (2) sum over the averaging kernels, where the shaded region indicates the range 0.75–1.25, and the dotted line indicates 1; (3) ensemble spread for the raw forecast ensemble (black) and the forecast ensemble after the observation operator is applied (red); and (4) the mean innovation, i.e. the retrieved temperature profile minus the ensemble mean, showing this for the raw ensemble mean (black) and the ensemble mean after the observation operator is applied (red).

The retrieved temperature profile and the retrieval prior converge at the top of the profiles, indicating that almost all the information in the retrieval comes from the prior there, and so we should avoid assimilating this part of the profile. After the observation operator is applied to the forecast ensemble (Eq. 6), at the top of profile the ensemble collapses onto the retrieval prior, so differences between the forecast and retrieved temperature profiles are artificially small in this region (note also the change in ensemble spread in the 4th column of Fig. 3 before and after Eq. 6 is applied). Conversely, lower in the profiles, most or all of the spread in the forecast ensemble is preserved by the observation operator, hence in this region the difference between the forecast temperature profile and the retrieved temperature profile is real, and so we should include such observations in the assimilation.

To distinguish between these two cases, and quantify how much information comes from the prior and the true atmospheric state, we use the sum over the averaging kernels at each level (Fig. 3, 3rd column). Where the averaging kernel sum is near zero, all the information in the retrieval comes from the prior. Where the averaging kernel sum is close to one (e.g. at 10 Pa), almost all the information in the retrieval comes from the observed radiances. Therefore a suitable way to distinguish between these two cases is to impose a filter on the averaging kernel sum at each level of the temperature profile, and only assimilate observations whose averaging kernel sum falls within a particular range. This ensures that observations which are assimilated contain as much information from the observed radiances as possible, and as little information from the retrieval prior. The grey band in the 3rd column of Fig. 3 identifies the levels that would be retained if a filter of 0.75–1.25 were applied. The main practical difference between the two cases in Figs 3a and 3b is that the vertical range of retained observations extends higher in the atmosphere when the retrieval degrees of freedom are higher.

Figure 4 shows averaging kernel sum statistics for TIRVIM temperature retrievals between MY34 sols 376–424. Between 5–40 km a large majority of the profiles have averaging kernel sums close to one, and even between 1–50 km in many cases. Figure 4c shows how much of the dataset remains when various filters are applied. Between 10–30 km almost all observations are kept, whatever the condition. The largest differences are between 40–60 km, where imposing strict limits of 0.85–1.15 removes at least 50% of the observations, and also in a narrow band of altitudes around 5–10 km. If the fil-

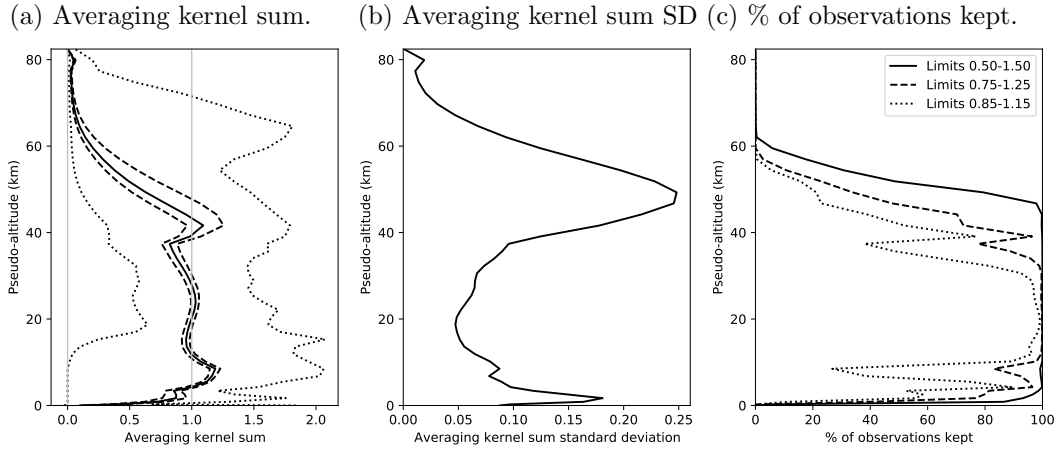


Figure 4. Averaging kernel sum statistics for TIRVIM temperature retrievals between MY34 $L_s = 182.288 - 211.388°$, as a function of pseudo-altitude. (a) Sum over averaging kernels. The black dotted lines are the minimum and maximum, the solid line is the median, the dashed lines are the 25th and 75th percentiles, and the vertical grey lines indicate 0 (all weight on the prior) and 1 (all weight on the measured radiances). (b) Standard deviation of averaging kernel sums. (c) Percentage of observations kept when an averaging kernel sum filter is applied.

ter is relaxed slightly to 0.75–1.25, then most of the observations that were rejected for the strict limits are included.

3.4 Column Dust Optical Depth Assimilation

The second major update to our scheme is to assimilate column dust optical depths (denoted CuD, i.e column-updates-dust). Column dust optical depths have been assimilated by others previously (Lewis & Barker, 2005; Montabone et al., 2014; Ruan et al., 2021), but were not used by Navarro et al. (2014, 2017) in previous work with this scheme. They used MCS observations, and MCS is a limb scanner that does not measure CDOD directly as it rarely sounds the dust profile all the way to the ground. TIRVIM dust retrievals do have to make some assumptions about the vertical dust distribution, which is a limitation of TIRVIM data, but as a nadir instrument it does view the full atmospheric column.

Guerlet et al. (2022) assume dust is well-mixed in the lowest two scale heights and that its mixing ratio decreases linearly with $\log(p)$ above. A quality flag is assigned to each CDOD observation, related to the sensitivity of the radiance to changes in dust load.

Above warm surfaces, dust retrievals are quite reliable, but above colder surfaces, there can be systematic biases and non-unique solutions depending on the assumed dust vertical distribution. Only retrievals with a good quality flag are assimilated; this excludes about a third of the CDOD retrievals (Fig. 2).

We forecast the column dust optical depth at the TIRVIM frequency of 1090 cm^{-1} using the forecast dust mass mixing ratio profile. For each available column dust observation we interpolate the forecast dust MMR field to the observation longitude, latitude, and time. We then integrate the total dust column opacity in the vertical based on Eq. 1 of Madeleine et al. (2011):

$$\tau_{\mathrm{dust}} = \sum_{k=1}^{K} \frac{3}{4} \frac{Q_{\mathrm{ref}} q_k}{\rho_p r_{\mathrm{eff}} g} \Delta p_k \tag{7}$$

where τ_{dust} is the column dust optical depth; K is the total number of model levels; $Q_{\mathrm{ref}} = 1.53112$ is the effective dust extinction coefficient at the TIRVIM wavelength, that accounts for integration over a dust particle size distribution with dust effective radius $r_{\mathrm{eff}} = 1.5$ μm and effective variance $\nu_{\mathrm{eff}} = 0.3$; q_k is the dust mass mixing ratio at level k; $\rho_p = 2500$ kg m^{-3} is the dust density; and Δp_k is the pressure thickness of model level k. We made some of the same assumptions as in the retrieval, such as using the same fixed Q_{ref} and r_{eff}, but in the GCM r_{eff} is a function of location, and the GCM does not assume any particular vertical dust distribution, so these restrictions could be relaxed in future.

Once CDOD is forecast at each observation location, observations are assimilated into the GCM CDOD field using the LETKF in the same way as for atmospheric temperature (Eq. 2), except vertical weighting is not required. As CDOD is not a prognostic variable in the LMD Mars GCM, once we have the CDOD analysis we adjust the dust MMR profile so that its column dust optical depth is the same as the CDOD analysis. As CDOD varies linearly with dust MMR this is a simple linear rescaling of the atmospheric dust profile. The dust number is also updated, as described at the end of Sect. 3.2.

We ran two versions of the dust assimilation. In the first case (CuD) we just used the column dust optical depth observation to update the dust column abundance. In this case we might expect the vertical dust distribution not to reflect the real vertical dust distribution, as the model does not generate detached dust layers spontaneously. We try to avoid this problem with our second version (TuD-CuD), which combines the direct

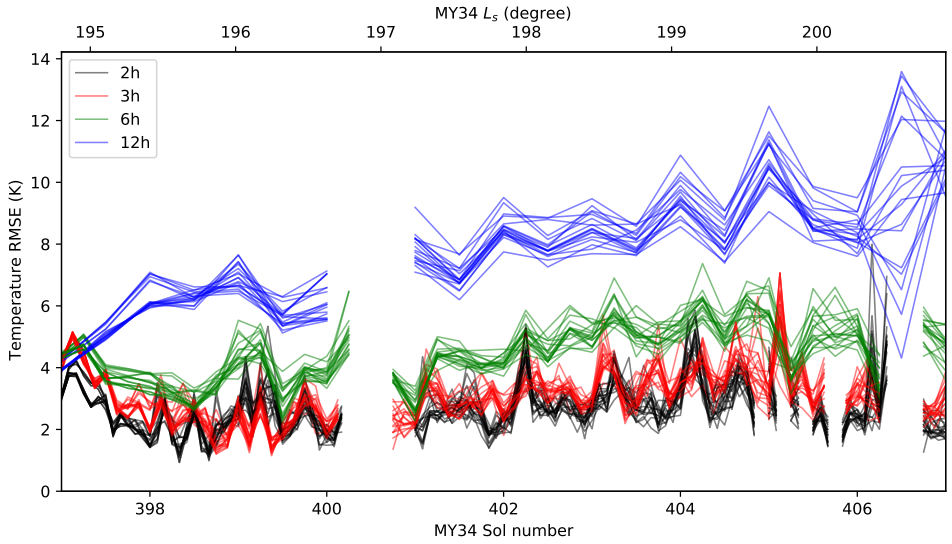
assimilation of CDOD with an indirect update of the dust profile using temperature (TuD). First we perform the TuD step, and then assimilate CDOD. The TuD step should produce a more accurate vertical dust distribution than the model alone, and the column dust assimilation constrains the total amount of dust in each column using real observations. The order of operations is important here: TuD must be done first and CuD second, because otherwise the total amount of dust in the column will not be the amount that is assimilated from observations; if we assimilate CDOD first there is no guarantee that the TuD step will conserve the total amount of dust in the column. This is similar to Ruan et al. (2021), who assimilated MCS dust profiles alongside TES and THEMIS dust column abundances, but note we do not assimilate the dust vertical distribution directly, but only infer it from changes in temperature.

4 Sensitivity Analysis

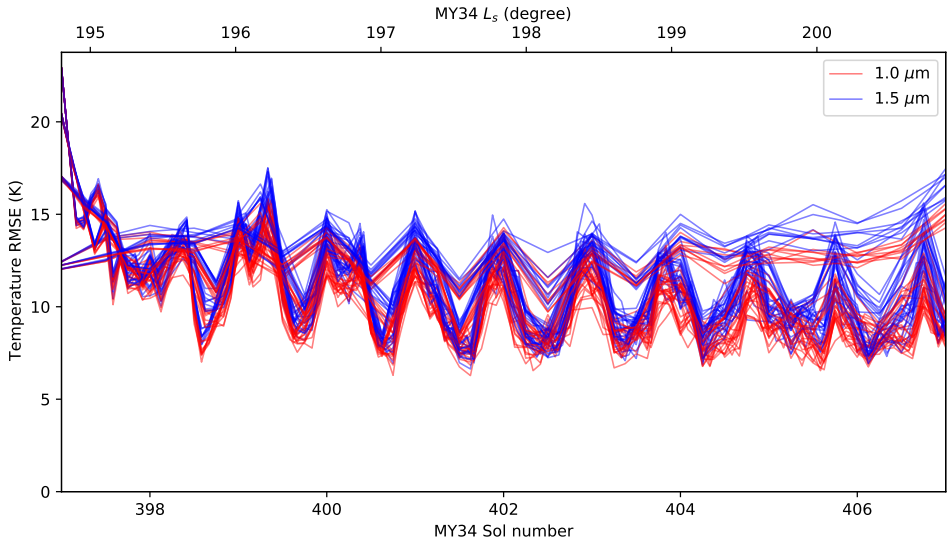
The assimilation scheme has several tuneable parameters. To optimise the temperature assimilation we performed sensitivity tests running short assimilations during the onset of the MY34 GDS (i.e. the period during which the dust concentrations are changing most rapidly). We varied four parameters, and ran an assimilation for each combination of values, 72 in total. We varied the number of ensemble members N (18, 36, and 72), the length of the assimilation cycle Δt (2, 3, 6, and 12 hours), the threshold on the averaging kernel sum filter (0.50–1.50, 0.75–1.25, and 0.85–1.15), and the effective radius of new dust added to the model r_{eff} (1.0 μm and 1.5 μm).

The goal of the ensemble is to sample the range of uncertainties in the forecast model. However, the ensemble is finite, and the mean and covariance of a finite ensemble will differ from the mean and covariance of an infinitely large ensemble. If an infinitely-large ensemble spanning all possible atmospheric states has mean $\bar{x}$ and variance σ^2, then an ensemble of size N will have an expected error in the ensemble mean of $\sigma/\sqrt{N}$, and an expected error in the ensemble variance of $\sigma^2\sqrt{2/(N-1)}$ (Ahn & Fessler, 2003). Hence if the ensemble is too small then outlying values and noise will have a disproportionately large influence on the ensemble mean and variance, so we expect the assimilation to be more accurate for larger ensembles. This is an important part of the rationale for using localization in LETKF assimilation (Hamill et al., 2001). Conversely, larger ensembles consume more resources, and we might expect the forecast step to be more unstable, as the simulations span a larger (and hence more extreme) range of atmospheric states.

(a) Verification against TIRVIM at 10–30 km pseudo-z, coloured by cycle length.



(b) Verification against MCS at 10-100 Pa, coloured by new dust effective radius.



476

477 **Figure 5.** Time series showing RMS error for short test assimilations verified against (a)
478 TIRVIM temperature observations at 10–30 km pseudo-altitude, and (b) MCS temperature ob-
479 servations at 10-100 Pa. In (a) the lines are colour-coded by the length of the assimilation cycle,
480 and in (b) they are colour-coded by the dust effective radius used for new dust. The different
481 lines for a given colour correspond to variations in the other three parameters.

There is a trade-off when choosing the assimilation cycle length between the forecast length and the number of grid points that are updated during each cycle. When the cycle is short, the model has less time to diverge from reality. For Mars the forecast error can saturate in 3–6 hours (Navarro et al. (2017), Fig. 2). But when the analysis cycle is longer a larger fraction of the model grid points are updated during each analysis. This is particularly important for assimilation of tracer quantities such as dust, because the only way that new dust is added to the model state is during the assimilation step.

Figure 4c shows different averaging kernel sum filter limits. The very narrow range (0.85–1.15) excludes many observations near 10 km and 40 km pseudo-altitude, the moderate range (0.75–1.25) encompasses those observations, and the wide range (0.5–1.5) includes a lot of observations, but some will have 50% of their information content from the prior profile.

The dust effective radius r_{eff} for new dust added to the analysis state is tested for 1.0 μm, because Navarro et al. (2017) use this value, and 1.5 μm, because the dust retrievals we assimilate use that value (Guerlet et al., 2022). Dust with smaller radius takes longer to sediment and fall out onto the surface.

We ran each test assimilation from sols 397–407 ($L_s = 194.8{-}200.9°$), using the TuTD configuration (i.e. no CDOD assimilation), and we compared the results with TIRVIM and MCS temperature profiles. We interpolated the analysis in time and space to the TIRVIM (on σ levels) and MCS (on pressure levels) observations. To compare with TIRVIM we applied the averaging kernel operator and the averaging kernel sum filter. We did not apply any vertical smoothing before making the comparison with MCS observations, as the vertical resolution of those observations is comparable with the grid level spacing. We then computed the bias and the root-mean squared error (RMSE) over various pseudo-altitude ranges for TIRVIM (0–10, 10–30, and 30–50 km) and various pressure ranges for MCS (100–1000, 10–100, 1–10, 0.1–1, and 0.01–0.1 Pa). Time series for each of the 72 test assimilations are shown in Fig. 5. Because we ran every possible combination of parameters, we were also able to plot individual comparisons between cases varying just one parameter with all others held constant. These are shown in Fig. 6. In both figures we only show select cases where there were significant differences (gauged by eye) between parameter values.

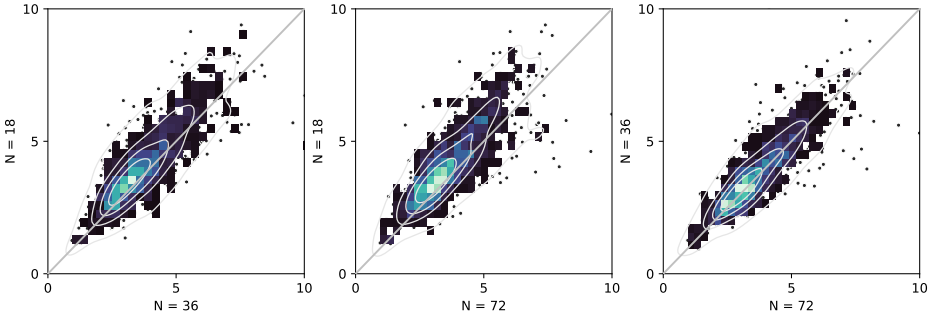
The clearest difference is between the assimilation cycle lengths. Over altitude ranges where TIRVIM is sensitive (i.e. up to 50 km), the shorter the cycle the better the match to observations. Figure 5a shows a representative set of RMSE time series, and Fig. 6b shows scatter plots, both verifying against TIRVIM. This trend remains when compared against MCS, although it is not as strong. A clearer trend when verifying against MCS is that $r_{\text{eff}} = 1.0$ μm for new dust is generally better than 1.5 μm, particularly for 1–100 Pa (i.e. where the TIRVIM observations are), as illustrated in Figs 5b and 6d.

The ensemble size makes a smaller difference, but overall both $N = 36$ and $N = 72$ produce lower RMSE than $N = 18$ (Figure 6a). The differences between the averaging kernel sum filters are smaller still (Fig. 6c); 0.50–1.50 clearly performs the worst, but there is no visible difference between 0.75–1.25 and 0.85–1.15.

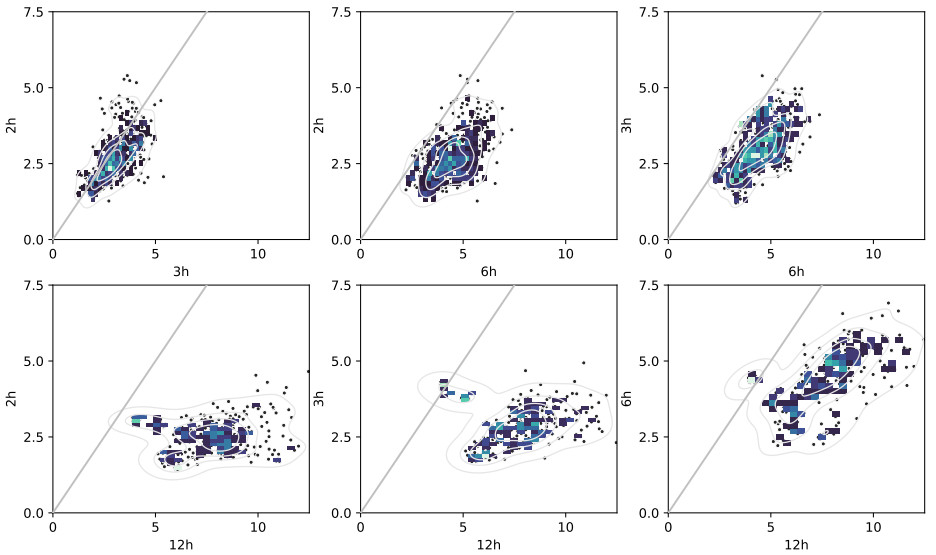
The only parameter that had any effect on the practical aspects of running the assimilation was the ensemble size. The clock time to assimilate 1 sol of data was strongly correlated with ensemble size, with $N = 72$ taking 38–58 min (5–95th percentiles over all $N = 72$ cases) while $N = 36$ took 25–39 min and $N = 18$ took 22–34 min. The forecast step run time depends on the maximum time for any one ensemble member to finish, so the more ensemble members there are, the longer this is. The number of times individual forecast members became unstable was significant only for $N = 72$ (1.2–13.6 times per sol, compared with 0.1–1.2 for $N = 36$). We had planned to run a series of tests with $N = 108$, but chose not to do so as a result.

In conclusion, $r_{\text{eff}} = 1.0$ μm for new dust was generally better than 1.5 μm. An averaging kernel sum filter of 0.75–1.25 was better than 0.5–1.5 in most cases, and marginally better than 0.85–1.15 in some cases. As it allows more observations to be assimilated over a wider range of altitudes, we chose the more inclusive 0.75–1.25 condition. An ensemble size of $N = 36$ was better than $N = 18$ in most cases. $N = 36$ and $N = 72$ were very similar, but $N = 72$ required lots of manual restarts and had more model instability in the forecast step, so we chose $N = 36$ ensemble members. Finally, a 2-hour cycle length was better than 6- and 12-hours in most cases. It was also better than 3 hours in some cases, when compared with TIRVIM observations, but 2- and 3-hour cycles were generally similar. As a 2-hour cycle provides higher time resolution in the final data product, we chose $\Delta t = 2$ hrs.

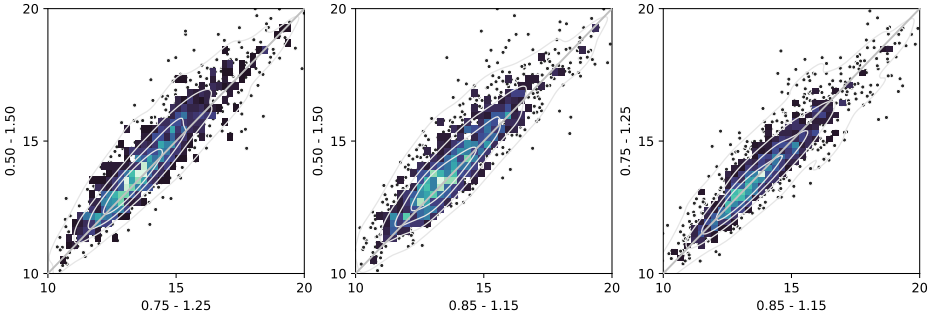
(a) Varying ensemble size N, verifying against TIRVIM at 30–50 km.



(b) Varying cycle length Δt, verifying against TIRVIM at 10–30 km.



(c) Varying averaging kernel sum filter, verifying against MCS at 10–100 Pa.



(d) Varying new dust effective radius r_{eff}, verifying against MCS at 1–10 Pa.

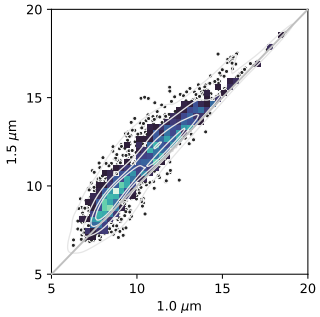


Figure 6. Scatter plots showing temperature RMS error comparisons between 10-sol test assimilations, varying one parameter at a time. Each dot corresponds to a single analysis. Filled and line contours indicate point density, and individual points are outliers. Equality is the diagonal grey line, and all axes are in K (note not all go to zero). In (a) ensemble size N is varied, verifying against TIRVIM at 30–50 km; (b) varies the assimilation cycle length Δt, verifying against TIRVIM at 10–30 km; (c) varies the averaging kernel sum filter range, verifying against MCS at 10-100 Pa; and (d) varies the effective radius for new dust r_{eff}, verifying against MCS at 1-10 Pa.

554 ## 5 Results

555 We ran three versions of the assimilation between sols 376–424 of MY34, encom-
556 passing the onset and peak of the Global Dust Storm. These were TuTD (temperature
557 updates temperature and dust profile), TuT-CuD (temperature updates temperature,
558 and then column dust updates the dust column and hence the dust profile), and TuTD-
559 CuD (temperature updates temperature and dust profile, and then column dust updates
560 the dust column and hence the dust profile). The assimilation period splits nicely into
561 three distinct segments: before the storm (sols 376–385, $L_s = 182.3 - 187.6°$); during
562 the onset of the storm (sols 385–410, $L_s = 187.6 - 202.7°$); and at the peak of the storm
563 (sols 410–424, $L_s = 202.7 - 211.4°$).

564 In assimilation experiments it is usual to compare the analysis with a free running
565 model as a control. Alongside the three assimilations we ran a "GCM ensemble" of GCM
566 simulations without assimilation, where the dust column abundances are constrained to
567 match the MY34 dust scenario (Montabone et al., 2015, 2020). This is not a true free-
568 running model, but it is the standard configuration for this GCM; if the LMD Mars GCM
569 is run without such constraints then dust will sediment out and the model will crash af-
570 ter O(10 sols) as the lack of airborne dust leads to extremely low atmospheric temper-
571 atures. When the GCM is run with a dust scenario, dust is added and removed from the
572 atmosphere by rescaling to the column abundances, by dust lifting from the surface (which
573 occurs at a constant rate), and by sedimentation onto the surface. The results below use
574 the mean of this ensemble when referring to the "GCM ensemble".

575 One additional complication is that the dust scenario used in the GCM ensemble
576 is based on MCS observations. So when our analyses are verified against MCS observa-
577 tions and the assimilation's performance compared against the GCM ensemble, the com-
578 parison between the GCM ensemble and MCS observations is more like an "in-sample
579 comparison", while the comparison between the analyses and MCS observations is a com-
580 pletely independent verification. When we assimilate column dust optical depths directly,
581 the comparison between our analyses and the GCM ensemble is a fairer comparison than
582 when they are not assimilated.

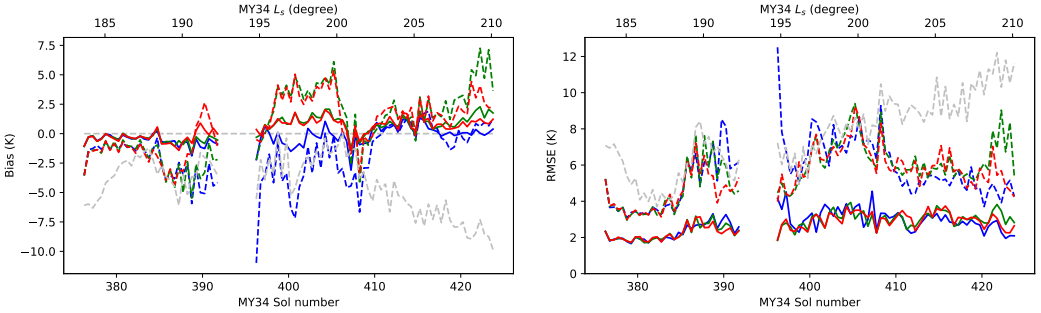
5.1 In-sample Verification against TIRVIM Observations

We computed the forecast and analysis bias and RMSE with respect to the TIRVIM observations that were assimilated, as described in Sect. 4, and over the same vertical ranges. We also computed CDOD at 1090 cm^{-1} from the forecasts and analysss using Eq. 7, and verified them against TIRVIM CDOD observations (i.e. with fixed Q_{ref} and r_{eff}).

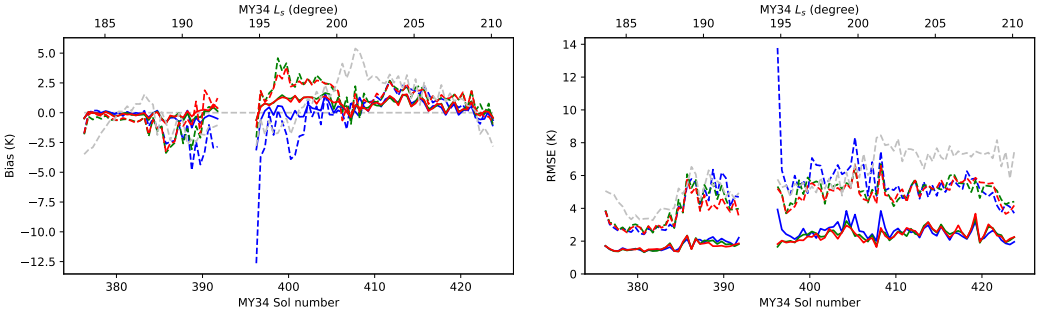
Figure 7 shows the temperature bias and RMSE, and Fig. 8 shows the CDOD bias and RMSE. In both fields the analysis RMSE is typically 50% of the forecast RMSE, and the analysis bias is close to zero, showing that the assimilation reduces the error by adding observational information to the system. The temperature RMSE is close to the observed uncertainty (Fig. 1). There is no systematic bias in the analyses at any altitude. The forecasts typically have a cool bias below 10 km altitude, and between 30–50 km in the TuTD case there is a cool bias while in the other two cases there is a warm bias from about sol 398 onwards. The GCM ensemble is generally biased cool before the dust storm begins, and once the dust storm begins it is biased warm at 0-10 km and cold at 30-50 km. There is a gap in the data between sols 391.7126–396.4636, after which the forecast RMSE is temporarily high, but returns to an equilibrium after about 1 sol (note this also quantifies the convergence time for the assimilation).

For the CDOD observations, when dust is assimilated the error reduction by the assimilation is about 50% before the GDS begins, and 20–40% once the GDS has begun. In the TuTD case (where no dust observations are assimilated, so this is an out-of-sample verification), the error reduction is close to zero, except at the peak of the storm. There is a significant negative bias in the TuTD case (i.e. significantly less dust in the analysis than in the TIRVIM observations). When dust is assimilated directly the analysis error is significantly less, leaving a small negative bias of about 0.1–0.2 in each case. This is comparable with the typical uncertainty in the CDOD measurements (mean 0.16 over this study period) The GCM ensemble has a strong positive bias (about 1.0) compared with the TIRVIM CDOD observations at the peak of the storm.

(a) 30–50 km.



(b) 10–30 km.



(c) 0–10 km.

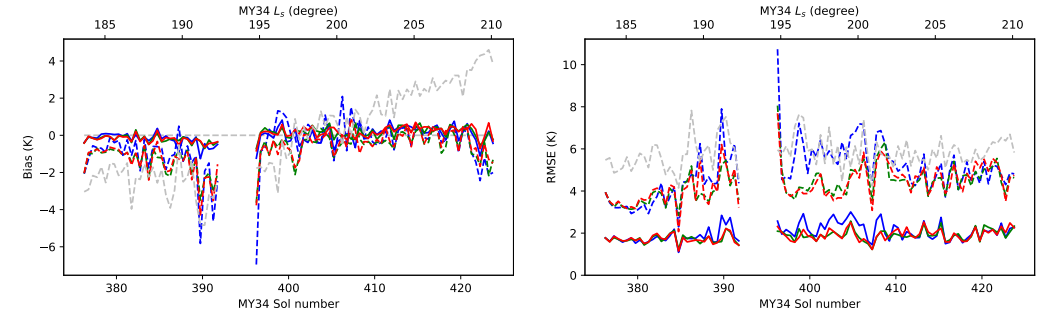


Figure 7. In-sample verification against TIRVIM temperature profiles. The left column shows the bias (analysis/forecast – observations) and the right column shows the RMS error. Data are split by pseudo-altitude: (a) 30–50 km; (b) 10–30 km; and (c) 0–10 km. Solid lines show the analysis mean, and dashed lines show the forecast mean. TuTD is blue, TuT-CuD is green, TuTD-CuD is red, and the GCM ensemble is grey. Data are averaged over 0.5-sol bins.

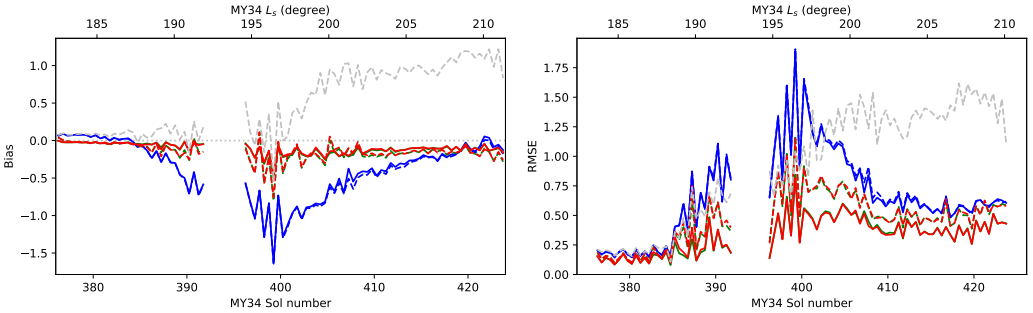


Figure 8. As Fig. 7, but showing in-sample verification against TIRVIM extinction column dust optical depths.

5.2 Out-of-sample Verification against MCS Observations

We verified the analysis and forecast means against MCS temperature observations as described in Sect. 4. We also computed the dust density-scaled opacity (DSO) in extinction at 21.6 μm (Eq. 7 from Madeleine et al. (2011)):

$$\tau_{\mathrm{DSO}} = \frac{3}{4}\frac{Q_{\mathrm{ref}}(r_{\mathrm{eff}})q}{\rho_{\mathrm{dust}}r_{\mathrm{eff}}} \tag{8}$$

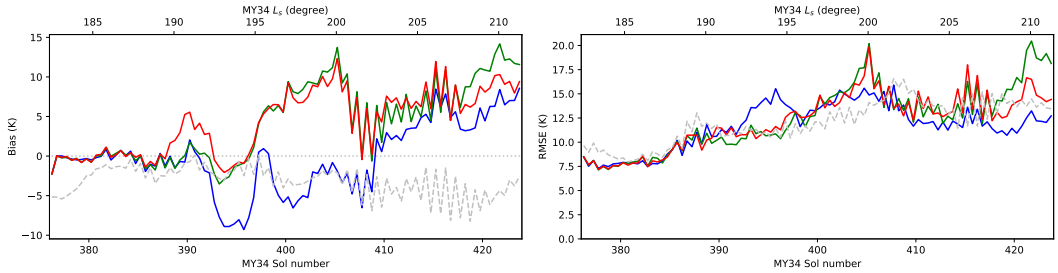
where in this case both r_{eff} and Q_{ref} take the local dust effective radius into account.

Figures 9 and 10 show bias and RMSE time series for temperature and dust DSO, over pressure ranges where the assimilated TIRVIM observations typically lie. These results are clearly poorer than the in-sample verification against TIRVIM observations, but that is not surprising. However, it is important to identify and understand the degree to which they are poorer.

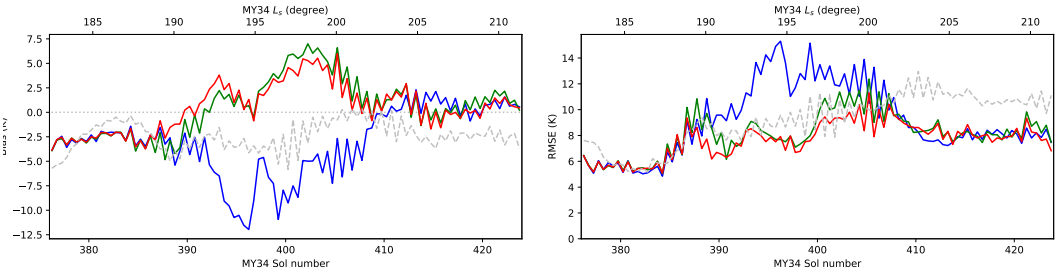
The GCM ensemble, which is constrained by CDOD derived from MCS dust opacities, has a bias and RMSE that falls within the range of values from the three cases assimilating TIRVIM observations. This is encouraging because it implies that, while the absolute errors are larger than when compared with TIRVIM observations, the out-of-sample comparison is comparable with what is effectively an in-sample comparison between the GCM ensemble and MCS observations.

Before the storm there is little difference between the bias and RMSE in the three cases where TIRVIM observations are assimilated. Lower in the atmosphere than 100 Pa, the temperature analysis (as well as the GCM ensemble) has a cold bias of 2–8 K

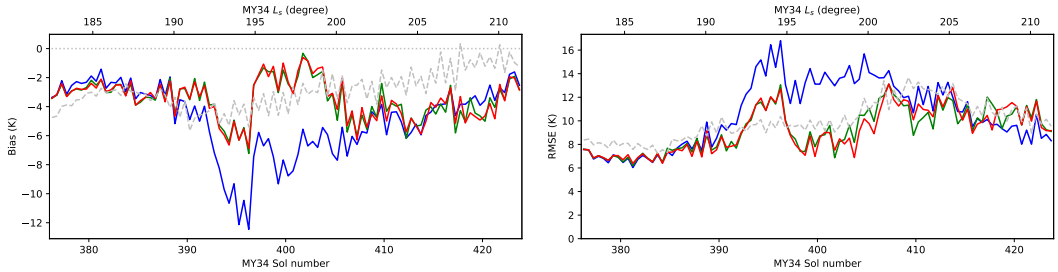
(a) 1–10 Pa



(b) 10–100 Pa



(c) 100–1000 Pa



631

632

633 **Figure 9.** Independent verification against MCS temperature profiles. The left column shows
634 the bias (analysis/forecast – observations) and the right column shows the RMSE. Data are split
635 by pressure ranges: (a) 1-10 Pa; (b) 10-100 Pa; and (c) 100-1000 Pa. Solid lines show analysis
636 means, and the dashed line shows the mean of the GCM ensemble. TuTD is blue, TuT-CuD is
637 green, TuTD-CuD is red, and the GCM ensemble is grey. Data are averaged over 0.5-sol bins.

in all cases; we note that when dust observations are assimilated this bias is closer to zero. Higher in the atmosphere than 100 Pa, the TuTD case has a cold bias during the onset of the storm, while the TuT-CuD and TuTD-CuD cases have a warm bias during that period, with the three cases converging at the peak of the storm. Similarly, during the onset of the storm the RMSE in the TuTD case is significantly poorer than in the two other assimilation cases, showing that it is crucial to assimilate some information about dust when it is changing quickly.

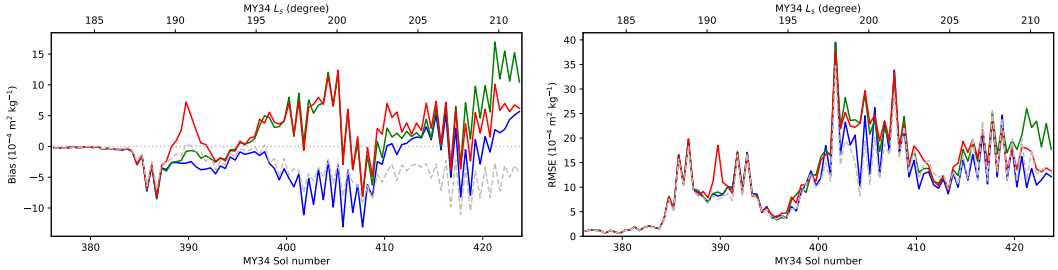
There is significantly lower dust DSO during the onset of the storm in the TuTD case than in the MCS observations, at all pressure ranges shown (Fig. 10). For TuT-CuD and TuTD-CuD there is, on average, a higher dust DSO than in the MCS observations during the onset of the storm, and by the peak of the storm the bias is close to zero between 10-100 Pa and positive in the other two ranges. The GCM ensemble overestimates the amount of dust below the 100 Pa pressure level, and underestimates it above. At all altitudes the RMSE increases from a low level before the onset of the storm, and increases thereafter. As for temperature, we note that the RMSE in the three analyses is comparable with the RMSE in the GCM ensemble.

The amount of dust in the atmosphere is considerably more when dust observations are assimilated, and so the heating effect of more dust helps to explain why, on average, the cases assimilating dust are warmer than the TuTD case. At the peak of the dust storm the analysis temperature bias in all three cases between 10-100 Pa is close to zero, while the GCM ensemble has a cold bias. This is possibly because the GCM ensemble does not exhibit detached dust layers, with its dust distribution monotonically decreasing upwards, so the dust is more concentrated at the surface, heating the lowest atmospheric layers and cooling the higher layers, compared with when dust is assimilated. This is analysed further in the next section.

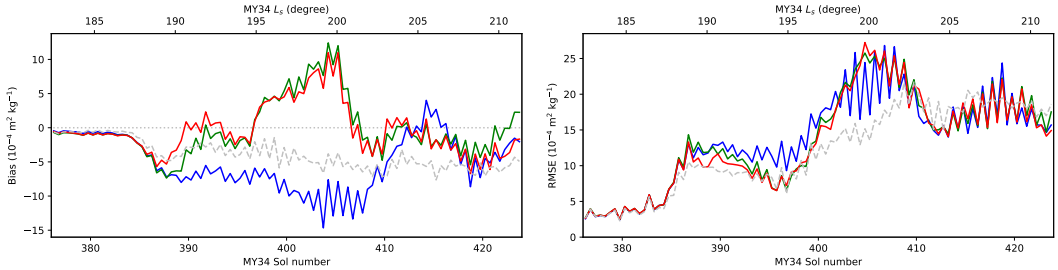
5.3 Temperature and Dust Structure during the Onset of the MY34 GDS

We now compare the atmospheric temperature and dust structure between the analyses, the GCM ensemble, and MCS observations. Figures 11 and 12 show Hovmöller diagrams for temperature and dust DSO at 21.6 μm, at 30 Pa (where MCS dust concentrations are highest during the storm). These cover all latitudes starting before the storm and ending just after its peak. Figure 13 shows vertical cross-sections of temperature at

(a) 1–10 Pa.



(b) 10–100 Pa.



(c) 100–1000 Pa.

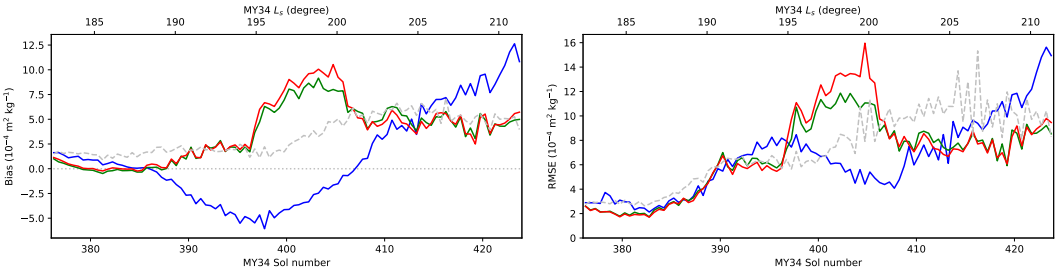


Figure 10. As Fig. 9, but showing independent verification against MCS extinction dust density-scaled opacity profiles.

3 PM and 3 AM local mean solar time (chosen because that is the observation time for MCS), and Fig. 14 shows the same for dust density-scaled opacity.

There is significant heating associated with the global dust storm, with equatorial temperatures rising from 180 K to 225 K over this period (Fig. 11). Assimilating dust observations significantly improves the temperature analysis with respect to the independent MCS observations. When no dust observations are assimilated (Fig. 11a) the temperature change due to the storm over this L_s range is somewhat slower than observed by MCS (although 225 K is reached by the end). But when column dust optical depths are assimilated (Fig. 11 b-c) the change in temperature better matches the MCS observations. Other details also match better, such as the hemispheric asymmetry at the peak of the storm. The differences between the case with and without the TuD step are small, but TuTD-CuD gives marginally better results, particularly for sols 391–396 (Fig. 11 c).

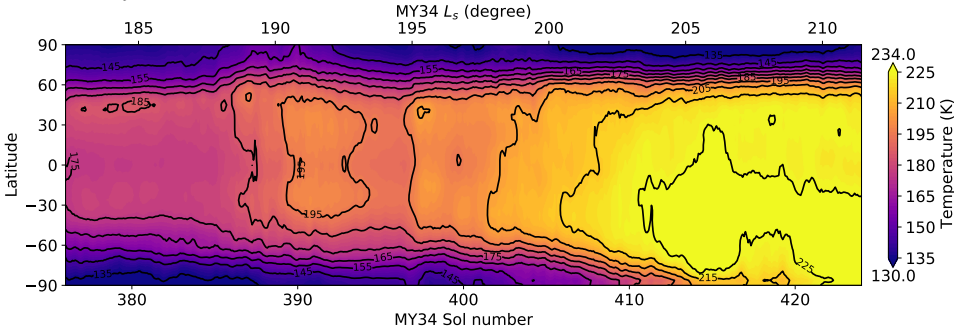
The GCM ensemble captures the minimum at the equator around sol 395 better than the analyses, but overall the analyses match the MCS observations better. The former underestimates the atmospheric temperature at most latitudes, and does not fully reproduce the observed hemispheric asymmetry at the peak of the storm.

The temperature results at 3 AM are similar, although the differences between the GCM ensemble and the analyses are smaller. Hovmöller diagrams for 3 AM are shown in Supporting Fig. S1.

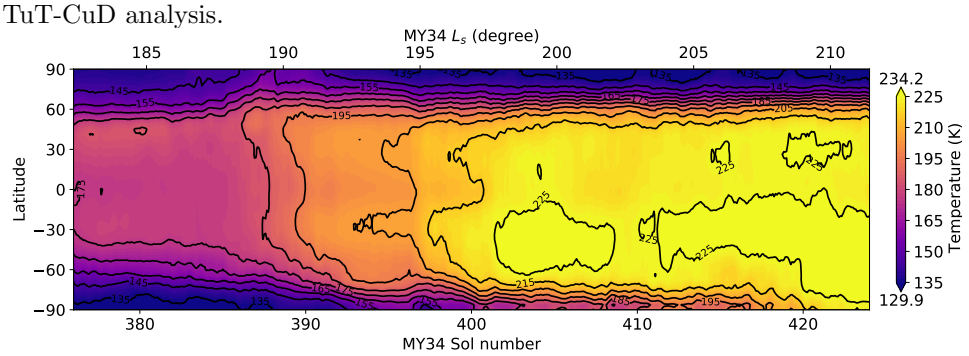
The differences in dust DSO between the various configurations (Fig. 12) are larger than for temperature. MCS observations show a peak in the dust DSO around 60×10^{-4} m^2 kg^{-1}. The GCM ensemble (Fig. 12d), whose column abundance is constrained by MCS observations, matches the dust DSO reasonably well, at least in terms of its evolution during the onset of the storm up to the peak, although the distribution is more strongly concentrated at the equator than for the MCS observations.

The three analyses each contain overall less dust than do the MCS observations. The largest difference is between TuTD and the other two cases. For TuTD (Fig. 12a), dust opacity increases slowly during the onset of the storm, and only begins to change rapidly close to its peak. The other two cases (Figs 12 b-c) follow the increase in dust during the onset of the storm better, although neither reach the peak opacity measured by MCS at this pressure by the peak of the storm. None of the three analyses reproduce

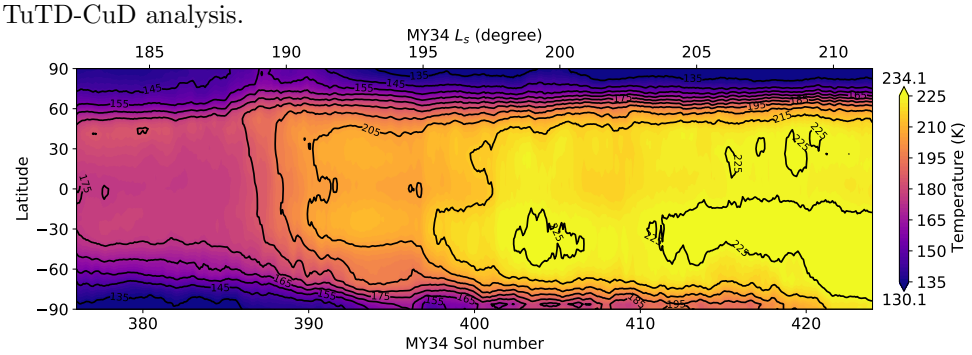
(a) TuTD analysis.



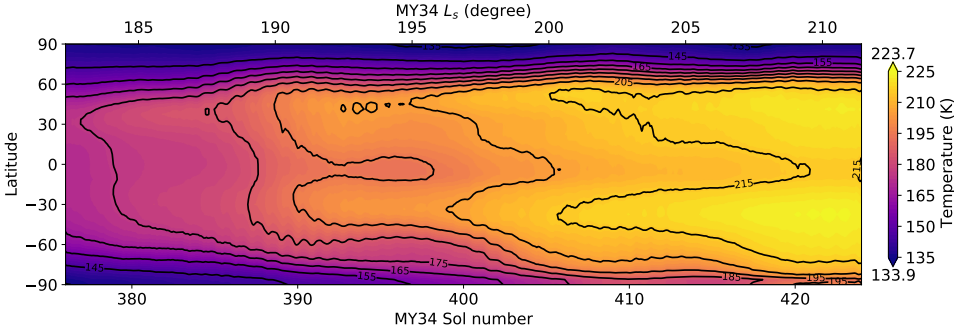
(b) TuT-CuD analysis.



(c) TuTD-CuD analysis.



(d) GCM ensemble.



(e) MRO-MCS observations.

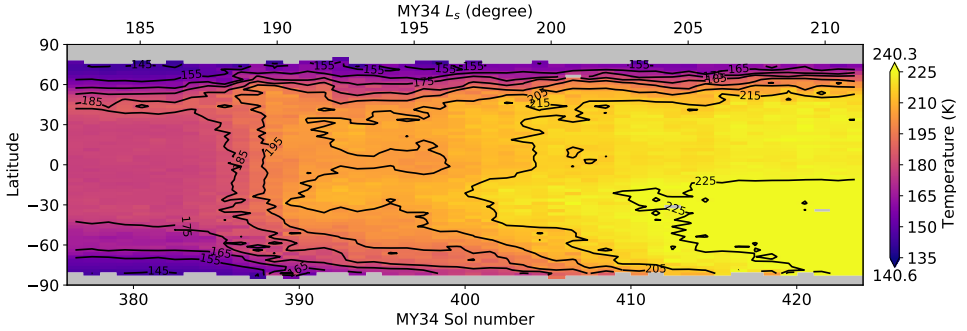
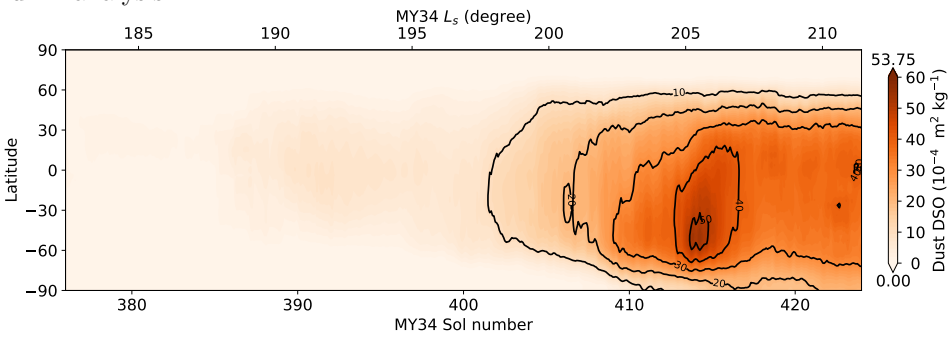
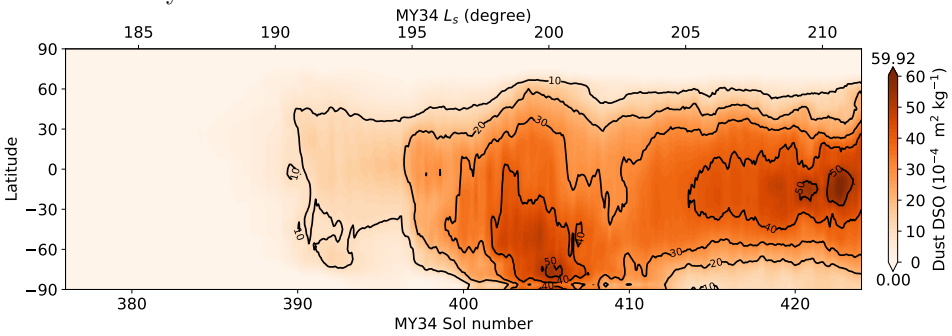


Figure 11. Hovmöller diagrams showing analysis temperature at 30 Pa at 3 PM local mean solar time (LMST). (a) TuTD analysis; (b) TuT-CuD analysis; (c) TuTD-CuD analysis; (d) GCM ensemble; (e) MRO-MCS observations. Data are smoothed using a 1-sol running mean. Grey indicates missing data. Line contours are every 10 K. MCS observations between 2-4 PM LMST are used; analyses are interpolated to 3 PM LMST. Maxima and minima are shown above and below the colour bar.

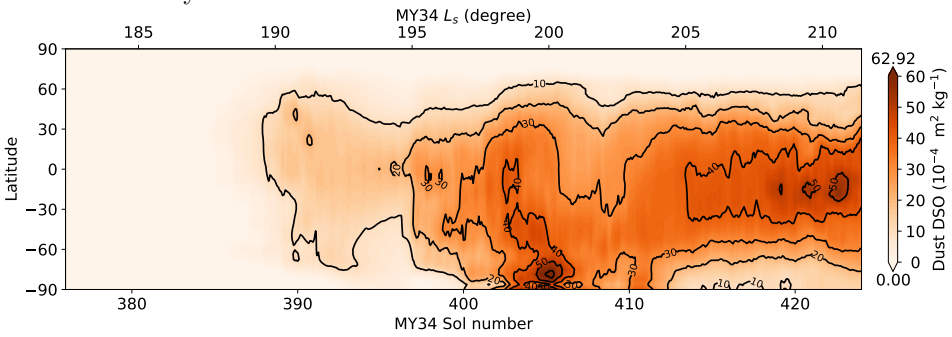
(a) TuTD analysis.



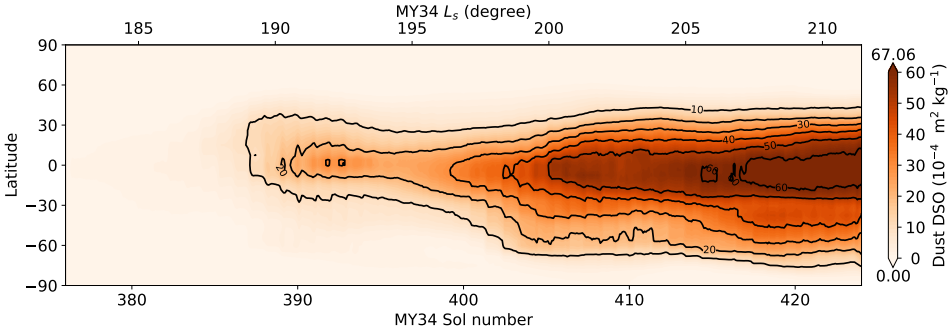
(b) TuT-CuD analysis.



(c) TuTD-CuD analysis.



(d) GCM ensemble.



(e) MRO-MCS observations.

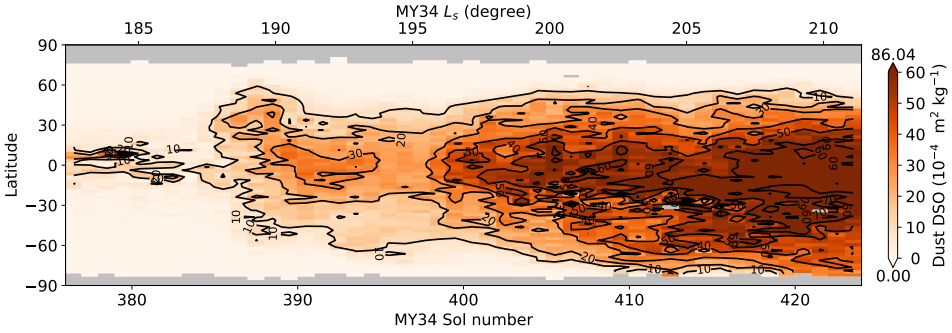


Figure 12. As Fig. 11 but showing Hovmöller diagrams for dust density-scaled opacity at 21.6 μm at 30 Pa. Line contours are every 10×10^{-4} m^2 kg^{-1}.

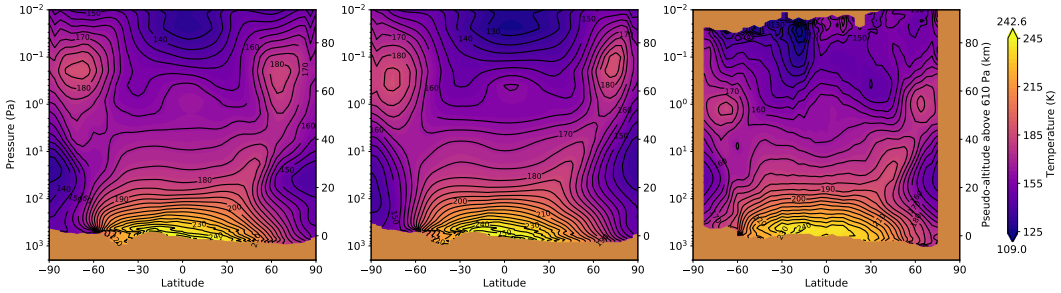
the short-lived peak at the equator between sols 389–393, but we note this partially coincides with a gap in the TIRVIM data. Nevertheless, it is clearly better to have assimilated the dust information directly, both in terms of the temperature analysis and the dust analysis.

The equivalent plots at 3 AM, shown in Supporting Fig. S2, show a more favourable comparison between the TuT-CuD / TuTD-CuD analyses and MCS observations. Both reach the peak dust opacity seen in MCS observations by the peak of the storm (which we note the GCM ensemble does not do at 3 AM). They also reasonably reproduce the growth in dust concentration from sol 397 onwards. Kleinböhl et al. (2020) discuss the significant diurnal variations of temperature and dust opacity in the MCS observations during the dust storm in some detail.

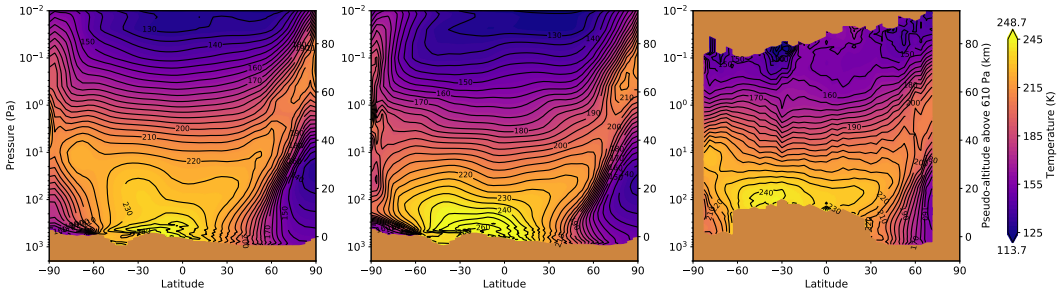
Figures 13 and 14 show vertical cross-sections of the temperature and dust density-scaled opacity before and at the peak of the global dust storm. These figures show how well the analysis reproduces the vertical structure, compared with the GCM ensemble and MCS observed profiles. In these figures and hereafter we only use our "best case" assimilation TuTD-CuD, alongside the GCM ensemble.

Both the temperature and dust opacity increase significantly during the storm, during daytime and nighttime. Unfortunately it is not possible to compare directly with daytime MCS observations at the peak of the storm near the surface, as the high dust concentration prevents a reliable retrieval of both temperature and dust (Fig. 13b, right). Before the storm, both the analysis and the GCM ensemble match the MCS observations well during nighttime. During the daytime the temperatures below 100 Pa are generally lower than the MCS observations. This may be because at low altitudes during daytime the MCS observations are biased warm compared with TIRVIM due to a lack of retrievals in cloudy regions; see Guerlet et al. (2022), Fig. 19. The analysis and GCM ensemble both have a temperature maximum of 250–260 K close to the surface during the daytime at the peak of the storm. The maximum MCS temperature is at about 100 Pa, and appears to be a local maximum. In northern midlatitudes the observed profiles do reach the ground, and the analysis reasonably reproduces the observed meridional temperature gradient from about 30° poleward (Fig. 13b). Both the analysis and the GCM ensemble are cooler than MCS at its maximum temperature point around 100 Pa, 30–50°S, with the analysis about 10–20 K cooler, and the GCM ensemble about 5–10 K cooler

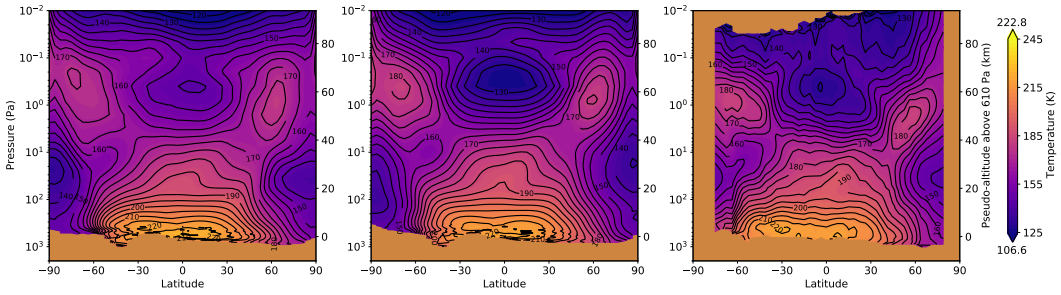
(a) 3 PM LMST, before the storm (sols 380–384).



(b) 3 PM LMST, at the peak of the storm (sols 416–420).



(c) 3 AM LMST, before the storm (sols 380–384).



(d) 3 AM LMST, at the peak of the storm (sols 416–420).

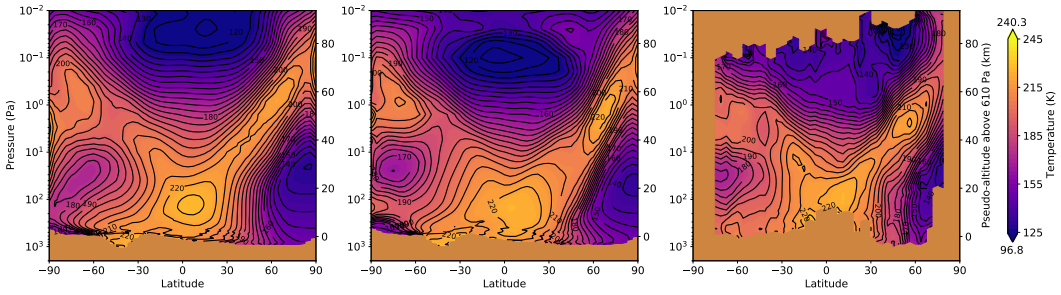


Figure 13. Temperature cross-sections at 3 PM and 3 AM local mean solar time. The left column shows the TuTD-CuD analysis, the middle column shows the GCM ensemble, and the right column shows MCS observations (up to one hour either side of the target local time). (a-b) show 3 PM, and (c-d) show 3 AM. (a) and (c) show before the storm, averaged over sols 380–384, and (b) and (d) show the peak of the storm, averaged over sols 416–420. The right vertical axis shows the pseudo-altitude above 610 Pa. The numbers at the end of each colour bar indicate the maximum and minimum values in the MCS panels. Line contours are spaced every 5 K. Flat brown areas show missing data (either below the surface, or where there are no observations).

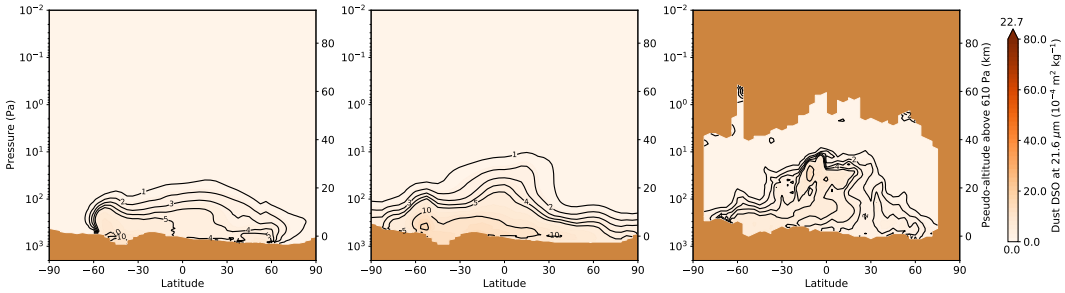
than observed. Both the analysis and the GCM ensemble reproduce the tongue of warm air in the middle atmosphere at northern hemisphere midlatitudes, which appears at the peak of the storm.

At 3 AM the analysis and GCM ensemble reproduce the MCS observations better at the peak of the storm than during the daytime (Fig. 13d). The analysis reproduces the local temperature maximum around 100 Pa that appears in the MCS observations better than does the GCM ensemble. The analysis also better reproduces the meridional temperature variation below 10 Pa. The analysis and GCM ensemble have quite different temperature structures near the top of the domain near the equator, with the local temperature minimum in the GCM ensemble being about a decade in pressure below where it is in the analysis. However, the MCS observations do not allow us to distinguish between these two, as the observations cease just above 0.1 Pa, but there does not appear to be an obvious temperature minimum as the GCM ensemble predicts. Note also that the assimilated TIRVIM observations typically go up to about 1 Pa, so above this any differences are due to the GCM's response to assimilated data rather than being constrained directly by observations.

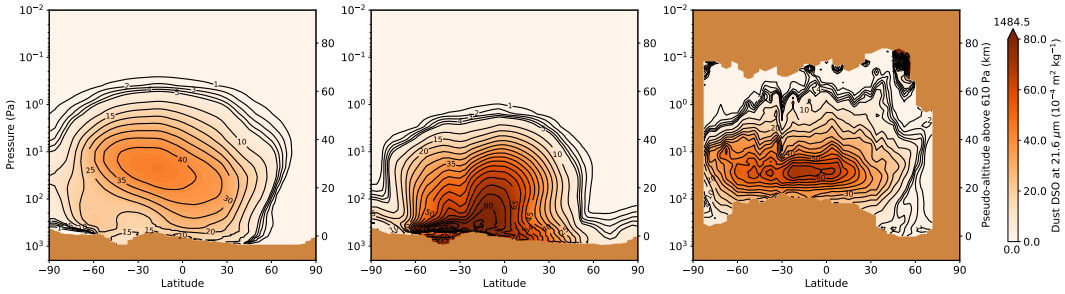
As in the Hovmöller diagrams, there is a larger difference between the analysis and the GCM ensemble in the dust opacity cross-sections (Fig. 14) than in temperature (Fig. 13). Before the storm, the amount of dust peaks around 5×10^{-4} m^2 kg^{-1} in the analysis and 10×10^{-4} m^2 kg^{-1} in the GCM ensemble, with little diurnal variation (Fig. 14 a, c). The MCS observations peak at about 5×10^{-4} m^2 kg^{-1}, but the main difference is that the dust is more vertically extended in the observations, with a clear "detached dust layer" peak near the equator during the night-time around 100 Pa of about 5×10^{-4} m^2 kg^{-1}, which is not visible in either the analysis or the GCM ensemble.

At the peak of the storm there are larger differences between daytime (Fig. 14b) and nighttime (Fig. 14d), as well as between the analysis, GCM ensemble, and MCS observations (Fig. 14, different columns). In general, the GCM ensemble contains significantly more dust than either the analysis or the MCS observations. The peak in the GCM ensemble is close to the ground in the southern hemisphere, about 90×10^{-4} m^2 kg^{-1} at both 3 PM and 3 AM, while the MCS observations peak at about 30 Pa, with maxima around 60×10^{-4} m^2 kg^{-1} at 3 PM and a concentrated peak of 50×10^{-4} m^2 kg^{-1} at 3 AM.

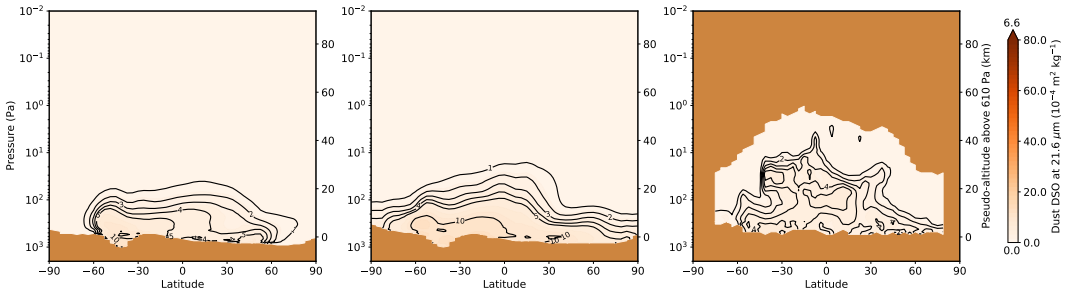
(a) 3 PM LMST, before the storm (sols 380–384).



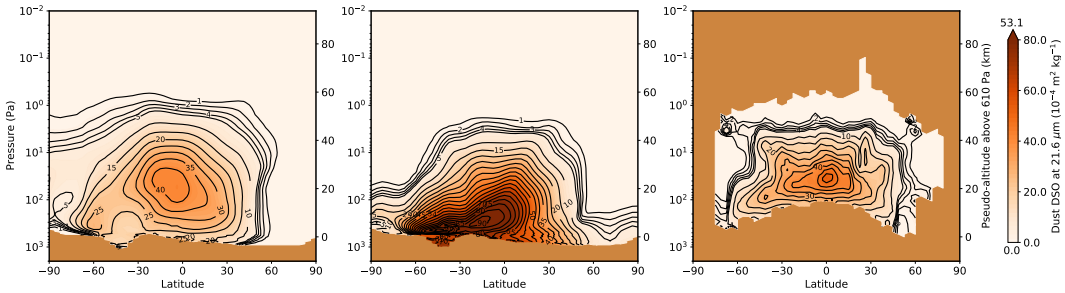
(b) 3 PM LMST, at the peak of the storm (sols 416–420).



(c) 3 AM LMST, before the storm (sols 380–384).



(d) 3 AM LMST, at the peak of the storm (sols 416–420).



780

781 **Figure 14.** As Fig. 13, but showing dust density-scaled opacity at 21.6 μm at 3 PM and 3
782 AM local mean solar time. Line contours are spaced every 10^{-4} m^2 kg^{-1} from 1–4 $\times$ 10^{-4} m^2
783 kg^{-1}, and then every 5 $\times$ 10^{-4} m^2 kg^{-1} thereafter.

The analysis reproduces the vertical and latitudinal distribution of dust opacities significantly better than does the GCM ensemble. In particular, it very clearly contains a detached dust layer at both 3 PM and 3 AM, with a broad peak between 10–100 Pa, as seen in the MCS observations, while the GCM ensemble peaks near the surface (although note not exactly at the surface). The peak dust DSO in the analysis is about 40×10^{-4} m^2 kg^{-1} at both 3 PM and 3 AM. During the daytime this is an underestimate, as the dust opacity peak in the MCS observations is above 50×10^{-4} m^2 kg^{-1} between 60°S and 20°N. During nighttime the MCS observations peak at 50×10^{-4} m^2 kg^{-1} near the equator. However, there is significantly less dust than during the daytime, with the opacity above 40×10^{-4} m^2 kg^{-1} between 20°S and 20°N. This distribution is matched well by the analysis dust opacity at this time, which is above 40×10^{-4} m^2 kg^{-1} between 20°S and the equator.

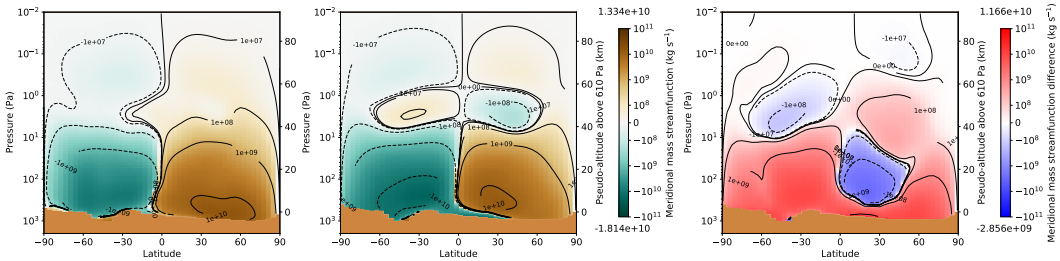
5.4 Wind Structure Retrieved by Assimilation

Having demonstrated that our reanalysis state fits the independent MCS observations reasonably well, we now turn to wind, a quantity that is not observed but which is retrieved by the assimilation. This is one of the main strengths of the data assimilation process.

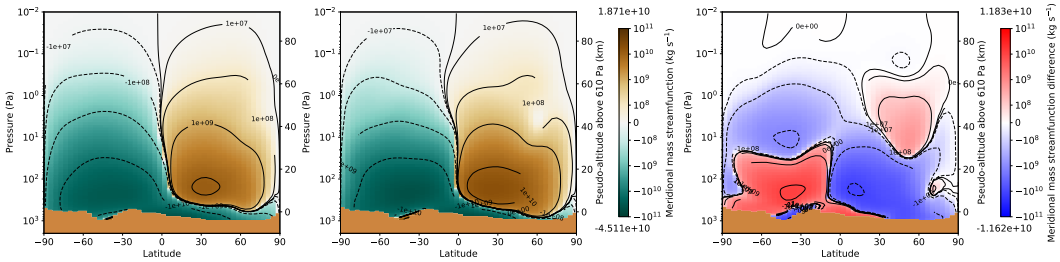
Figures 15 and 16 show the meridional mass stream function and zonal velocity at 3 PM and 3 AM local mean solar time averaged over two time periods before the storm and at its peak. Before the storm, the daytime overturning circulation (Fig. 15 a-b) shows a Hadley cell covering the whole of each hemisphere. There is a weak counter-rotating cell higher than 10 Pa, which only appears in the GCM ensemble. The study period is close to the autumnal equinox, so the Hadley cell is approximately symmetric about the equator. During nighttime (Fig. 15 c-d) the overturning circulation is reversed, with the descending branch of the main overturning cell at the equator, with a weaker thermally direct cell between 1–10 Pa. The meridional overturning circulation is generally weaker in the analysis than in the GCM ensemble.

The overturning circulation strengthens during the daytime as the storm progresses, and extends higher into the atmosphere, significantly strengthening the overturning circulation aloft. The thermally-indirect cell present in the GCM ensemble disappears, leaving a single Hadley cell spanning the whole model domain from the surface to 100 km

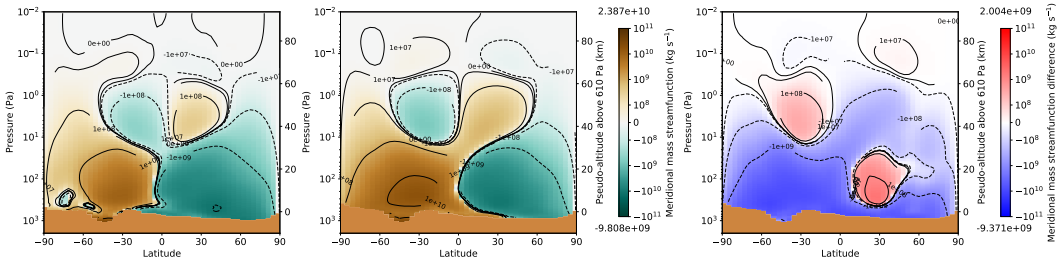
(a) 3 PM LMST, before the storm (sols 380–384).



(b) 3 PM LMST, at the peak of the storm (sols 416–420).



(c) 3 AM LMST, before the storm (sols 380–384).



(d) 3 AM LMST, at the peak of the storm (sols 416–420).

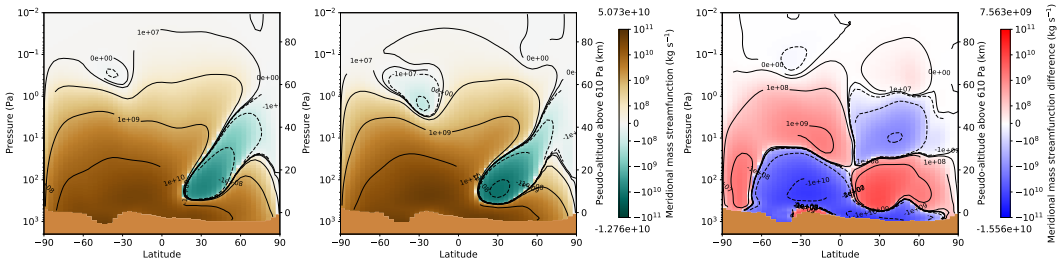


Figure 15. As Fig. 13, but showing the meridional mass streamfunction at 3 PM and 3 AM local mean solar time. The right hand column shows the difference between the analysis and GCM ensemble (analysis — GCM). In the left two columns brown is clockwise motion and blue is anticlockwise motion. In the right column red means more clockwise, and blue means more anticlockwise. Line contours are spaced every factor of 10 between 10^8 and 10^{11} kg s^{-1}.

pseudo-altitude. Close to the surface, the southern hemisphere cell extends about 40° into the northern hemisphere. At nighttime the circulation changes considerably. The symmetric Hadley cells are severely disrupted, and a cross-equatorial cell develops which covers most latitudes with a weak Hadley circulation in the northern polar region.

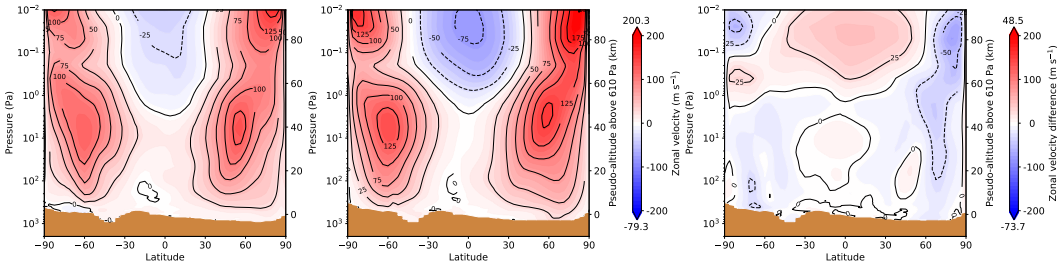
The zonal velocity response to the storm (Fig. 16) is noticeably asymmetric about the equator, both during daytime and nighttime. Like the meridional overturning circulation, this change is more profound during nighttime. Before the storm there are prograde jets in both hemispheres with a retrograde jet above 1 Pa at low latitudes, and these jets are slightly weaker at 3 AM. By the peak of the storm, the equatorial retrograde jet and the prograde northern hemisphere jet increase in speed by about 25 m s^{-1} and 75 m s^{-1} respectively, while the southern hemisphere prograde jet remains approximately the same speed but its vertical extent shrinks, becoming a more focused jet around 75°S / 10 Pa. During nighttime, the northern prograde jet strengthens by about 50 m s^{-1}, but the southern prograde jet weakens significantly by about 75 m s^{-1}, almost disappearing. At the peak of the storm, the two prograde jets differ in speed by 125 m s^{-1}. The strong strengthening of the northern hemisphere jet at both 3 PM and 3 AM is consistent with the steepening of the latitudinal temperature gradient near 60°N at the peak of the storm (Fig. 13 b-d), via thermal wind balance. A similar argument explains the weakening of the southern prograde jet: the temperature contours are flattened in the southern hemisphere around 10 Pa (Fig. 13 a-b)

Differences between the analysis and the GCM ensemble are relatively small compared with the magnitude of the jets, with typical differences up to ± 25 m s^{-1} over most of the domain. The analysis winds are typically more eastward at lower latitudes than the GCM ensemble, in particular above 10 Pa, and they are typically more westward at higher latitudes.

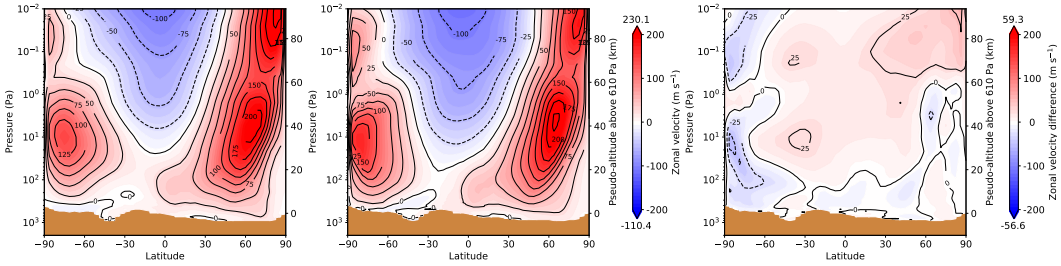
5.5 Diurnal Cycle and Thermal Tides

One of the reasons for the design of the ExoMars TGO orbit is that it samples the Martian atmosphere at different local times of day. In this section we present how this affects the analysis, and analyse how the diurnal cycle changes during the global dust storm. Assimilation constrains the atmosphere at particular local times, and then the forecast model transfers that information to other local times during model integration,

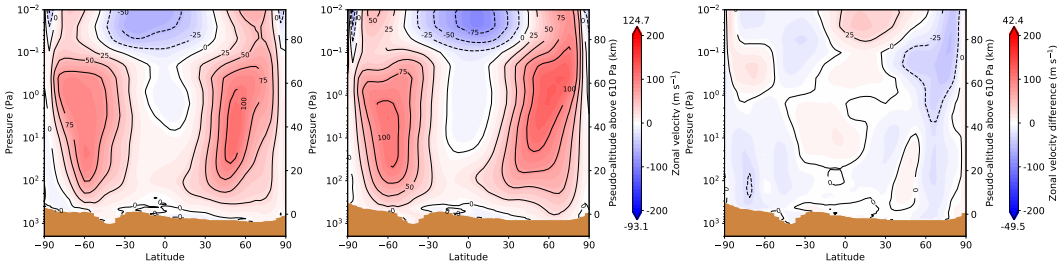
(a) 3 PM LMST, before the storm (sols 380–384).



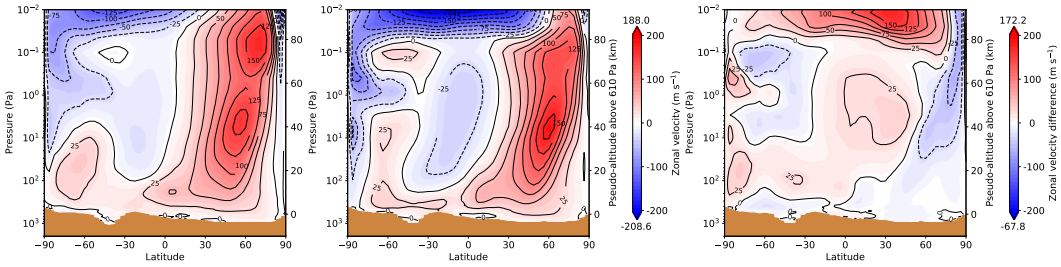
(b) 3 PM LMST, at the peak of the storm (sols 416–420).



(c) 3 AM LMST, before the storm (sols 380–384).



(d) 3 AM LMST, at the peak of the storm (sols 416–420)



841

842 **Figure 16.** As Fig. 15, but showing the zonal velocity at 3 PM and 3 AM local mean solar

843 time. Line contours are spaced every 25 m s^{-1}.

871 further constraining the atmospheric structure at local times that were not observed. For
872 sols 380–384 the bulk of the TIRVIM observations are at 1–2 AM / PM, and for sols 416–
873 420 they are at 5–6 AM / PM (Fig. 2c). First we look at the diurnal cycle in two dif-
874 ferent ways (Figs 17–18) and then analyse the migrating Sun-synchronous tide modes
875 (Figs 19–20). Note that where MCS observations are not part of the analysis, we do our
876 computations in LTST, as that is more dynamically relevant than LMST, and it is the
877 local time used in the GCM.

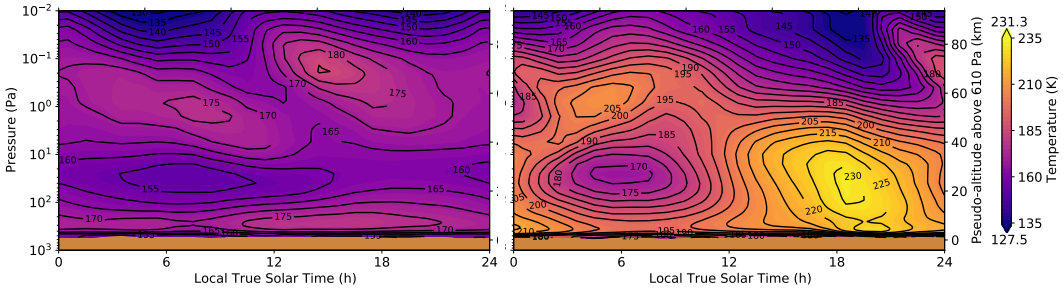
878 Figure 17 shows the temperature diurnal cycle averaged over two different latitude
879 bands before and at the peak of the global dust storm. We interpolated the analysis and
880 the GCM ensemble to fixed pressure levels, and then interpolated to each local time in
881 1-hour steps. We then averaged the temperature at each pressure and local time over
882 latitude and sol ranges. This was done at 60–65° S (to match Fig. 4.13 of M. D. Smith
883 et al. (2017)), and over a latitude band at the equator.

892 There is a significant difference between the diurnal cycle before and at the peak
893 of the storm. At 60–65° S (Fig. 17 a-b) the range of temperatures at all altitudes sig-
894 nificantly increases with the peak to peak diurnal variability increasing from about 10
895 K at 30 Pa before the storm to about 60 K at the peak of the storm. In the middle at-
896 mosphere (0.1–1 Pa), the 12-hour oscillation is overwhelmed by a 24-hour oscillation that
897 peaks between 3-6 AM.

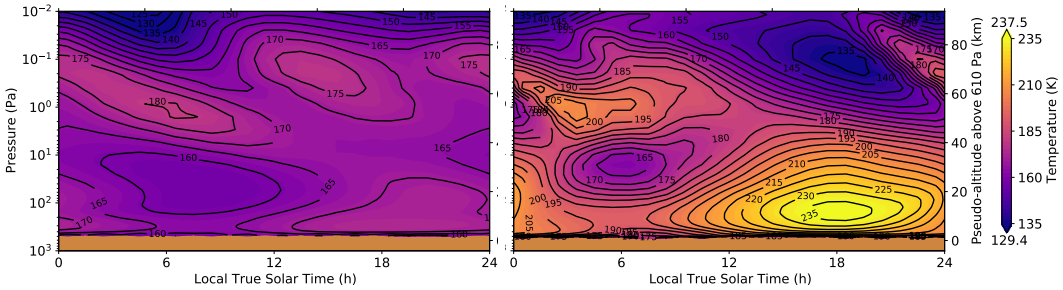
898 The main difference between the analysis and GCM ensemble is in the lower at-
899 mosphere temperature maximum around 6 PM. In the analysis this is concentrated in
900 the local time direction between 6-8 PM, and is extended in the pressure direction up
901 to about 10 Pa. In the GCM ensemble the peak is concentrated closer to the ground,
902 up to about 100 Pa, but spans a longer range of local times. The peak during the night
903 time in the lower atmosphere is quite similar in the analysis and the GCM ensemble.

904 At the equator (Fig. 17 c-d) we also see the 12-hour oscillation in the upper part
905 of the domain weaken. In the analysis the peak around 12 PM between 0.1–1 Pa almost
906 completely disappears during the peak of the storm, while in the GCM ensemble it is
907 still distinguishable. Lower in the atmosphere the 12-hour oscillation is strengthened both
908 in the the analysis and the GCM ensemble at the peak of the storm, and the 24-hour
909 oscillation is less clear, although in the analysis the 12-hour oscillation is dominant while
910 in the GCM ensemble the 24-hour oscillation is dominant.

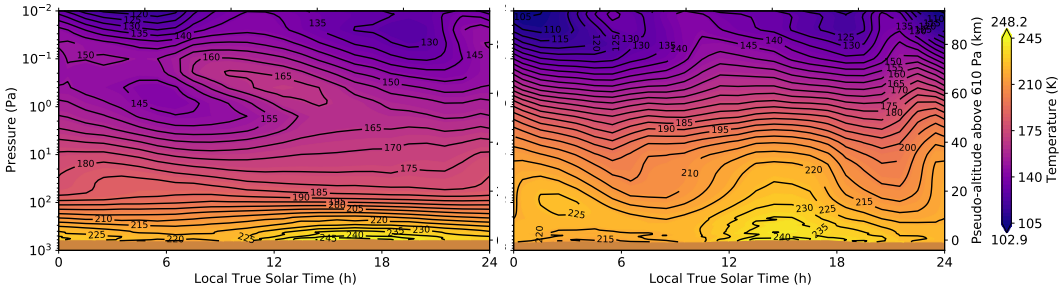
(a) TuTD-CuD analysis averaged over 60–65°S.



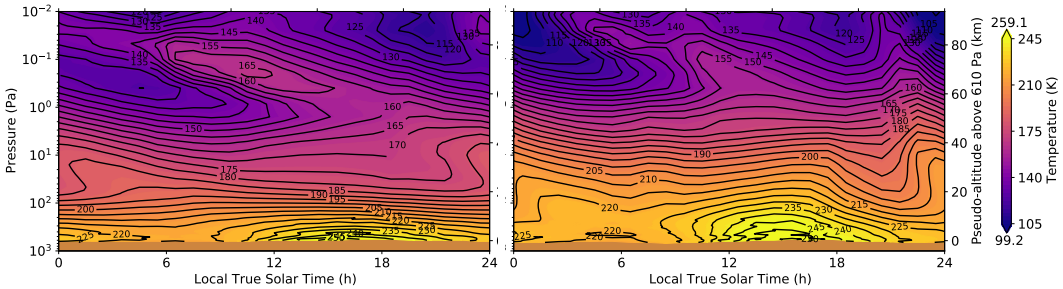
(b) GCM ensemble averaged over 60–65°S.



(c) TuTD-CuD analysis averaged over 10°S–10°N.



(d) GCM ensemble averaged over 10°S–10°N.



884

885 **Figure 17.** Vertical temperature structure against local true solar time. The left column
886 shows a mean over MY34 sols 380–384 (before the storm) and the right column shows sols 416–
887 420 (at the peak of the storm). The colour scale is the same for both columns, but note the
888 colour scales for (a-b) and (c-d) are slightly different. From top: (a) TuTD-CuD analysis aver-
889 aged over 60–65°S; (b) GCM ensemble for 60–65°S; (c) TuTD-CuD analysis for 10°S–10°N; (d)
890 GCM ensemble for 10°S–10°N. Numbers above and below the colour bars indicate the minimum
891 and maximum temperature for sols 416–420, and line contours are every 5 K.

(a) TuTD-CuD analysis.

(b) GCM ensemble.

(c) MCS observations.

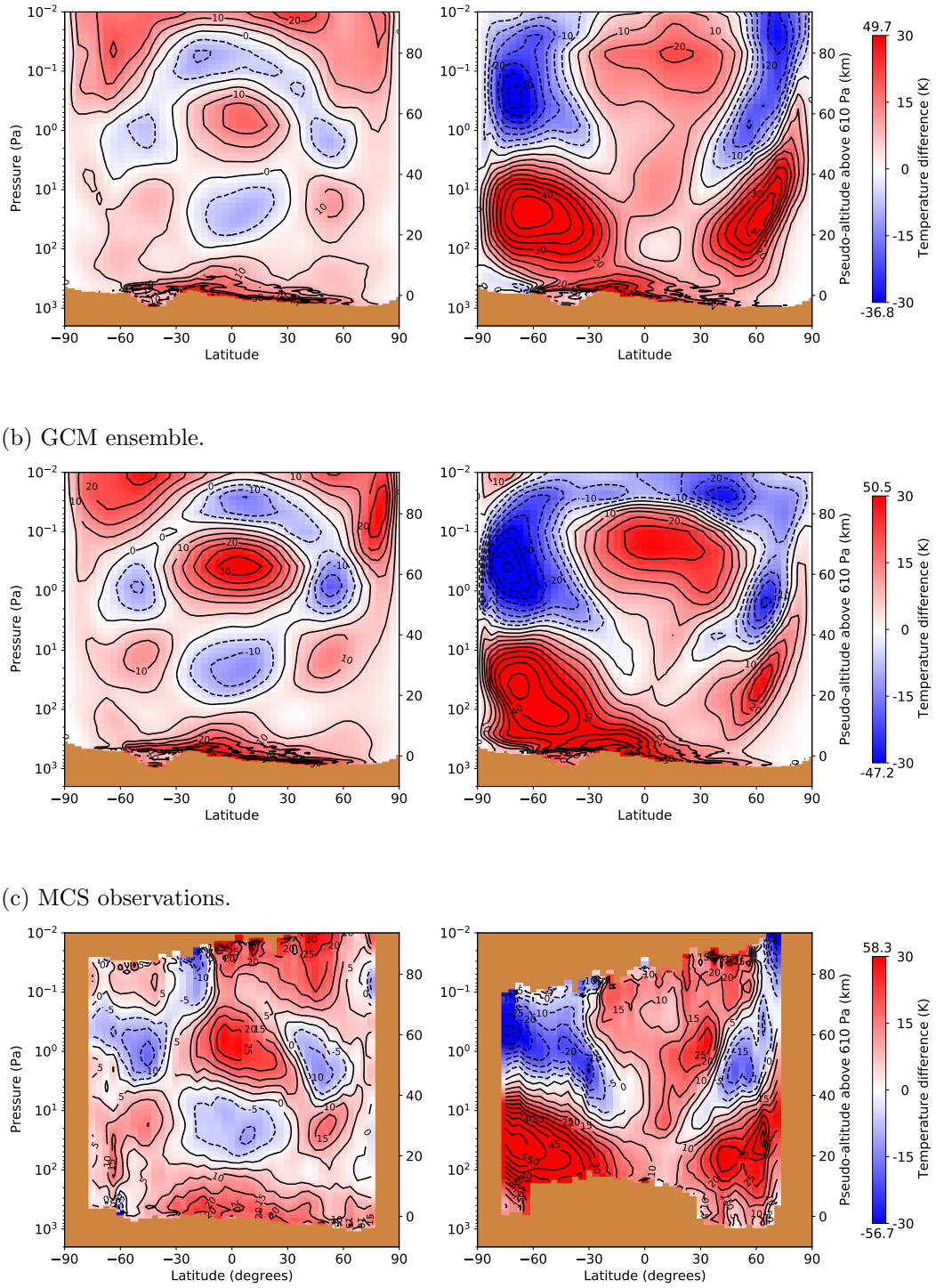


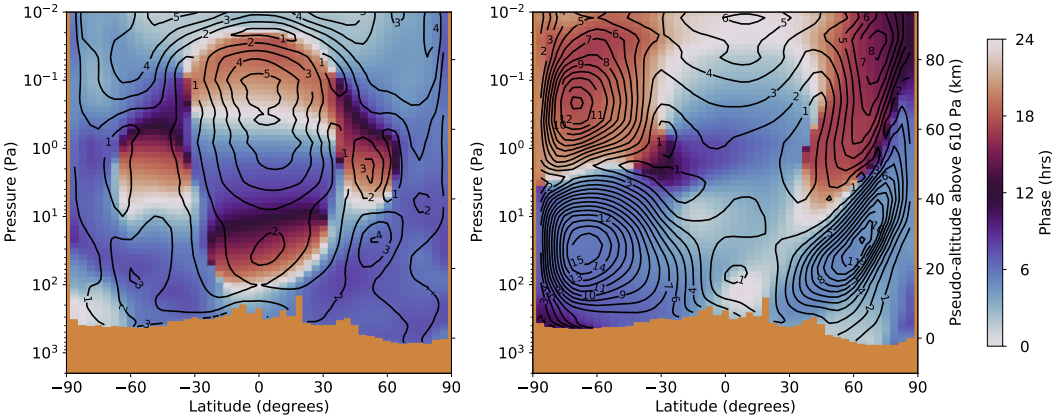
Figure 18. Difference between 3 PM and 3 AM temperatures (local mean solar time) for (a) TuTD-CuD analysis; (b) GCM ensemble; and (c) MCS observations (using observations between LMST 2-4 AM/PM). The left column shows sols 380–384, before the storm, and the right column shows sols 416–420, at the peak of the storm. Brown is missing data (either below the surface, or not observed). Numbers above and below the colour bars indicate the minimum and maximum values for sols 416–420. Line contours are at 5 K intervals.

We can compare diurnal variations in our analysis against independent MCS observations by measuring the temperature differences between 3 PM and 3 AM. This comparison can only be made between those two local times, but it is one aspect of the diurnal cycle that can be compared directly with observations. This is shown in Fig. 18. Both the analysis and the GCM ensemble reproduce most basic aspects of the MCS observations, both before the storm and at its peak. Differences between 3 PM and 3 AM temperatures are generally smaller in the analysis compared with the MCS observations. The GCM ensemble matches the MCS observations reasonably well before the storm, while at the peak of the storm it tends to overestimate differences in the upper atmosphere and underestimate them in the lower atmosphere. The general position of the peaks and troughs in the MCS observations are reproduced well by both the analysis and the GCM ensemble.

At the peak of the storm the quadrupole structure disappears and is replaced at mid-to-high latitudes by a strong positive anomaly in the lower atmosphere and a strong negative anomaly in the middle atmosphere. Near the equator the temperature difference is small in the lower atmosphere and increases significantly in the middle atmosphere. In most areas the analysis is closer than the GCM ensemble to the MCS observations. In the northern hemisphere the analysis better matches the anomaly magnitudes and vertical structure. In the equatorial region the positive anomaly in the GCM ensemble is more compact than in the MCS observations. Because the MCS observations don't reach the ground, we cannot tell whether the analysis or GCM ensemble better matches reality near the ground in the southern hemisphere.

Figures 19 and 20 show the amplitude and phase of the westward-propagating Sun-synchronous migrating diurnal and semi-diurnal tides, respectively, before and at the peak of the storm. We interpolated the analysis and the GCM ensemble temperature fields to a fixed set of pressure coordinates. At each pressure and latitude point this gave temperature as a function of longitude and absolute time (over the 4-sol periods we have used throughout), which we then decomposed into its Fourier modes to give a two-dimensional spectrum as a function of longitudinal wavenumber and frequency. We then extracted the amplitude and phase of the period-(1 sol), wavenumber-(-1) mode, which is the westward-propagating Sun-synchronous migrating diurnal tide S_1; and the period-(0.5 sol), wavenumber-(-2) mode, which is the equivalent semi-diurnal tide mode S_2. The phase is expressed in terms of hours (0–24 for the diurnal tide, 0–12 for the semi-diurnal tide).

(a) TuTD-CuD analysis.



(b) GCM ensemble.

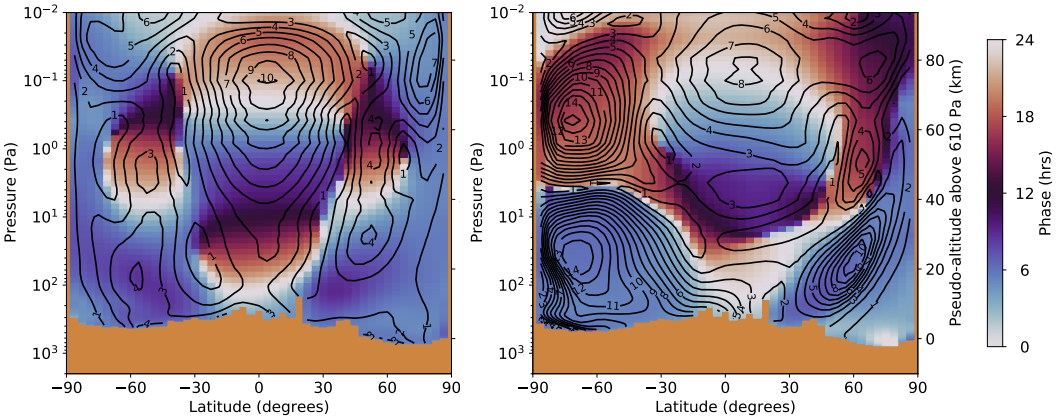


Figure 19. Amplitude and phase of the sun-synchronous migrating diurnal tide S_1. (a) TuTD-CuD analysis; (b) GCM ensemble. The left column is for sols 380–384, and the right column is for sols 416–420. Filled colour contours show the phase (scaled to local true solar time between 0–24 hours), and line contours show the amplitude (spaced at 1 K intervals).

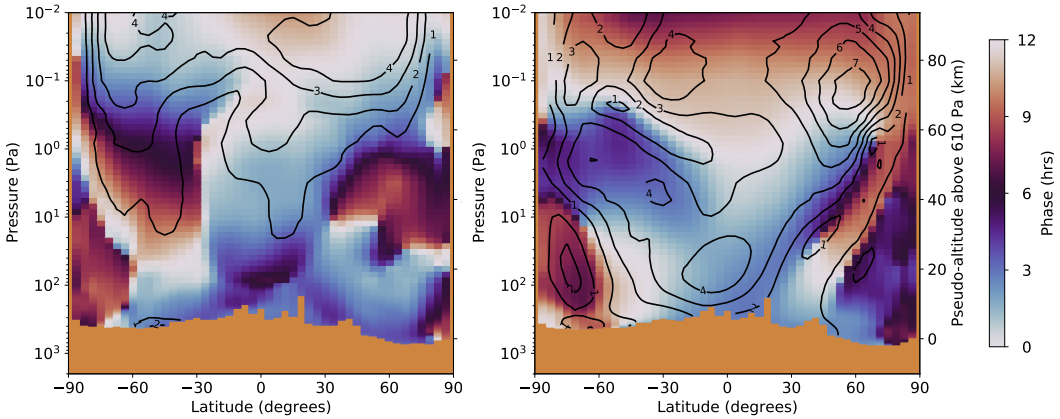
Before the storm the the spatial patterns of the diurnal tide (Fig. 19) are similar to those in Fig. 18. The phase data exhibit a quadrupole structure before the storm, and at the peak of the storm the vast majority of the power in the diurnal tide is at mid- and high latitudes, with a weaker diurnal tide at the equator. This is consistent with what we showed in Fig. 17. Before the storm, the diurnal tide has a maximum amplitude near the equator of about 5 K in the analysis and 10 K in the GCM ensemble (note that Fig. 18 shows the difference between two local times separated by 12 hours, not a wave amplitude). This difference suggests that the amplitude of the diurnal tide is overestimated in the GCM ensemble.

At the peak of the storm the diurnal tide amplitudes in the GCM ensemble and analysis are closer, although near the equator the analysis is again weaker than the GCM ensemble, particularly at high altitudes. The analysis peaks at about 6 K right at the top of the domain, while the GCM ensemble peaks at 8K around 0.1 Pa. At mid and high latitudes, however, the analysis and the GCM ensemble agree within about 1 K. One difference between the analysis and the GCM ensemble at the peak of the storm is the phase of the diurnal tide near the equator below 1 Pa. In the analysis this phase is 3-9 hours, whereas in the GCM ensemble it is 9–24 hours, a significant phase shift.

Before the storm the amplitude of the semi-diurnal tide (Fig. 20) is small in the lower atmosphere, increasing to about 4 K in the upper parts of the domain. The amplitude is largely independent of latitude in the analysis, but there is a slightly larger amplitude and one peak in each hemisphere around 50° N/S in the GCM ensemble. We may compare these figures before the storm directly with equivalent figures in Kleinböhl et al. (2013). Before the storm the general pattern is similar except for the phase near the equator. The phase generally increases downwards, starting around 0 hours at the top of the domain, and reaching 6 hours around 1 Pa. In Kleinböhl et al. (2013) this is only a weak function of latitude, but in our analysis (Fig. 20a, left) the phase of the semi-diurnal tide is 0–3 hours throughout most of the column, while in the GCM ensemble it displays the full range of phases, but goes though 6 hours much higher in the atmosphere than at higher latitudes.

At the peak of the storm the amplitude of the semi-diurnal tide increases at all altitude and latitudes, particularly near the surface, which increases in the analysis from less than 1 K to about 4 K. In the upper part of the domain the amplitude increases from

(a) TuTD-CuD analysis.



(b) GCM ensemble.

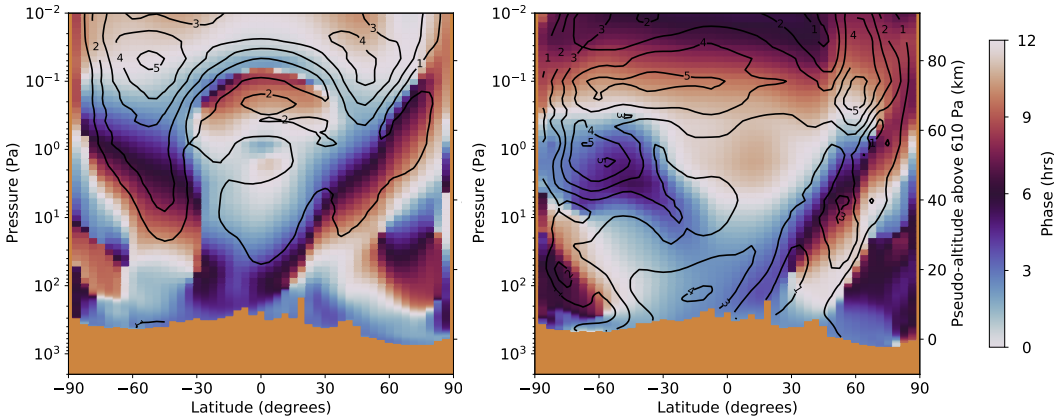


Figure 20. Same as Fig. 19, but showing the amplitude and phase of the sun-synchronous migrating semi-diurnal tide S_2. In this case the filled colour contours show the phase scaled to local true solar time between 0–12 hours.

4 K to about 7 K at 0.1 Pa around 60° N. Again the amplitudes in the analysis are generally smaller than in the GCM ensemble, but not by much.

5.6 Verification against Independent Surface Pressure Observations from Curiosity-REMS

The final verification of our analysis is to compare surface pressures with high time cadence measurements taken by the Mars Science Laboratory (MSL) Curiosity rover Rover Environmental Monitoring Station (REMS). We interpolated the analysis surface pressure field from the GCM grid to Curiosity's longitude and latitude before the storm (sols 380–384) and at the peak of the storm (sols 416–420). We then corrected the surface pressure to account for the difference in surface elevation between the GCM grid and Curiosity. We used Mars Orbital Laser Altimeter (MOLA) data with 32 points per degree (D. E. Smith et al., 2001) to obtain the altitude of the Curiosity rover, interpolating the MOLA surface elevation to Curiosity's location at each time it made a pressure measurement during these two periods. We corrected the analysis surface pressure assuming hydrostatic balance and a pressure scale height based on the temperature at 1 km pseudo-altitude (Spiga et al., 2007). The correction factor is

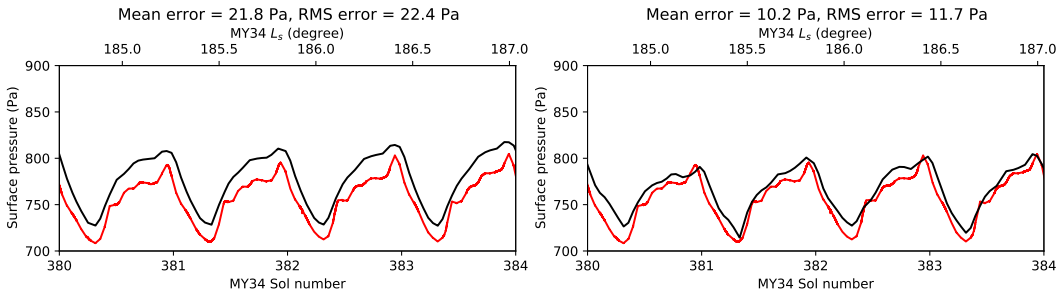
$$p_{\text{corrected}} = p_{\text{original}} \exp\left(-\frac{z_{\text{Curiosity}} - z_{GCM}}{H}\right) \tag{9}$$

where $H = RT/g$ is the scale height. Because Curiosity is deep within Gale Crater, which is not resolved by the LMD Mars GCM, the typical change in elevation from the GCM grid to Curiosity's location was an enormous -2,720 m, typically +160 Pa.

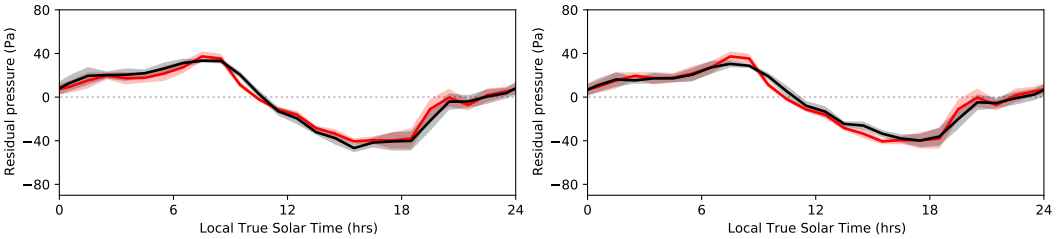
There are uncertainties associated with the finite resolution of the MOLA grid (about 2 km), and we also assume that errors due to using Curiosity's location in the rover landing frame of reference (which is inclined to the direction of gravity by about 3°) are small. However, our estimated uncertainty due to these assumptions is smaller than the uncertainty in the pressure measurements themselves (5.8 Pa).

Figure 21 shows surface pressure time series and diurnal cycles at the Curiosity rover location over two 4-sol periods before and at the peak of the storm, for the GCM ensemble and our best-case assimilation (TuTD-CuD). Both the analysis and the GCM ensemble generally fit the measured Curiosity observations well. For the analysis, the match

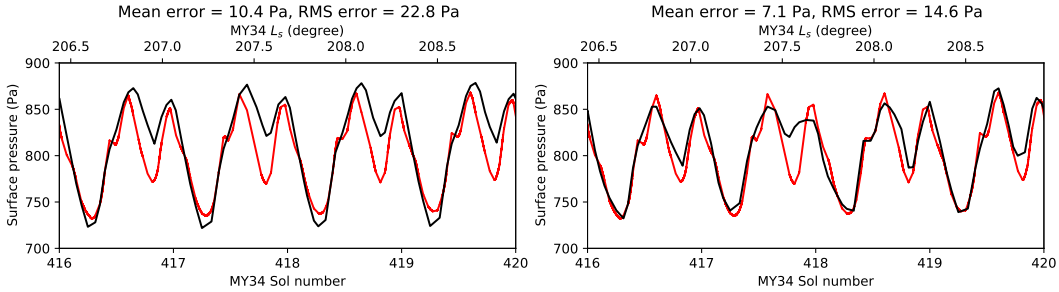
(a) Time series for sols 380–384.



(b) Diurnal cycle for sols 380–384.



(c) Time series for sols 416–420.



(d) Diurnal cycle for sols 416–420.

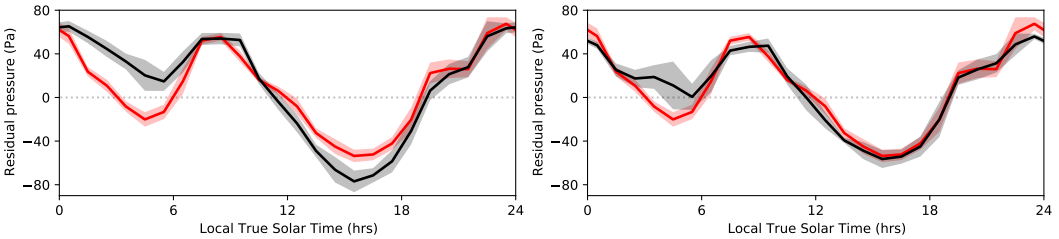


Figure 21. Surface pressure time series and diurnal cycle at the Curiosity rover location. The left column shows the GCM ensemble, and the right column shows assimilation TuTD-CuD. In each case the red line shows the Curiosity data, and the black line shows the analysis or GCM ensemble. (a-b) are before the storm, sols 380–384, and (c-d) are at the peak of the storm, sols 416–420. (a) and (c) show surface pressure time series, corrected for the difference in surface elevation between the GCM grid and the rover location on the surface, but otherwise unmodified. (b) and (d) show the same data binned into 30-minute local true solar time bins, where the mean in each bin is the thick line, and the shaded region shows +/- the standard deviation. Before binning the data the diurnal mean pressure was removed from the time series using a 1-sol moving average. The mean and RMS surface pressure error (analysis/GCM – Curiosity) are listed above each time series.

to the time series at the peak of the storm is particularly good, falling within the Curiosity measurement uncertainty. Given that we are down-sampling a climate resolution model (which does not resolve Gale Crater) with grid boxes approximately 300 × 200 km to a single location within the crater, and then correcting the surface pressure for a difference in surface elevation close to 3 km, a mean error close to 10 Pa is remarkable.

The surface pressure rises during the storm by about 50 Pa, and the diurnal mode becomes stronger, with the peak-to-peak diurnal variation increasing from 80 Pa to 120 Pa. The analysis is closer to the Curiosity observations than the GCM ensemble, reflecting the additional constraints on surface pressure from the assimilated temperatures. The analysis is better both in its absolute match to the Curiosity time series, and in how well the diurnal cycle is reproduced, both before and at the peak of the storm. The GCM ensemble overestimates the absolute pressure, and overestimates the peak-to-peak diurnal variability, compared with the analysis and the Curiosity measurements. This is reflected in the mean and RMS errors, which are significantly smaller for the analysis than for the GCM ensemble: about 50% before the storm and about 70% at the peak of the storm. Only between 2–7 AM does the analysis poorly track the diurnal surface pressure cycle at the Curiosity location at the peak of the storm, where it overestimates the observations. The same is seen in the GCM ensemble, although the error is larger. This is most likely due to mesoscale meteorology specific to Gale Crater, which is not accounted for in the model (see e.g. Rafkin et al. (2017) for a review). Our analysis' good agreement with observations, despite the difficulties associated with interpolating into a crater, is encouraging for possible future assimilation of surface pressure observations.

As the analysis verifies well against Curiosity observations, Fig. 22 shows surface pressure diurnal cycles at several other locations; we may reasonably believe the analysis reproduces the surface pressure diurnal cycle there. These points of interest are: the final location of the Opportunity rover (which succumbed to dust loading during the study period); the landing sites for Perseverance and Insight, both currently operational; the Viking 2 landing site, chosen as it is one of the locations used to tune the pressure cycle in the LMD Mars GCM; Acidalia, where flushing dust storms often originate, and within the northern hemisphere baroclinic zone; Hellas, where global dust events have originated (Mulholland et al., 2013), and the location of the highest surface pressures on Mars; and Tharsis, the location of the lowest surface pressures on Mars.

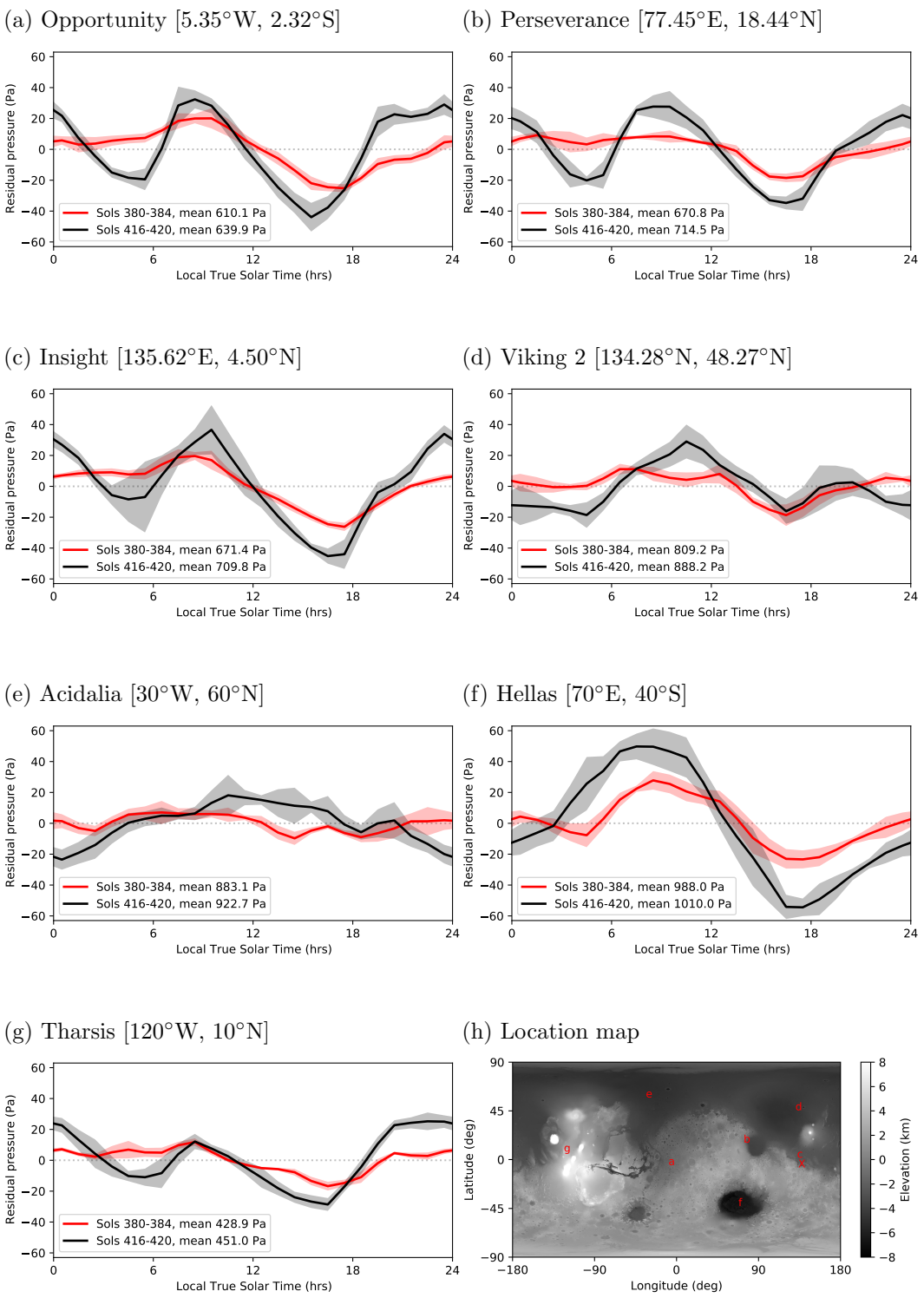


Figure 22. Surface pressure diurnal cycles at several locations during assimilation TuTD-CuD. Red lines are from the period before the storm (sols 380–384) and black lines are from the peak of the storm (sols 416–420). The diurnal cycle plots are otherwise the same as in Fig. 21. The mean pressure over each 4-sol period is given in each legend. The locations are (a) Opportunity final location; (b) Perseverance landing site; (c) Insight landing site; (d) Viking 2 landing site (these locations were all obtained using NASA GISS' Mars24 software); (e) Acidalia; (f) Hellas; and (g) Tharsis. (h) shows these seven locations with an additional X at the Curiosity landing site, for reference [137.44°E, 4.59°S]; topography is from MOLA (D. E. Smith et al., 2001).

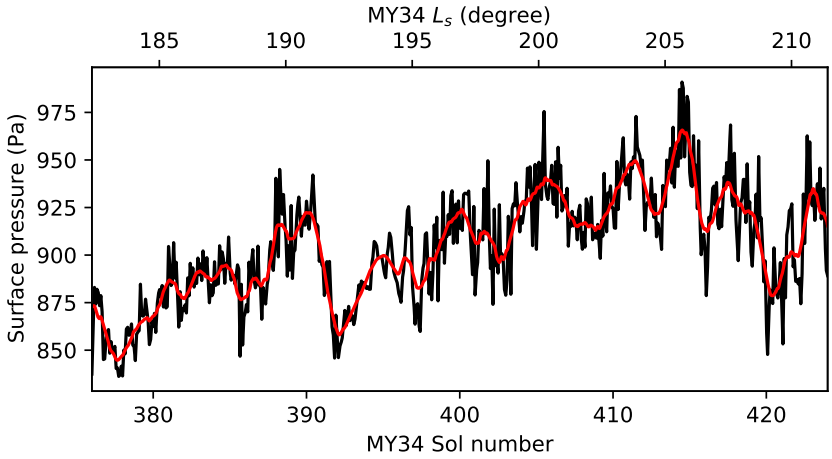
In all cases the surface pressure increases from before the storm to the peak of the storm, ranging from about +20 Pa in Hellas to about +80 Pa at the Viking 2 landing site. The magnitude of the diurnal mode also increases at the peak of the storm, and at most locations the magnitude of the semi-diurnal mode also increases. The day-to-day variability (quantified by the width of the shaded region accompanying each line) is noticeably higher in the analysis than in the GCM ensemble (an equivalent figure to Fig. 22 for the GCM ensemble is included as Supporting Fig. S3).

The mid-afternoon low at each location is typically weaker in the analysis than in the GCM ensemble, as it was at the Curiosity location, and the mean pressure difference between the analysis and the GCM ensemble is typically 5–15 Pa. One exception is Acidalia (Fig. 22e), where the analysis pressure is about 30 Pa higher than the GCM ensemble, and the diurnal cycle is qualitatively different. The diurnal mode is relatively weak compared with other locations, about 20 Pa peak-to-peak before the storm, and 40 Pa at the peak of the storm, and the phase of the diurnal pressure mode is shifted in the analysis by about 3–6 hours relative to the GCM ensemble. Figure 23 shows the pressure time series in both cases for the whole assimilation period. In the GCM ensemble the baroclinic waves are clear, with a period of about 10 sols. In the analysis, however, while some signatures of the baroclinic wave remain (peaks at sols 390 and 405, for example), there are shorter-period oscillations that do not appear in the GCM ensemble. Variability on diurnal timescales is also noisier, better reflecting real conditions; we saw this in the Curiosity dataset, where the day-to-day variability in the analysis diurnal cycle (Fig. 21d, right) is similar to the day-to-day variability in the Curiosity observations.

6 Summary and Conclusion

In this paper we have assimilated observations from ACS-TIRVIM, the thermal infrared spectrometer on board ExoMars Trace Gas Orbiter, into the LMD Mars Global Climate Model during the onset and peak of the MY34 global dust storm. We assimilated both temperature profiles and dust column optical depth measurements using the LETKF in three configurations, with our nominal case TuTD-CuD assimilating temperature profiles to update temperature and dust profiles, followed by column dust optical depths to update the total dust column abundance. We ran an ensemble of GCM simulations alongside the assimilations, constrained by MY34 MCS dust observations. Our

(a) TuTD-CuD analysis.



(b) GCM ensemble.

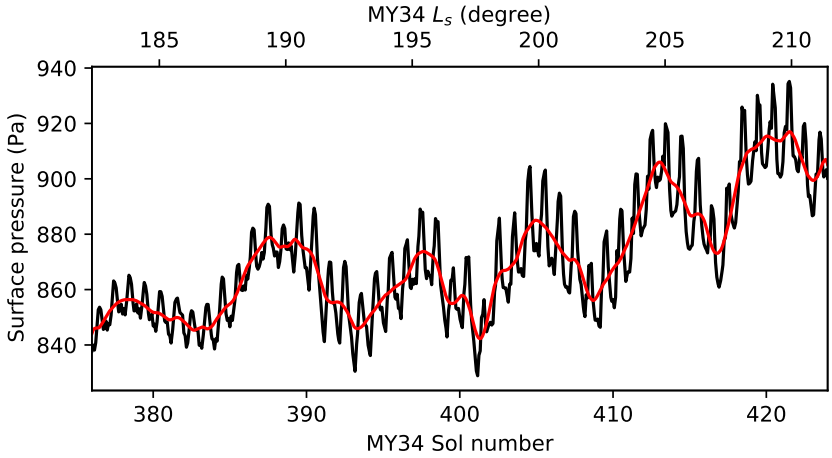


Figure 23. Surface pressure time series in Acidalia [30°W, 60°N] throughout the full assimilation period for (a) analysis TuTD-CuD and (b) the GCM ensemble. Black shows the instantaneous surface pressure (every two hours) and red shows a running diurnal mean. Note the different y-axis scales.

other improvement over previous work was to compare forecast with observations using the retrieval averaging kernels. This allowed a like-by-like comparison to be made between forecast and observation, and by application of a filter to the sum over the averaging kernels at each vertical level, we ensured that information assimilated from the retrievals was strongly weighted towards information from the real atmospheric state rather than the retrieval prior. This is important for assimilating such temperature profiles derived from nadir-viewing geometry, which have rather coarse vertical resolution.

Our nominal assimilation verified well against in-sample observations from TIRVIM, and was at least as good as the GCM ensemble when verified against independent MCS observations. This was encouraging because the GCM ensemble was itself constrained by dust opacities measured by MCS. There was a significant reduction in RMS error when verifying temperatures against MCS once we assimilated column dust optical depth observations.

The atmospheric temperature structure at 30 Pa followed the MCS observations closely in our TuTD-CuD case, and was noticeably closer to the MCS values than the GCM ensemble (Fig. 11 c-e). While the 3PM analysis dust density-scaled opacity did not reach the maximum observed by MCS at the peak of the storm, there was particularly good correspondence with MCS at 3AM. The analysis reproduced detached dust layers, while the GCM ensemble did not, and the latter also overestimated the dust DSO at the storm's peak (Fig. 14). The assimilation retrieved the (unobserved) atmospheric wind structure before and during the storm, which showed that the meridional overturning circulation significantly strengthened during the storm (Fig. 15), as well as the development of a 125 m s^{-1} asymmetry in the midlatitude jets between the two hemispheres (Fig. 16).

Assimilating observations over a range of local times gives us confidence in our results related to the diurnal cycle. The diurnal tide strengthened considerably at midlatitudes during the global dust storm, at the expense of the diurnal tide near the equator (Fig. 19), and the semi-diurnal tide increased in strength everywhere, particularly in the lower atmosphere (Figs 17c, 20). The diurnal cycle was also verified against independent Curiosity-REMS surface pressure measurements. Despite the large difference in surface elevation between the GCM topography and the rover, the analysis surface

pressures at the peak of the storm accurately reproduced the diurnal cycle measured by Curiosity (Fig. 21).

Assimilation during the onset of a Mars global dust storm is a particularly hard stress test of our assimilation scheme, so it is encouraging that we are able to reproduce many features of independent observational datasets. But one shortcoming of the assimilation has been made particularly clear when the total amount of dust in the atmosphere changes quickly over time. There is a lag in changes to the global amount of dust (and hence the onset of the global dust storm) by 6–8 sols (Fig. 12c). This is also present in temperature, but it is only 1–2 sols in that case. We expect this is because not all model grid points are updated during each assimilation cycle. If only a small fraction of grid points are updated at any one time, trends in the total amount of dust and heat in the analysis will lag behind such trends in the real atmosphere, because in reality the dust and heat content change at both unobserved and observed locations at the same time. This is a fundamental limitation in the observations, particularly when assimilating data from a single polar-orbiting satellite where the number of model grid columns updated in a single 2-hour assimilation cycle is typically 20–30% of the total.

This could be overcome in a number of ways. First, by simultaneously assimilating more observations at different locations, such as by assimilating MCS in addition to ACS observations. Second, by increasing the length of the assimilation window to include more observations, but this increases the forecast time. Third, by assimilating single observations multiple times, as in Analysis Correction (Lorenc et al., 1991). Fourth, one could use a longer assimilation window for dust than for temperature. Finally, one could update non-observed locations in some ad hoc way, for example by assuming that dust lingers in place over some timescale, or by assuming some spatial distribution of dust linking observed locations. This is already done in a sophisticated way for the GCM dust scenarios, which are complete maps but which use incomplete observations with gaps filled in by kriging (Montabone et al., 2015). Only the first of these options avoids some ad hoc intervention or a substantial change to the assimilation method, so the next step will be to focus on joint assimilation of multiple instruments' observations.

Acknowledgments

ExoMars is a space mission of ESA and Roscosmos. The ACS experiment is led by IKI, the Space Research Institute in Moscow, Russia, assisted by LATMOS in France. This work, exploiting ACS/TIRVIM data, acknowledges funding by CNES. The science operations of ACS are funded by Roscosmos and ESA. The ACS/TIRVIM team at IKI acknowledges the subsidy of the Ministry of Science and Higher Education of Russia.

RMBY acknowledges funding from UAE University grants G00003322 and G00003407. Supercomputing resources were provided by UAE University High Performance Computing, with technical support from Anil Thomas and Asma Alneyadi, and at LMD by the IPSL mesocentre. The authors thank Luca Montabone for access to processed versions of Mars Climate Sounder temperature and dust observations, and Thomas Navarro and Claus Gebhardt for useful discussions.

Data Availability Statement

The processed TIRVIM data and retrievals used in this paper are available in NetCDF format on the Institut Pierre Simon Laplace (IPSL) data server, see: https://doi.org/10.14768/ab765eba-0c1d-47b6-97d6-6390c63f0197. The authors acknowledge Sandrine Guerlet and the ACS/TGO team for supplying the data and the data center ESPRI/IPSL for their help in accessing the data. The observations use version 7 of the retrieval algorithm, and version 4 of the IKI calibration algorithm. Mars Climate Sounder observations are available on the NASA Planetary Data System Atmospheres node at https://pds-atmospheres.nmsu.edu/data_and_services/atmospheres_data/MARS/mars_reconnaissance_orbiter.html. Mars Science Laboratory Curiosity rover Rover Environmental Monitoring Station observations are available on the NASA Planetary Data System Atmospheres node at https://atmos.nmsu.edu/data_and_services/atmospheres_data/MARS/curiosity/rems.html. MOLA topography data are available on the NASA Planetary Data System Geosciences node at https://pds-geosciences.wustl.edu/missions/mgs/megdr.html. The LMD Mars GCM is available from http://www-mars.lmd.jussieu.fr/; we used GCM subversion revision r2533. The assimilation uses letkfmars git revision f28935f6472a280394a98ac64f99213d1cbb2e01. Mars24 is available from NASA GISS at https://www.giss.nasa.gov/tools/mars24/. The data produced by our assimilation will be deposited after review using Zenodo or an equivalent repository.

References

Ahn, S., & Fessler, J. A. (2003). *Standard Errors of Mean, Variance, and Standard Deviation Estimators* (Technical Report 413 No. 48109-2122). Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI.

Amiri, S., Brain, D., Sharaf, O., Withnell, P., McGrath, M., Alloghani, M., ... Yousuf, M. (2022). The Emirates Mars Mission. *Space Science Reviews*, *218*, 4. doi: 10.1007/s11214-021-00868-x

Capderou, M., & Forget, F. (2004). Optimal orbits for Mars atmosphere remote sensing. *Planetary and Space Science*, *52*, 789–798. doi: 10.1016/j.pss.2004.03.006

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. doi: 10.1007/s10236-003-0036-9

Forget, F., Hourdin, F., Fournier, R., Hourdin, C., Talagrand, O., Collins, M., ... Huot, J.-P. (1999). Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research*, *104*(E10), 24155–24175. doi: 10.1029/1999JE001025

Giuranna, M., Wolkenberg, P., Grassi, D., Aronica, A., Aoki, S., Scaccabarozzi, D., ... Formisano, V. (2021). The current weather and climate of Mars: 12 years of atmospheric monitoring by the Planetary Fourier Spectrometer on Mars Express. *Icarus*, *353*, 113406. doi: 10.1016/j.icarus.2019.113406

Greybush, S. J., Wilson, R. J., Hoffman, R. N., Hoffman, M. J., Miyoshi, T., Ide, K., ... Kalnay, E. (2012). Ensemble Kalman filter data assimilation of Thermal Emission Spectrometer temperature retrievals into a Mars GCM. *Journal of Geophysical Research*, *117*(11), E11008. doi: 10.1029/2012JE004097

Guerlet, S., Ignatiev, N., Forget, F., Fouchet, T., Vlasov, P., Bergeron, G., ... Korablev, O. (2022). Thermal Structure and Aerosols in Mars' Atmosphere From TIRVIM/ACS Onboard the ExoMars Trace Gas Orbiter: Validation of the Retrieval Algorithm. *Journal of Geophysical Research - Planets*, *127*, e2021JE007062. doi: 10.1029/2021JE007062

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, *129*(11), 2776–2790. doi: 10.1175/

1520-0493(2001)129⟨2776:DDFOBE⟩2.0.CO;2

Heavens, N. G., Richardson, M. I., Kleinböhl, A., Kass, D. M., McCleese, D. J., Abdou, W., ... Wolkenberg, P. M. (2011). Vertical distribution of dust in the Martian atmosphere during northern spring and summer: High-altitude tropical dust maximum at northern summer solstice. *Journal of Geophysical Research*, *116*(1), E01007. doi: 10.1029/2010JE003692

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J. M., Nicolas, J., ... Dee, D. (2019). Global reanalysis: Goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, *159*, 17–24. doi: 10.21957/vf291hehd7

Hoffman, M. J., Greybush, S. J., Wilson, R. J., Gyarmati, G., Hoffman, R. N., Kalnay, E., ... Szunyogh, I. (2010). An Ensemble Kalman Filter Data Assimilation System for the Martian Atmosphere: Implementation and Simulation Experiments. *Icarus*, *209*(2), 470–481. doi: 10.1016/j.icarus.2010.03.034

Hoffman, R. N. (2011). A retrieval strategy for interactive ensemble data assimilation. *arXiv*, 1009.1561.

Holmes, J. A., Lewis, S. R., Patel, M. R., & Lefèvre, F. (2018). A reanalysis of ozone on Mars from assimilation of SPICAM observations. *Icarus*, *302*, 308–318. doi: 10.1016/j.icarus.2017.11.026

Holmes, J. A., Lewis, S. R., Patel, M. R., & Smith, M. D. (2019). Global analysis and forecasts of carbon monoxide on Mars. *Icarus*, *328*, 232–245. doi: 10.1016/j.icarus.2019.03.016

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, *230*(1-2), 112–126. doi: 10.1016/j.physd.2006.11.008

Kahre, M. A., Murphy, J. R., Newman, C. E., Wilson, R. J., Cantor, B. A., Lemmon, M. T., & Wolff, M. J. (2017). The Mars Dust Cycle. In *The Atmosphere and Climate of Mars* (pp. 295–337). Cambridge University Press.

Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press.

Kass, D. M., Schofield, J. T., Kleinböhl, A., McCleese, D. J., Heavens, N. G., Shirley, J. H., & Steele, L. J. (2019). Mars Climate Sounder Observation of Mars' 2018 Global Dust Storm. *Geophysical Research Letters*, *46*, e2019GL083931. doi: 10.1029/2019GL083931

Kleinböhl, A., Schofield, J. T., Kass, D. M., Abdou, W. A., Backus, C. R., Sen, B., ... McCleese, D. J. (2009). Mars Climate Sounder limb profile retrieval of atmospheric temperature, pressure, and dust and water ice opacity. *Journal of Geophysical Research*, *114*(10), E10006. doi: 10.1029/2009JE003358

Kleinböhl, A., Spiga, A., Kass, D. M., Shirley, J. H., Millour, E., Montabone, L., & Forget, F. (2020). Diurnal Variations of Dust During the 2018 Global Dust Storm Observed by the Mars Climate Sounder. *Journal of Geophysical Research - Planets*, *125*, e2019JE006115. doi: 10.1029/2019JE006115

Kleinböhl, A., Wilson, R. J., Kass, D., Schofield, J. T., & McCleese, D. J. (2013). The semidiurnal tide in the middle atmosphere of Mars. *Geophysical Research Letters*, *40*, 1952–1959. doi: 10.1002/grl.50497

Korablev, O., Montmessin, F., Trokhimovskiy, A., Fedorova, A. A., Shakun, A. V., Grigoriev, A. V., ... Zorzano, M. P. (2018). The Atmospheric Chemistry Suite (ACS) of Three Spectrometers for the ExoMars 2016 Trace Gas Orbiter. *Space Science Reviews*, *214*(1), 7. doi: 10.1007/s11214-017-0437-6

Lewis, S. R., & Barker, P. R. (2005). Atmospheric tides in a Mars general circulation model with data assimilation. *Advances in Space Research*, *36*(11), 2162–2168. doi: 10.1016/j.asr.2005.05.122

Lewis, S. R., & Read, P. L. (1995). An operational data assimilation scheme for the Martian atmosphere. *Advances in Space Research*, *16*(6), 9–13. doi: 10.1016/0273-1177(95)00244-9

Lewis, S. R., Read, P. L., Conrath, B. J., Pearl, J. C., & Smith, M. D. (2007). Assimilation of thermal emission spectrometer atmospheric data during the Mars Global Surveyor aerobraking period. *Icarus*, *192*(August), 327–347. doi: 10.1016/j.icarus.2007.08.009

Lorenc, A. C., Bell, R. S., & Macpherson, B. (1991). The Meteorological Office analysis correction data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, *117*(497), 59–89. doi: 10.1002/qj.49711749704

Madeleine, J.-B., Forget, F., Millour, E., Montabone, L., & Wolff, M. J. (2011). Revisiting the radiative impact of dust on Mars using the LMD Global Climate Model. *Journal of Geophysical Research*, *116*(11), E11010. doi: 10.1029/2011JE003855

Montabone, L., Forget, F., Millour, E., Wilson, R. J., Lewis, S. R., Cantor, B., ...

Wolff, M. J. (2015). Eight-year climatology of dust optical depth on Mars. *Icarus*, *251*, 65–95. doi: 10.1016/j.icarus.2014.12.034

Montabone, L., Marsh, K., Lewis, S. R., Read, P. L., Smith, M. D., Holmes, J., ... Pamment, A. (2014). The Mars Analysis Correction Data Assimilation (MACDA) Dataset V1.0. *Geoscience Data Journal*, *1*(2), 129–139. doi: 10.1002/gdj3.13

Montabone, L., Spiga, A., Kass, D. M., Kleinböhl, A., Forget, F., & Millour, E. (2020). Martian Year 34 Column Dust Climatology from Mars Climate Sounder Observations: Reconstructed Maps and Model Simulations. *Journal of Geophysical Research - Planets*, *125*, e2019JE006111. doi: 10.1029/2019JE006111

Mulholland, D. P., Read, P. L., & Lewis, S. R. (2013). Simulating the interannual variability of major dust storms on Mars using variable lifting thresholds. *Icarus*, *223*(1), 344–358. doi: 10.1016/j.icarus.2012.12.003

Navarro, T., Forget, F., Millour, E., & Greybush, S. J. (2014). Detection of detached dust layers in the Martian atmosphere from their thermal signature using assimilation. *Geophysical Research Letters*, *41*(19), 6620–6626. doi: 10.1002/2014GL061377

Navarro, T., Forget, F., Millour, E., Greybush, S. J., Kalnay, E., & Miyoshi, T. (2017). The challenge of atmospheric data assimilation on Mars. *Earth and Space Science*, *4*, 690–722. doi: 10.1002/2017EA000274

Pottier, A., Forget, F., Montmessin, F., Navarro, T., Spiga, A., Millour, E., ... Madeleine, J.-B. (2017). Unraveling the martian water cycle with high-resolution global climate simulations. *Icarus*, *291*, 82–106. doi: 10.1016/j.icarus.2017.02.016

Rafkin, S. C. R., Spiga, A., & Michaels, T. I. (2017). Mesoscale Meteorology. In *The Atmosphere and Climate of Mars* (pp. 203–228). Cambridge University Press.

Rodgers, C. D., & Connor, B. J. (2003). Intercomparison of remote sounding instruments. *Journal of Geophysical Research*, *108*(D3), 4116. doi: 10.1029/2002JD002299

Ruan, T., Young, R. M. B., Lewis, S. R., Montabone, L., Valeanu, A., & Read, P. L. (2021). Assimilation of Both Column- and Layer-Integrated Dust Opacity Observations in the Martian Atmosphere. *Earth and Space Science*, *8*,

1332 e2021EA001869. doi: 10.1029/2021EA001869

1333 Smith, D. E., Zuber, M. T., Frey, H. V., Garvin, J. B., Head, J. W., Muhleman,
1334 D. O., ... Sun, X. (2001). Mars Orbiter Laser Altimeter: Experiment sum-
1335 mary after the first year of global mapping of Mars. *Journal of Geophysical*
1336 *Research*, *106*(E10), 23689–23722. doi: 10.1029/2000JE001364

1337 Smith, M. D., Bougher, S. W., Encrenaz, T., Forget, F., & Kleinböhl, A. (2017).
1338 Thermal Structure and Composition. In *The Atmosphere and Climate of Mars*
1339 (pp. 42–75). Cambridge University Press.

1340 Spiga, A., Forget, F., Dolla, B., Vinatier, S., Melchiorri, R., Drossart, P., ...
1341 Gondet, B. (2007). Remote sensing of surface pressure on Mars with the
1342 Mars Express/OMEGA spectrometer: 2. Meteorological maps. *Journal of*
1343 *Geophysical Research: Planets*, *112*(E8), E08S16. doi: 10.1029/2006JE002870

1344 Steele, L. J., Lewis, S. R., & Patel, M. R. (2014). The radiative impact of water ice
1345 clouds from a reanalysis of Mars Climate Sounder data. *Geophysical Research*
1346 *Letters*, *41*(13), 4471–4478. doi: 10.1002/2014GL060235

1347 Steele, L. J., Lewis, S. R., Patel, M. R., Montmessin, F., Forget, F., & Smith,
1348 M. D. (2014). The seasonal cycle of water vapour on Mars from assimi-
1349 lation of Thermal Emission Spectrometer data. *Icarus*, *237*, 97–115. doi:
1350 10.1016/j.icarus.2014.04.017