Icarus 317 (2019) 649–668

Contents lists available at ScienceDirect

Icarus

journal homepage: www.elsevier.com/locate/icarus

An initial assessment of the impact of postulated orbit-spin coupling on Mars dust storm variability in fully interactive dust simulations

C.E. Newman[a,*], C. Lee[b,a], M.A. Mischna[c], M.I. Richardson[a], J.H. Shirley[c]

[a] Aeolis Research, Pasadena, CA, United States
[b] University of Toronto, Toronto, Canada
[c] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States

ABSTRACT

A weak coupling between the rotational and orbital angular momenta of Mars has been postulated to produce a 'coupling term acceleration' (CTA) that accelerates the wind field and is asynchronous with the seasonal cycle of solar forcing (Shirley, 2017). This paper presents the first GCM simulations of a fully interactive dust cycle with the CTA included, enabling storm sizes, onset times and locations to be predicted. The inclusion of the CTA greatly augments interannual variability in the occurrence and timing of GDS, with the nature of the storm season strongly linked to the phasing and amplitude of the orbit-spin coupling. This dramatically improves the model's skill at predicting GDS and non-GDS Mars Years (MY) compared to a GCM without CTA forcing. The model is clearly wrong in only 4 out of 22 well-observed storm seasons and is able to capture the general onset time of most observed storms as well as some onset locations.

In years when the CTA forcing has large positive amplitudes around perihelion, GDS with onset near perihelion occur due to a net strengthening of the single-cell Hadley circulation at this time, while earlier (or later) GDS are likely produced by more localized constructive interference between the CTA and tidal/topographic flows at a time of peak forcing amplitudes. The latter may be more sensitive to errors in the assumed surface dust availability, which may explain why a late GDS observed in MY 10 is not predicted. Depletion of surface dust in source regions by GDS in prior years may have prevented a GDS from occurring in the real MY 17, when a large GDS is incorrectly predicted. Early GDS are observed but not predicted in two MYs (12 and 25) with large negative CTA forcing amplitudes around perihelion, which may be associated with a lack of water cycle coupling in these simulations. Other missing physical processes, imperfect dust parameterizations or parameter values, the assumption of unlimited surface dust availability, or the wrong CTA strength may account for other mismatches.

A GDS is predicted close to perihelion in the current storm season, MY34 (2018), with a smaller GDS predicted later next Mars year, MY35 (2020). The CTA forcing in MY 34 is very similar to that of MY 21, in which a GDS is correctly predicted by the model.

1. Introduction

Dust raised from the surface has a large impact on radiative transfer through the thin Martian atmosphere and thus on the atmosphere's thermal structure and hence its circulation, with increased dust loading resulting in a stronger global circulation overall. Most global modeling studies to date have focused on two primary mechanisms for raising dust from the surface of Mars (see extensive introductory discussion in Newman and Richardson, 2015, henceforth referred to as NR15). The first is associated with model-resolved winds (on synoptic scales, ~500 km) and is initiated when wind stresses exceed a threshold. The second is associated with vertical convective vortices (dust devils) that are assumed to dominate convective planetary boundary layer (PBL) dust lifting on microscales (less than about 5 km). The model-resolved wind stress lifting depends largely on wind speed and air density, while the dust devil lifting also depends on the surface-to-air temperature difference and PBL thickness (see Section 2.1).

Dust lifting may result in local, regional, or global storms, where local storms typically cover less than about 10^6 km^2 in area, regional storms cover areas larger than this but do not encircle the planet in longitude, while global dust storms (GDS) are planet-encircling in longitude and may cover most latitudes too. Regional storms may have more than one lifting center within a region, but GDS always consist of several lifting centers and often begin as a grouping of two or more regional storms, which may or may not combine depending on the global circulation's response to the additional dust loading. To date, all GDS have been observed to begin during the so-called 'storm season' between areocentric solar longitude (L_s) ~180° and 310°, i.e., between southern spring equinox and mid-summer. Curiously, while the general pattern of regional storms typically repeats from year to year, GDS have only been observed in roughly one year out of three, although no GDS had occurred for five Mars years as of mid-2018..

The general seasonal timing of GDS is likely due to the eccentricity of Mars's orbit, as well as the hemispheric topographic dichotomy. Perihelion (the time of closest approach to the sun, hence the peak in solar forcing) occurs at L_s ~ 251°, shortly before southern summer

* Corresponding author.
E-mail address: claire@aeolisresearch.com (C.E. Newman).

https://doi.org/10.1016/j.icarus.2018.07.023
Received 6 March 2018; Received in revised form 6 June 2018; Accepted 31 July 2018
Available online 17 September 2018

solstice (at $L_s = 270^\circ$). Around solstice, the zonal-mean circulation across much of Mars is dominated by a single overturning meridional Hadley cell, with its rising branch at summer mid-latitudes. Due to the timing of perihelion, and also due to the slope in planetwide topography which puts the southern hemisphere overall higher than in the north, the southern summer Hadley cell is more intense than its northern summer counterpart. Strong wind stresses are also associated with baroclinic wave activity around the polar cap edge, resulting in increased local and regional storm activity at high northern latitudes from $L_s \sim 0$–180°, and high southern latitudes from $L_s \sim 180$–360°. The role of the very large thermal tide in boosting both the influence of the tropical overturning circulation and in interaction with baroclinic wave activity is also notable for Mars.

The net result is that the strongest solar forcing and the strongest planetwide winds occur around in southern late spring / early summer, leading to a peak in dust devil and wind stress lifting, with the resulting increase in dust loading further boosting the seasonal asymmetry in circulation intensity. As the dust optical depth increases, the surface-to-air temperature contrast decreases and dust devil lifting tends to shut down, whereas wind stress lifting enjoys a strong positive feedback to much higher optical depths, due to the link between dust loading and circulation strength. Eventually, however, dust storms cease due to some as-yet-unknown combination of the exhaustion of surface dust source regions, the decoupling of surface winds from upper level flows at high opacities, and the changing circulation with season.

1.1. Understanding Martian dust storms using general circulation models

In the past few decades, strides have been made in understanding the patterns and timings of Martian dust storms by simulating them using general circulation models (GCMs). The earliest models prescribed some amount of dust lifting then followed the transport of radiatively active atmospheric dust by the circulation and sedimentation back to the surface (e.g. Murphy et al., 1990; Wilson and Hamilton, 1996). Subsequent models went a step further by parameterizing wind stress and dust devil dust lifting, as discussed above and in Section 2.1, enabling the pattern of dust lifting to be self-consistently predicted based on the GCM circulation, with feedbacks between the two activated. The actual lifting mechanisms may be considerably more complex (e.g. Kok et al., 2012), yet global models using such parameterizations have been successful in capturing the general character of the global dust cycle, such as spontaneously generating quasi-realistic storms and producing some degree of interannual variability (e.g. Newman et al., 2002a,b; Basu et al., 2004, 2006; Kahre et al., 2006; Newman and Richardson, 2015).

Although the general seasonal pattern of dust storms could be explained, however, several outstanding questions remained. For example, the early GDS observed to begin soon after $L_s \sim 180^\circ$ have long been hard to predict in GCMs with an active dust cycle, occurring as they do far before the time of peak solar insolation and circulation strength. Recent work suggests that radiatively active water ice clouds in early southern spring may be responsible for triggering major dust events at this time, although more study is required (e.g. Kahre et al., 2012; Lee et al., 2018).

But the largest outstanding question continues to be: what causes the huge interannual variability in observed GDS? In other words, what causes one storm season to contain only regional storms, while in another storms grow to become global? And why do some storms begin early in the dust storm season, while others begin much later? Several explanations have been proposed in recent years. Basu et al. (2004, 2006) demonstrated that sufficient intrinsic interannual variability exists in the pattern of surface wind stress to produce GDS in some years and not others, provided that the threshold wind stress required to produce dust lifting is close to the peak wind stresses in the GCM. However, NR15 demonstrated that in such cases the majority of dust lifting occurs in very few regions, which is unrealistic compared to the observed onset of GDS. In addition, while setting a high threshold wind stress does enable interannual variability in terms of the size of GDS, the timing and onset locations of these storms are far less variable than observed. For example, the Basu et al. (2006) GDS all began either in the Hellas basin or as northern hemisphere 'flushing' storms between $L_s \sim 240$ and 270°, while the NR15 storms assuming unlimited surface dust availability all began between $L_s \sim 265$ and 280°, with primary lifting in northern Hellas or on the slopes of Alba Patera (leading to a flushing storm also).

Coupling the dust and water cycles appears to increase intrinsic interannual variability in the system, as shown in Kahre et al. (2012) and Lee et al. (2018). As noted above, it is possible that such interactions are vital for permitting early 'equinoctial' GDS. However, more realistic variability in the dust cycle at present appears to require unrealistically strong variability in the water cycle (Lee, personal communication, 2018), thus this may not be the complete answer.

Another possibility is that realistic dust cycles require a specific pattern of surface dust availability. The simulations described above all assume unlimited surface dust everywhere, yet observations show that while some regions lack surface dust cover year-round, others gain or lose most of their surface dust during and after major storms, and still others do so on a seasonal basis (Szwast et al., 2006). Newman and Richardson (2015) found that a finite surface dust source enhanced the interannual variability of storms, also increasing the range of storm types, onset times and locations. In part, this was due to the additional 'memory' component of surface dust cover - e.g., if a major dust source region for a particular storm type is cleared during such a storm, it may be several years before it is sufficiently replenished to act as a source region again, thus reducing the likelihood of that storm type occurring until then. But another factor was that the primary source regions were no longer linked only to regions of peak wind stress. Indeed, regions with peak wind stresses typically had lifting rates that far exceeded deposition rates, hence lost all of their dust cover in a few years. Thus the new primary source regions became the highest wind stress regions that experienced no long-term loss of dust - i.e., regions for which lifting was balanced by replenishment over periods of no more than a few years. The simulations of NR15 set a lower wind stress threshold and allowed the finite dust supply to self-consistently arrange itself over the surface. Once the simulation reached a steady state, rapid onset dust storms still occurred but with far greater variability in timing and pattern than in the original high threshold, unlimited surface dust simulations. Difficulties remained, however, including a lack of equinoctial GDS (with only regional storms predicted then), differences between the predicted pattern of surface dust cover and observed surface albedo (although the two are not expected to map exactly), and the need to include a variable threshold (Pankine and Ingersoll, 2002, 2004) to reach a true steady state with no long-term change in surface dust cover at any grid point.

1.2. Orbit-spin coupling as a driver of increased interannual variability

In the past few years, another mechanism for interannual variability has been proposed. Shirley (2015) noted a statistically significant correlation between years with GDS on Mars and the phasing of the orbital angular momentum of Mars with respect to the Solar System barycenter (**L**). Shirley (2017) proposed that $d\mathbf{L}/dt$ may be a forcing function for atmospheric variability, and derived a mathematical expression for a weak 'orbit-spin' coupling between the orbital and rotational angular momenta of extended bodies that produces a small horizontal acceleration within the Mars atmosphere. Referred to as the "coupling term acceleration" (CTA), this force per unit mass takes the following form:

$$CTA = -c\left(\frac{d\mathbf{L}}{dt} \times \boldsymbol{\omega}_\alpha\right) \times \mathbf{r} \tag{1}$$

where $\boldsymbol{\omega}_\alpha$ is the angular velocity of Mars's rotation about its spin axis, **r** is a position vector identifying a location in the Mars body-fixed

coordinate system, and the coefficient c is a measure of the efficiency of the coupling between the orbital and rotational angular momentum reservoirs. The value of c is unknown but is constrained by solar system observations to be quite small (Shirley, 2017). Mischna and Shirley (2017), henceforth referred to as MS17, found that using $c = 5 \times 10^{-13}$ produced noticeable influence of the CTA on wind patterns and the circulation, yet did not produce unrealistically high wind speeds. For consistency with MS17, this value was used for all simulations here. However, future work should explore whether a different c value is more suitable when radiatively active dust and lifting feedbacks are included. For example, while increasing c will tend to increase the number/size of storms, decreasing the wind stress lifting rate parameter and/or increasing the threshold will do the opposite. Hence both c and the lifting parameters should be varied until the best match is found to the observed GDS sequence.

1.2.1. The sign and amplitude of *dL/dt*

While the cycle of solar irradiance repeats every Mars year, ***dL/dt*** has peak values ~1.2 Mars years apart and also has significant phase and amplitude modulation due to solar system dynamics (see Shirley, 2015). Each Mars Year (MY) thus has a different pattern of CTA at the same L_s, where MY 1 is defined as the Mars Year that began in Earth Year 1956. Hence a physical mechanism connecting ***dL/dt*** and dust lifting (via Eq. (1)) may naturally produce strong interannual variability in dust lifting and thus in GDS occurrence. Fig. 1 through 5 of Shirley and Mischna (2017), henceforth referred to as SM17, show the relationship between ***dL/dt*** and solar irradiance in all well-observed MYs from MY −16 to MY 39. As a reference, ***dL/dt*** for all MYs from MY −18 to MY 40 are also shown in Supplementary Materials, Fig. S1.

Of most interest is how the sign and amplitude of the ***dL/dt*** forcing in a given MY relates to the amplitude of solar forcing, which always peaks at perihelion ($L_s \sim 251^\circ$). As described in SM17, the phase of the ***dL/dt*** waveform is defined analogous to the sine function, with the upward-crossing zero-amplitude transition assigned a phase of 0°, the positive extremum a phase of 90°, and so on. Following SM17, each MY is categorized as positive, negative, or transitional polarity (with respect to perihelion) according to whether perihelion occurs near the positive extremum, the negative extremum, or one of the two zero-

Table 1
Definition of polarity with respect to perihelion, using ranges described in SM17.

Polarity	Positive	Negative	Transitional
dL/dt phase at perihelion	44° − 144°	225° − 315°	315° − 44° and 144–225°

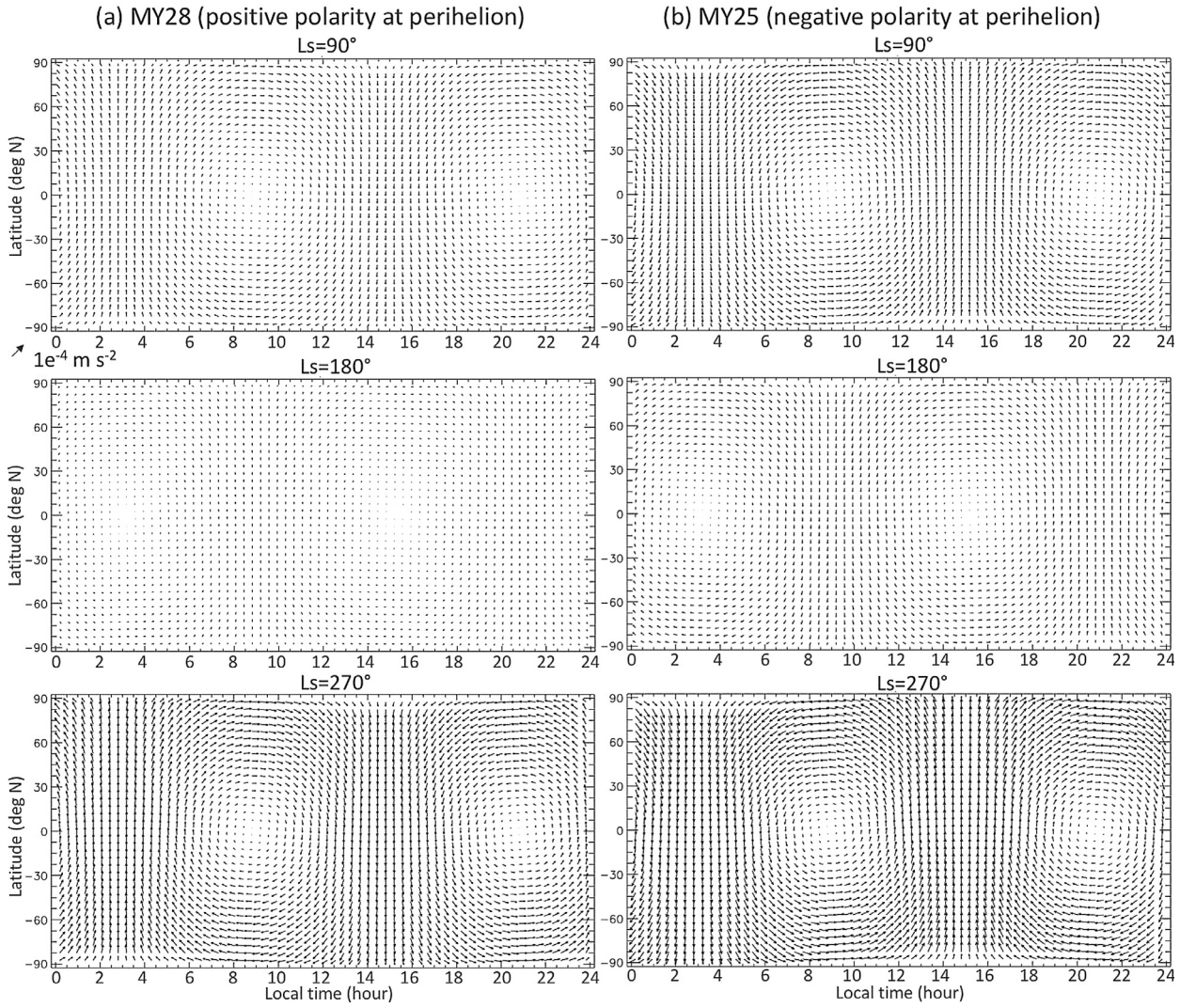

Fig. 1. The CTA as a function of local time at three times of year (L_s = 90°, 180°, and 270°) in MYs 28 (a positive polarity year; left column) and 25 (a negative polarity year; right column). Note that a ***dL/dt*** sign change between the L_s 90° and 270° periods in both years mostly compensates for the shifting of the pattern halfway around the planet every half year, hence for a given polarity the CTA directions are very similar in both periods.

Table 2
List of observed dust storm seasons, type and onset timing of dust storms, phase of ***dL/dt*** at perihelion, polarity, and comparison with type and onset timing of predicted dust storms in simulations CTA1 and CTA2. A tick or cross indicates the modeled storm respectively did or did not occur within 20° of an observed GDS. A question mark indicates a year for which the storm season is likely to be well observed but is wholly or partially in the future.

Mars Year	Earth Year at peri-helion	GDS onset (Ls in °)	φ of *dL/dt* at peri-helion (in °)	Pola-rity	CTA1 storm season	CTA1 mat-ches observ-ations	CTA2 storm season	CTA2 mat-ches observ-ations
-16	1924	**310**	92.6	+	**GDS, Ls ~ 240° ✗**	Yes	**GDS, Ls ~ 230° ✗**	Yes
-8	1939	None	309.1	-	None	Yes	None	Yes
1	1956	**249**	143.7	+	**Possible, Ls ~ 240° ✓**	Poss	**GDS, Ls ~ 200° ✗**	Yes
9	1971	**260**	92.5	+	**Possible, Ls ~ 250° ✓**	Poss	**GDS, Ls ~ 240° ✓**	Yes
10	1973	**300**	44.4	+	None	No	**Possible, Ls ~ 255° ✗**	Poss
11	1975	None	302.5	-	None	Yes	None	Yes
12	1977	**204, 268**	232.7	-	None	No	None	No
15	1982	**208**	98.7	+	**GDS, Ls ~ 200° ✓**	Yes	**GDS, Ls ~ 200° ✓**	Yes
17	1986	None	38.6	0	**GDS, Ls ~ 260°**	No	**GDS, Ls ~ 250°**	No
18	1988	None	1.5	0	None	Yes	None	Yes
21	1994	**254**	70.3	+	**GDS, Ls ~ 240° ✓**	Yes	**GDS, Ls ~ 240° ✓**	Yes
23	1998	None	1.6	0	None	Yes	**GDS, Ls ~ 265°**	No
24	1999	None	313.7	-	None	Yes	None	Yes
25	2001	**185**	272.4	-	None	No	None	No
26	2003	None	213	0	None	Yes	None	Yes
27	2005	None	134.5	+	None	Yes	**Possible, Ls ~ 250°**	Poss
28	2007	**262**	82.4	+	**GDS, Ls ~ 245° ✓**	Yes	**GDS, Ls ~ 240° ✗**	Yes
29	2009	None	37.8	0	None	Yes	**GDS, Ls ~ 255°**	No
30	2011	None	342.9	0	None	Yes	None	Yes
31	2013	None	272.1	-	None	Yes	None	Yes
32	2014	None	174.6	0	None	Yes	None	Yes
33	2016	None	105	+	**Possible, Ls ~ 240°**	Poss	**GDS, Ls ~ 235°**	No
34	2018	?	64.9	+	**GDS, Ls ~ 245°**	?	**GDS, Ls ~ 235°**	?
35	2020	?	35.5	0	**GDS, Ls ~ 260°**	?	**GDS, Ls ~ 250°**	?
36	2022	?	1.0	0	None	?	None	?
37	2024	?	297.2	-	None	?	None	?
38	2026	?	272.6	-	None	?	None	?

amplitude transitions of the ***dL/dt*** waveform. Table 1 shows the categorization used in SM17 and here, to which the reader is referred for more explanation of the precise choice of ranges. The polarity and phases at perihelion for all MYs for which the storm season was well observed, or is expected to be well observed in the near future, are given in Table 2.

1.2.2. The CTA patterns

The CTA at any given time can be visualized as a spherical harmonic order 1 pattern, as illustrated in Fig. 1 (see also SM17 for more details). For any sol at any L_s in any MY, the pattern consists of purely meridional accelerations at all latitudes at two local times spaced 12 h apart, with zero acceleration on the equator at the two points midway between these times. The pattern shifts in local time over the course of the year, with the zero points being at 9 a.m. and pm at $L_s = 90°$ and 270° but at 3 a.m. and pm at $L_s = 0°$ (not shown) and 180°, meaning that the influence of the CTA on other circulation components is highly specific to location, time of day, and season. Since neither the orbital plane nor the planetary rotation vector changes significantly over decadal/century time scales, the cross product of ***dL/dt*** and $\boldsymbol{\omega_a}$ is approximately along a fixed direction vector in inertial space. From Eq (1), this means that the shape of the CTA field is thus fixed in the inertial (sidereal) reference frame. Mars then rotates under this field once per sidereal day, which produces both a daily cycle that is roughly coincident with the solar day, but also a steady seasonal drift of the pattern such that it migrates fully around the planet in one year.

The CTA pattern is also modulated by the magnitude of ***dL/dt*** which varies little over timescales of a few sols but goes to zero and changes sign completely twice per ~1.2 Mars year cycle. This means that the CTA's interaction with other circulation components will have a different strength and sign in different years at the same season. Positive polarity years have southward CTA at most latitudes in the afternoon around perihelion, negative polarity years have northward CTA at most latitudes in the afternoon around perihelion, and transitional polarity years have small CTA amplitudes around perihelion. While GDS have been observed to begin as much as 65° of L_s away from perihelion, it lies roughly in the middle of most dust activity. Thus if the CTA affect surface wind stresses and hence wind stress dust lifting, one might expect them to have the least influence on dust activity in transitional years, whereas in other polarity years they may enhance or impede the existing circulation and hence increase or decrease dust lifting in some regions. This is explored in more detail in Section 5.

1.3. Prior dust-free modeling of the impact of the CTA in a Mars GCM

To explore the detailed influence of the CTA on Mars's circulation, MS17 included them in the MarsWRF GCM, as described in Section 2.2, and examined their impact on the Hadley circulation and wind stresses in dust-free simulations of 21 storm seasons. The likelihood of global dust storm occurrence was judged by considering the increase in predicted surface wind stress in a given year. This work demonstrated that years with positive polarity GDS often corresponded to an increase in predicted Hadley circulation strength and the global mean wind stress due to the presence of the CTA forcing. However, MS17 was unable to predict the onset timings or durations of storms, and although the global mean wind stress predicted for MY 33 was larger than some other positive polarity years in which GDS were predicted and observed, no GDS occurred in that year in reality. In addition, no wind stress enhancement due to CTA forcing was predicted prior to onset of negative polarity, equinoctial storms.

1.4. Aims and significance of this work

This work builds on that of MS17 by including the CTA in a version

of MarsWRF with parameterized dust lifting and radiatively active dust transport. It also builds on the dust cycle modeling study of NR15 by asking how the NR15 unlimited dust supply simulations are influenced by the inclusion of the CTA. As in NR15, dust is raised only where wind stresses exceed threshold, and feedbacks between dust lifting and the circulation are included. Hence it is possible to examine not only whether a GDS may be more likely to occur in a given year, but also the storm's predicted onset time, relative size, onset region(s), expansion pattern, etc. These characteristics can then be compared to the historical record of observed GDS, in order to better test the hypothesis that the CTA are largely responsible for controlling GDS activity on Mars.

The phasing of the CTA relative to the cycle of solar irradiance is highly predictable well in advance. Thus if such a coupling exists, and if it provides a major control on interannual variability in dust lifting and GDS on Mars, it has the potential to allow accurate predictions of likely dust storm activity decades in advance. This has major applications to future mission planning and the human exploration of Mars. It is possible that a range of mechanisms are needed to fully explain the pattern of Martian dust storms, such as water cycle coupling, finite surface dust availability, dust particle size evolution, and the CTA. Only the last is explored here, leaving the combination of multiple mechanisms to subsequent studies.

Section 2 describes the methodology of combining dust processes and the CTA in MarsWRF. Simulations and results are described in Section 3, including a comparison between observed and predicted storm seasons and how predictions differ from those of MS17. Section 4 discusses how the assumption of unlimited surface dust may have affected results. Section 5 explores how the CTA forcing produces different storm seasons in different MYs. Section 6 makes predictions for the current and future MYs, and Section 7 presents the paper's conclusions.

2. The MarsWRF general circulation model

MarsWRF is the Mars version of the planetary Weather Research and Forecasting (WRF) model (planetWRF) (Richardson et al., 2007; Toigo et al., 2012), which is based upon the terrestrial WRF model developed by the National Center for Atmospheric Research (NCAR) (Powers et al., 2017). In this paper, the model is used in the same configuration as that described by NR15. Richardson et al. (2007) provides details of the dynamical framework and grid setup, which includes a Runga-Kutta solver on an Arakawa C-grid, and discusses the model's conservation properties and basic setup for Mars. Richardson et al. (2007) also describes the vertical mixing scheme used here (referred to in that paper as the MRF PBL scheme; Hong and Pan (1996)), the radiative heating due to solar and thermal infrared radiation due to dust and CO_2 gas, and the model's CO_2 scheme which allows the CO_2 atmosphere to freeze out onto the surface if a grid cell becomes supersaturated and to sublime back into the atmosphere if the surface over CO_2 ice becomes sub-saturated. Toigo et al. (2012) provides details of the surface properties (albedo, emissivity, thermal inertia, topography, and roughness, derived from Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES) and Mars Orbiter Laser Altimeter (MOLA) datasets) used in the simulations.

Although MarsWRF may be run with nested higher-resolution domains, all simulations shown here use a single, global model domain with a grid spacing of 5.625° in longitude and 5° in latitude (64 × 36 horizontal grid points). In the vertical there are 21 levels of constant eta = $(P - P_{surf})/(P - P_{top})$, where P is pressure, P_{surf} is surface pressure, and P_{top} is pressure at the model top (which is close to zero). The eta values at the midpoint of all 20 model layers are the same as shown in Table 1 of NR15. MarsWRF may be run in non-hydrostatic mode, which is particularly important at mesoscale resolutions, but in these relatively low-resolution simulations the model is run in hydrostatic mode.

All simulations shown here used parameterized dust lifting with radiatively active dust transport. This means that dust is raised from the surface according to the local atmospheric conditions, mixed by the MRF PBL scheme, advected by model winds, and sedimented back towards the surface under gravity. A single particle size of 4 μm diameter is assumed for sedimentation. The evolving dust optical depths in MarsWRF impact radiative transfer and thus affect the atmospheric thermal structure and hence winds; in other words, dust lifting is consistent with the circulation, and feedbacks between dust lifting, atmospheric dust loading, and the circulation are enabled. Dust transport processes and parameterizations in MarsWRF are described in detail in NR15, but the dust lifting parameterizations are summarized in Section 2.1.

For consistency with prior multiannual dust cycle modeling, the simpler “wide band model” (WBM) radiative transfer scheme option available within MarsWRF is used (Toigo et al., 2012). For radiative heating due to CO_2 gas, the heritage WBM uses the Hourdin et al. (1992) scheme in the thermal infrared and the Forget et al. (1999) scheme in the solar bands. The WBM treats heating of the atmosphere due to the presence of dust using a scattering model in the visible (Briegleb, 1992) and an emissivity model in the thermal infrared (Haberle et al., 1982). The WBM in WRF is essentially identical to that used in Wilson and Hamilton (1996), Toigo and Richardson (2002), Basu et al. (2004, 2006), Toigo et al. (2012) and NR15. In the WBM's single solar channel, the optical properties are set following Clancy and Lee (1991) (single scattering parameter of 0.92, asymmetry factor of 0.55, and a forward scattering fraction of 0.3). In the WBM's single thermal infrared band, which excludes the CO_2 15-micron band (11.67–20.96 μm), the emissivity as a function of the optical depth was fit by Haberle et al. (1982) based on calculations using a fixed particle size distribution (see Fig. 4 and Table II of Haberle et al., 1982). In the WBM, the visible-to-optical depth ratio is set as a tunable parameter and a value of 1.43 is used here, which is lower than the value of 2–2.5 that is closer to estimates derived from observations (Clancy et al., 1995). The model value was found by optimizing the relationship between model predictions of observed air temperatures and dust optical depths. The lower value found for the model relative to that inferred directly from observations possibly compensates for the dust thermal emission that would occur in the real atmosphere within the 15-μm band. For both the visible and thermal infrared, dust optical properties are assumed fixed as a function of time and space.

The choice of fixed optical properties implies an unchanging aerosol particle size distribution and aerosol composition (i.e. only a consistent mineral dust is treated, with water ice and CO_2 ice aerosols neglected). The absence of realistic property variations will undoubtedly influence the circulation relative to that of the real Martian atmosphere. However, the work of NR15 suggests that a wide range of Mars-like dust storm and dust cycle behaviors still emerge from the model. Hence the setup is chosen to minimize inter-study differences, enabling an assessment of how the CTA forcing changes the skill of the model relative to its unforced state.

In addition, although interactions with the water cycle and finite surface dust availability are likely important in the real dust cycle, in this work water cycle processes are not included and unlimited surface dust availability is assumed at all locations and times. The impact of the CTA in more realistic dust cycle simulations will be explored in later work.

2.1. Parameterizations of dust lifting

Dust lifting from the surface is assumed to occur via two physical mechanisms: dust devils and wind stress.

2.1.1. Dust devil lifting

Dust lifting by dust devils is parameterized using the Rennó et al. (1998) treatment of them as convective heat engines. Lifting is set proportional to a ‘dust devil activity’ defined as:

$$\wedge \approx \eta F_s \quad (2)$$

where F_s is the surface sensible heat flux (the heat input to the base of the vortex) and η is the thermodynamic efficiency of the dust devil convective heat engine (the fraction of the input heat turned into work). The lifted dust flux, Q_{ddev} (in kg m^{-2} s^{-1}), is given by:

$$Q_{ddev} = a_D \eta F_s \quad (3)$$

where α_D is the dust devil lifting rate parameter (in units of kg J^{-1}). This formulation results in fluxes that increase predominantly with (a) the drag velocity and surface-to-air temperature difference, on which F_s largely depends, and (b) the thickness of the convective boundary layer, with larger boundary layers resulting in larger η. Predicted dust devil lifting thus peaks in local summer and is greatest in early afternoon when peak surface-to-air temperature gradients occur, as well as the most developed convective boundary layers.

The dependence on surface-to-air temperature difference results in a negative feedback on dust lifting by dust devils; in general, as more dust is lifted, the daytime atmosphere warms as more solar radiation is absorbed, while the surface cools as more solar radiation is blocked via absorption or scattering (Newman et al., 2002a; Kahre et al., 2006). This response to increased dust loading is opposite to that produced during the onset of storms, when dust lifting rapidly ramps up. In combination with observations showing no correlation between dust devils and dust storm occurrence (e.g. Cantor, 2007), this suggests that dust devils are key to the maintenance of the background dust cycle but do not contribute significantly to storms.

2.1.2. Wind stress lifting

The parameterization of dust lifting by wind stress assumes that dust particles are lifted during saltation of more easily moved sand-sized particles above a given saltation threshold. The lifted vertical dust flux is set proportional to the horizontal saltation flux using the formula of Kawamura (1951) and White (1979):

$$Q_{nsws} = \max[0, a_N(\rho/g){u_{drag}}^3(1 - {u^t}_{drag}/u_{drag})(1 + {u^t}_{drag}/u_{drag})^2] \quad (4)$$

where ρ is near-surface air density, g is acceleration due to gravity, u_{drag} is the drag (friction) velocity at the surface, u^t_{drag} is the threshold drag velocity required for saltation to occur, and α_N is the dimensionless wind stress lifting rate parameter. (Note that there is a typographical error in White's equation (22); the above follows White's equations (11)–(19) in which the form of the flux equation is derived.) The drag velocity is defined as:

$$u_{drag} \cong (\tau/\rho)^{1/2} \quad (5)$$

where τ is the wind stress at the surface. The value of τ is calculated in MarsWRF's surface layer parameterization, which links the net eddy stress (and fluxes of momentum, heat and tracers) to model-resolved winds and the thermal structure in the surface layer. The parameterization used for these simulations is the original version of the WRF SFCLAY Monin–Obukhov scheme (Jimenez et al., 2012). The surface layer is defined as the lowest portion of the atmosphere in which the fluxes are invariant with height (practically, where they vary by less than about 10% in magnitude). The roughness length scale, z_0, for the stress calculation comes from MOLA pulse width maps generated by Garvin et al. (1999), as discussed in detail by Heavens et al. (2008). A constant wind stress threshold, τ^t, is used for all locations and times, with u^t_{drag} found via Eq. (5).

Unlike dust devils, resolved wind stress lifting generally has a strong positive feedback response, both on local scales (as thermal gradients and hence winds increase at the edge of dust clouds) and in a global sense (as the global circulation intensifies for large dust loadings), see e.g. Newman et al. (2002a). This response is in line with the rapid onset of dust storms, suggesting that wind stress lifting is largely responsible for them.

2.2. Adding orbit-spin coupling term accelerations to MarsWRF

The dynamical core of MarsWRF was modified in MS17 to include updates to the zonal and meridional wind tendencies (dU/dt and dV/dt, respectively) due to the CTA as defined in Eq. (1). The value of $\boldsymbol{dL/dt}$ is obtained from solar system ephemerides using methods provided in Shirley (2017) and SM17, and $\boldsymbol{\omega_\alpha}$ is the angular velocity of axial rotation of Mars in inertial space. A table of the J2000 ecliptic Cartesian components of $\boldsymbol{dL/dt}$ with two-Julian-day resolution from 1920 to 2030 is included in the Supplementary Materials of MS17.

In MarsWRF, at each model dynamical timestep (typically every 3 min), $\boldsymbol{dL/dt}$ is looked up and interpolated to the current time. $\boldsymbol{dL/dt}$ is then calculated and transformed to the body-fixed coordinate system of most GCMs (such as MarsWRF), as detailed in Appendix A of MS17. The triple product $(\frac{\boldsymbol{dL}}{\boldsymbol{dt}} \times \boldsymbol{\omega_\alpha}) \times \boldsymbol{r}$ is then found for all 3-D grid point locations in the atmosphere and multiplied by the coupling efficiency, c, to give the vector acceleration term. In the version of MarsWRF used in this study, the implementation has been slightly modified such that the difference between the sidereal and solar rotation rate of the planet (discussed in Section 1.2) is properly accounted for. Finally, the zonal and meridional components are found and added to the corresponding wind tendencies for that timestep.

3. MarsWRF simulations with an active dust cycle and the CTA

The basic goal of these experiments is to assess the impact of adding the CTA into the MarsWRF GCM with parameterized dust lifting and active dust transport, and to determine whether adding the CTA increases the model's predictive skill with respect to dust storms.

3.1. Tuning the dust lifting parameters by comparison with metrics of dust storm activity

In order to run active MarsWRF simulations with parameterized dust lifting, the dust lifting parameters described in Section 2.1 must be chosen in some way. In this work, as in NR15, they are 'tuned' by comparing the seasonal evolution of global mean lower-middle atmosphere temperature in the model with that observed on Mars. Specifically, the 'T15' temperature is calculated, which was first used as a metric for dust storm activity by Basu et al. (2004). This is the atmospheric temperature vertically weighted to mimic observations in the Viking IRTM instrument's 15 μm band - which was sensitive to a region between ~15 and 35 km - and then averaged over the 40°S to 40°N latitude range. Since dust loading and major dust storms strongly affect temperatures aloft over low and mid-latitudes, this T15 temperature metric provides a very sensitive measure of a dust storm's size and the period over which it is active. Comparing modeled and observed T15 temperatures thus enables the realism of modeled dust storms to be assessed.

Fig. 2 shows the T15 curves for MY 24–32 based on Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES) and Mars Reconnaissance Orbiter (MRO) Mars Climate Sounder (MCS) observations. Rather than obtain T15 estimates based on TES spectra or MCS temperature retrievals, these T15 values were obtained by running a MarsWRF simulation with prescribed dust loading based on TES and MCS retrievals of column dust opacity as presented in the Montabone et al. (2015) dust maps. The advantages of this approach are two-fold: (1) these 'observational' T15 values may be calculated using the same sampling as is used to find T15 values for MarsWRF simulations, and (2) a second metric for observed dust storm activity is produced in the form of the increased intensity of the Hadley circulation predicted by the observationally-forced model during the peak of the storms.

The disadvantage is that the vertical dust distribution must be assumed in this prescribed dust simulation and will have some impact on the T15 values produced. The vertical dust distribution was set to have

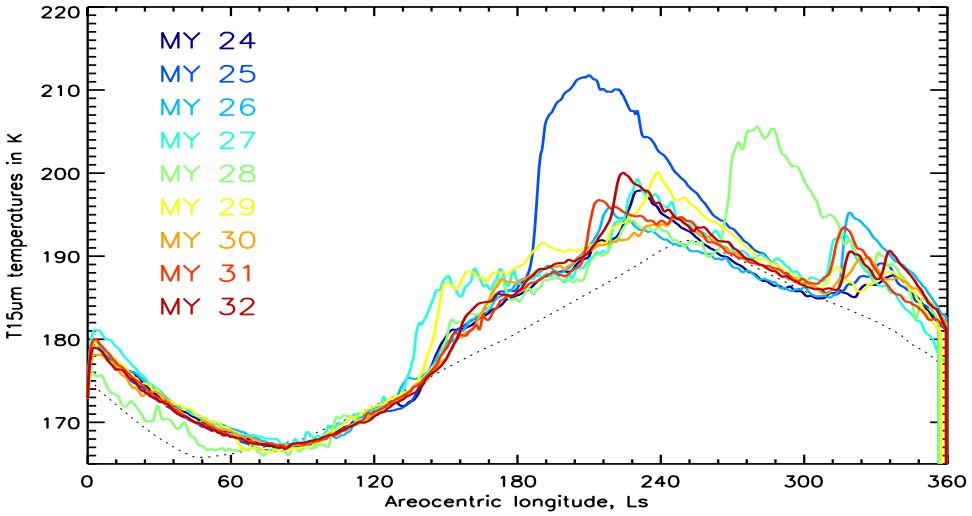

Fig. 2. T15 curves (see text) for nine TES / MCS Mars years, from an observationally-forced simulation with dust prescribed based on retrieved column dust opacities (see text). The faint dotted line is provided to assist the eye in comparing T15 curves between plots.

a constant mass mixing ratio from the surface up to an elevation z_{max}, above which it rapidly declines, with the latitudinal and seasonal dependence of z_{max} following that of the standard Mars Climate Database MGS dust scenario (Montmessin et al., 2004). The MGS dust scenario was designed to mimic a typical non-dust-storm year of MGS observations, but low latitude z_{max} still reaches over 70 km during the bulk of the storm season. Note, however, that Fig. 7 of Liu et al. (2003) shows T15 values at 2 a.m. and 2 p.m. obtained more directly from TES spectra, and these bracket the values shown in Fig. 2 (which are averaged over all local times of sol). Liu et al. (2003) also show T15 results for the Viking and Mariner 9 years, adding to the historical record that may be used for comparison.

Two global dust storms occurred during the TES/MCS period prior to the MY34 (current) storm season, one in MY 25 and one in MY 28, and both stand out very prominently in the T15 plot, with peak T15 temperatures exceeding 205 K in both years. The next highest dust loading occurs in MY 29, but peak T15 temperatures do not exceed 200 K. Perhaps more significant than the peak T15 value achieved during a storm is the increase in T15 compared to the 'background' in years when no major storm is active at this L_s; i.e., the difference between the GDS peak and the lowest T15 value at this L_s in any year. The increase at the peak of the MY 25 storm is ~21 K, while the increase at the peak of the MY 28 storm is ~16 K. In all other MY, the increase over the background never exceeds ~10 K at any time of year. Liu et al. (2003) Fig. 7 shows a comparable difference in peak T15 values and increase over the background for global dust storms in the Mariner 9 and Viking years.

As described in Section 2.1, two dust lifting parameterizations are used, one for dust devils and one for wind stress lifting. The dust devil parameterization has one parameter that scales the amount of dust lifted, α_D, while the wind stress parameterization has two parameters: the wind stress threshold for lifting to occur, τ_t, and a parameter that scales the amount of dust lifted, α_N. The tuning process, described in NR15, assumes that dust devils provide background dustiness while wind stress is responsible for storms. Hence α_D is first tuned such that the simulated T15 curves, with dust devil lifting only, match the observed 'background' (storm-free) T15 values. Following this, wind stress lifting is added and a threshold wind stress is chosen, with α_N then adjusted until the most realistic set of storms (based on comparison with the T15 in observed storm years, as shown in e.g. Fig. 2) are produced.

3.2. Criteria for determining whether a global dust storm has occurred

Using the T15 metric is the cleanest and simplest method of categorizing the size of a given storm in a simulation and also allows tens of simulation years to be presented and interpreted in one figure. Based on observations: years with peak T15 more than 16 K higher than the background are categorized as GDS years; years with peak T15 less than 10 K higher than the background are categorized as non-major storm years, and years with peak T15 between 10 K and 16 K above the background are categorized as possible GDS years.

Two other metrics for determining the category of simulated storms were also explored: the increased strength of the mass streamfunction during the peak of the storm compared to storm-free years, and the evolving dust spatial coverage in terms of visible column opacity maps. Both were produced for the observationally-forced simulation described above and are shown in Supplementary Material Figs. S2 and S3, respectively. The same metrics were also produced for all years of simulations CTA1 and CTA2 (described in Section 3.4 below), with selected years of CTA1 shown in Figures S4 and S5, respectively. However, as they essentially confirm the categorizations made using the T15 metric, they are not discussed further here.

3.3. Results of control experiments with no CTA included

Fig. 3 shows 14 years of T15 temperatures from control experiments CON1, CON2 and CON3 in which no CTA were included. All three simulations were run using the same wind stress threshold (0.036 Pa) and α_D value (5×10^{-9}), but with $\alpha_N = 1.5 \times 10^{-5}$, 2.5×10^{-5} and 4×10^{-5}, respectively. Using the criteria given in Section 3.2 (and using storm-free simulation CON1 as the 'no storms' background for simulation CON3), the α_N values used in CON1 and CON2 produce no major storms in any year, whereas the higher value used in CON3 produces a GDS in every year.

Note that none of these simulations show significant interannual variability in terms of the storm size, shape, or timing. This is expected given the results of NR15, who found that interannual variability only occurred for wind stress thresholds of $\gtrsim 0.047$ Pa (when surface dust was unlimited). In these control simulations the forcing is identical in all years and is purely due to seasonal variations in solar irradiance, which repeats exactly from year to year. Thus the slight year-to-year differences within each simulation are due entirely to intrinsic atmospheric variability, enhanced during the storm season in particular by non-linear feedbacks between the atmospheric dust distribution, circulation, and dust lifting.

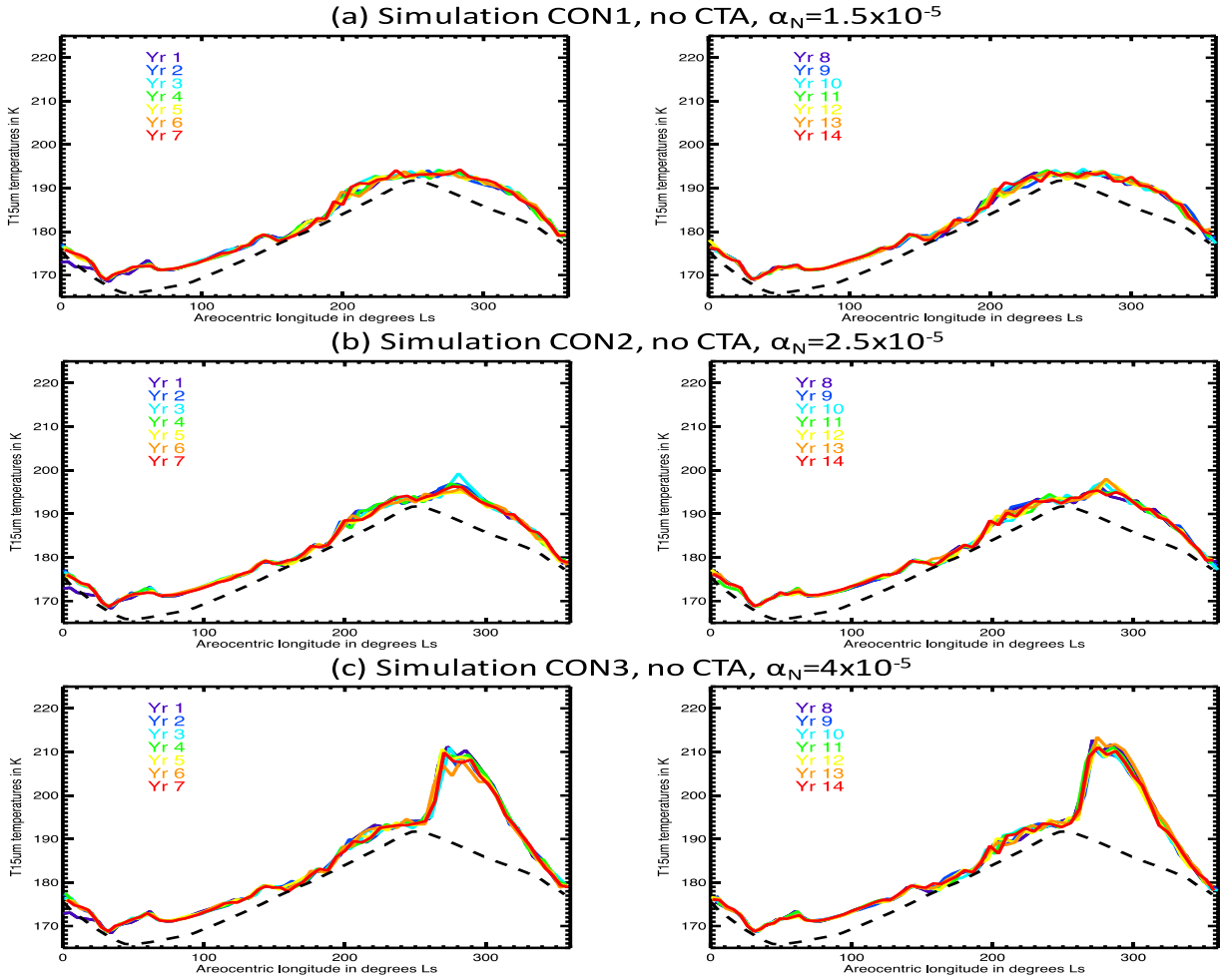

Fig. 3. 'T15' temperatures as a function of season for three 14-Mars-year dust cycle simulations with no CTA included. Each panel shows 7 Mars years. All three simulations have $\alpha_D = 5 \times 10^{-9}$ and $\tau_t = 0.036$ Pa, but simulations CON1, CON2 and CON3 have $\alpha_N = 1.5 \times 10^{-5}$, 2.5×10^{-5} and 4×10^{-5}, respectively.

3.4. Results of experiments with CTA included

Figs. 4 and 5 each show 56 years of simulations CTA1 and CTA2, which have the CTA included but otherwise identical settings (i.e., dust lifting parameters) to those used in simulations CON1 and CON2, respectively. Simulation CTA3 (identical to CON3 with the CTA included) produces GDS so large in some years that they massively increase atmospheric temperatures and circulation strength and crash the model, hence those results are not shown. Note that now the different years are labeled with MY, indicating that the CTA accelerations specific to that particular MY have been used in the simulation. The simulations begin at MY −17 (Earth year 1922) and continue through MY 38 (Earth year 2026).

By contrast with the Control runs, the CTA simulations clearly have far more interannual variability. While no CON simulations for this τ_t value show any significant interannual variability, CTA1 and CTA2 each have years with no major storms, years with large regional storms, and years with GDS of differing sizes. This is consistent with the CTA affecting the strength of the Hadley circulation and surface wind stress patterns during the storm season, especially in positive and negative polarity years, as discussed in MS17. The relationship between CTA polarity, its effect on the circulation and wind stress patterns, and the storm season produced in selected MY, is examined in detail for simulation CTA1 in Sections 5 and 6.

The onset time of GDS also varies from year to year in the CTA simulations, ranging from as early as $L_s \sim 200^\circ$ to as late as $L_s \sim 260^\circ$. While this is a smaller range of onset times than real storms (which have been observed to start as early as $L_s \sim 185^\circ$ and as late as $L_s \sim 310^\circ$), it is far broader than the onset times of global storms in simulation CON3, all of which start within a few degrees of $L_s \sim 260^\circ$. Thus the inclusion of the CTA appears to permit an earlier as well as more variable timing of dust storm onset than would otherwise be possible for the threshold chosen here, at least in the absence of a water cycle or limited surface dust cover. Despite the range of onset times, however, no GDS decays before the end of the storm season, which is contrary to observations and suggests that a process not included here is important for ending the storms sooner (e.g. source dust depletion or scavenging by ice particles).

One last point to note is that the timing, shape, and relative size of the storm simulated in a given MY is very similar between the two with-CTA simulations. For example, in both CTA1 and CTA2 (Figs. 4 and 5) the strongest storms occur in MYs 2, 3, 15 and 16; relatively early storms occur in MYs −17, −16, 1, 2, 14 and 15; and a small maximum in T15 occurs in northern spring in MYs 2, 3, 15 and 16. There are some exceptions - e.g., in simulation CTA1 the storm in MY −17 is weaker than that in MY −16, whereas in simulation CTA2 the MY −17 storm is marginally the stronger. In general, however, it is clear that the CTA are strongly influencing the timing, shape, and size of the dust storms produced in a given MY. In short, the CTA are dominating interannual variability in the timing and strength of dust storms in these simulations.

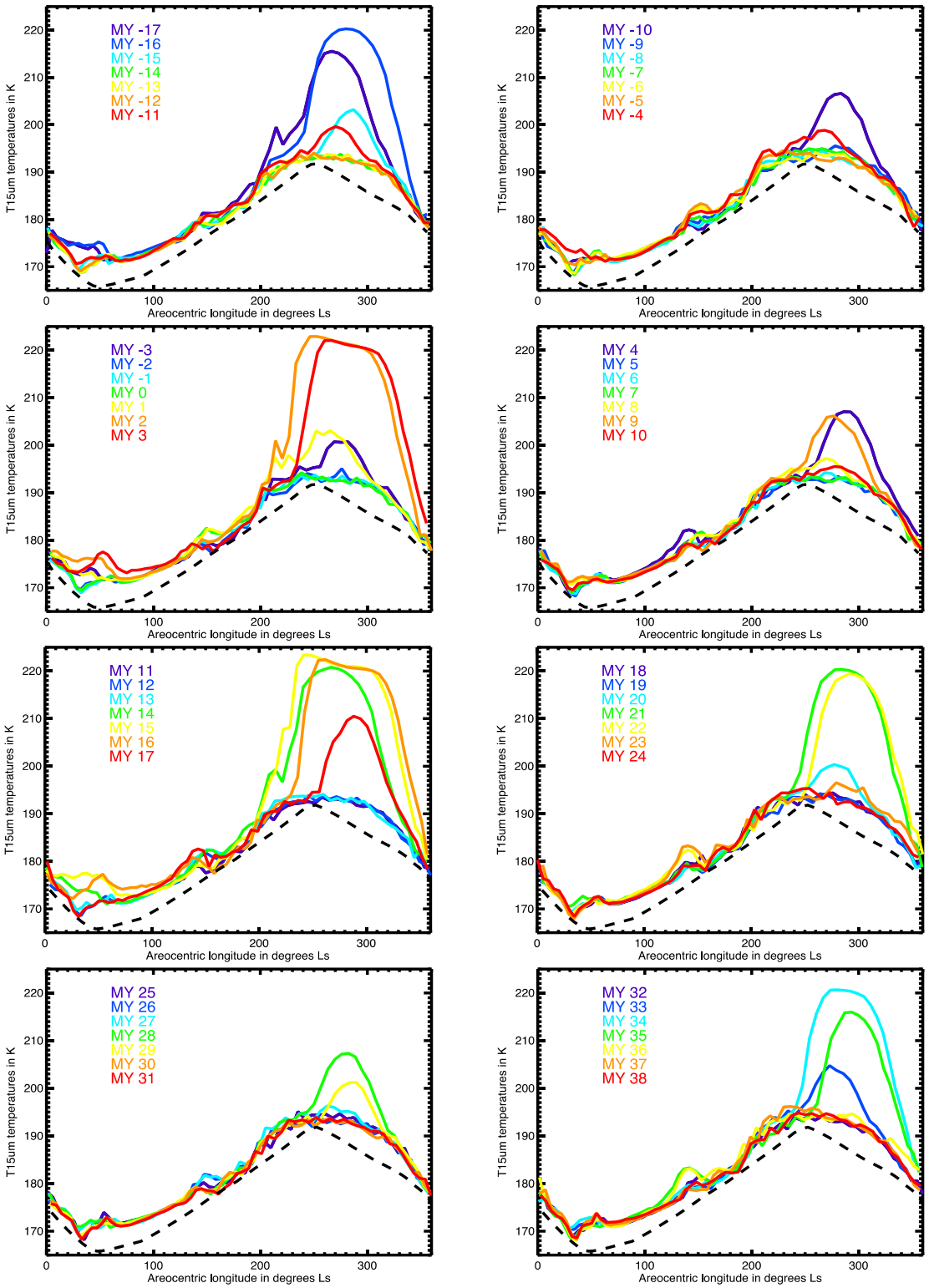

Fig. 4. Simulation CTA1 results. As in Fig. 3a, using the same dust lifting parameters as CON1 but with the CTA accelerations included and lasting 56 Mars years.

3.5. *Comparison with the pattern of observed storm seasons*

Table 2 shows a comparison between observed dust storm seasons and those simulated in CTA1 and CTA2. Simulation CTA2 predicts the type of storm season correctly in 14 years out of 22, with the timing of the GDS correct in 3 out of 6 years. Another 2 years have marginal results: MY 27 has a possible GDS where no GDS was observed, while MY 10 has only a possible GDS where a GDS *was* observed, and at the wrong time of year. In the remaining 6 years, either an observed GDS is not predicted (in MYs 12 and 25) or the model predicts a GDS when no GDS was observed. In summary, CTA2 tends to overpredict the amount of GDS activity, including the incorrect prediction of three GDS in transitional years, but does not predict the GDS in negative polarity storm seasons.

Simulation CTA1 does even better, predicting the type of storm season correctly in 15 out of the 22 observed storm seasons considered,

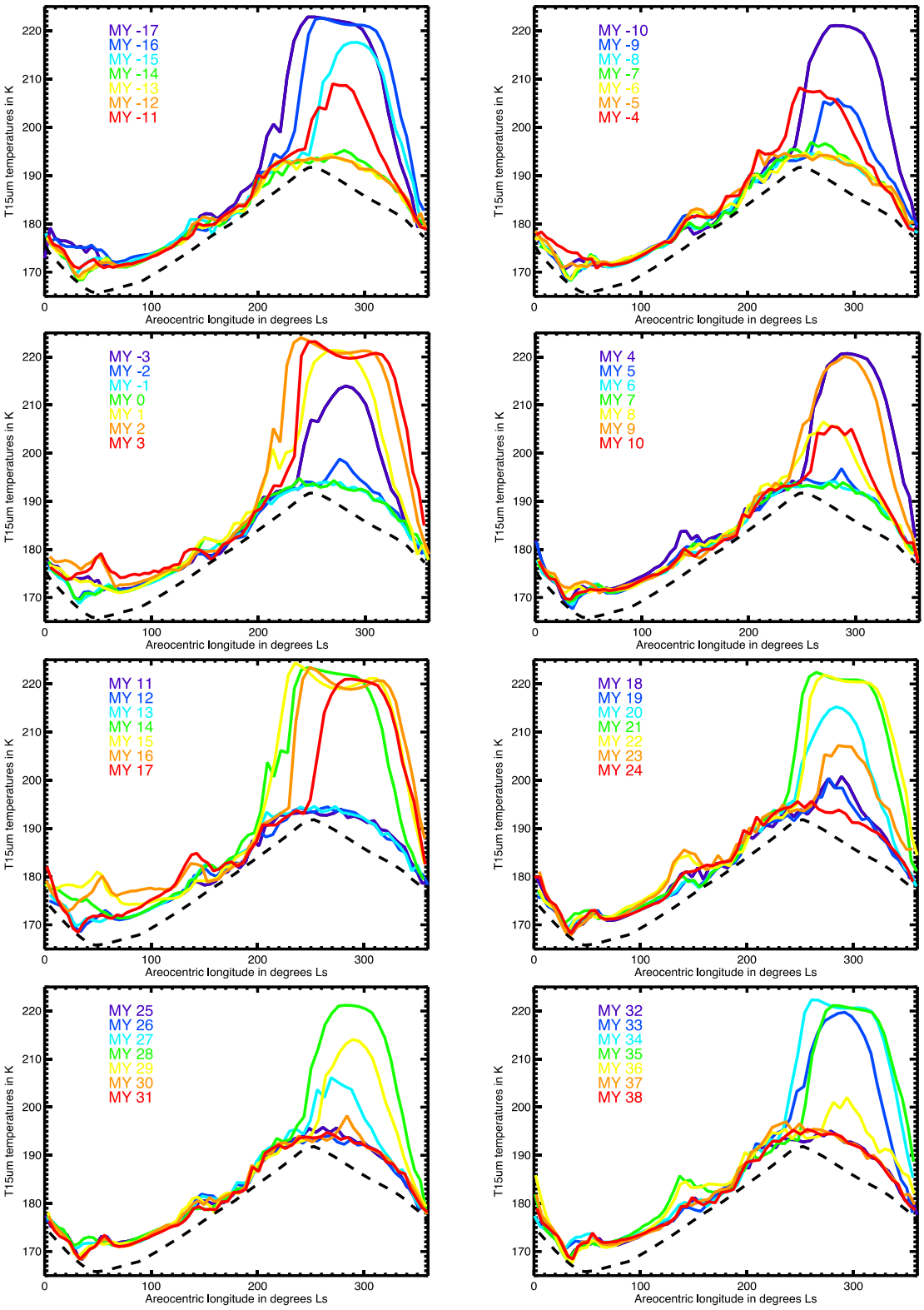

Fig. 5. Simulation CTA2 results. As in Fig. 3b, using the same dust lifting parameters but with the CTA accelerations included and lasting 56 Mars years.

with the timing of the predicted GDS correct in 4 out of 5 years. Another three years have marginal results: MY 33 has a possible GDS when no GDS was observed, while MYs 1 and 9 have only a possible GDS where a GDS *was* observed, although in this case the storm timing is correct in both of those years. In the remaining 4 years, either an observed GDS is not predicted (MY 10 and the two negative polarity years, MY 12 and 25) or a GDS is incorrectly predicted (in transitional year MY 17). Overall, this suggests that the smaller α_N value chosen for CTA1 is more realistic. A statistical analysis reveals that the number of successes obtained in simulation CTA1 differs from success rates obtained using stochastic model forcing at the 99% significance level (Shirley et al., 2019).

Several other simulations have been performed with differing values of α_N, but CTA1 produces the best overall match to observations. Hence

the remainder of this paper focuses on simulation CTA1's results.

4. Possible impact of unlimited dust cover on results

The simulations presented here all assumed unlimited surface dust availability, so it is useful to consider how instead allowing surface dust to deplete might have altered these results.

4.1. The effect of CTA in simulations with highly localized dust lifting centers

MS17 used the enhancement in Hadley circulation strength and global mean surface wind stress to infer whether a GDS would occur. By contrast, the simulations presented in this paper explicitly predict the dust lifting that results from the altered wind stress patterns, and include feedbacks between raised dust, the circulation, and future wind stress. However, these simulations still make one very big assumption: that dust is available to be lifted from the surface everywhere. This assumption results in the majority of dust lifting occurring in very few primary source regions in which the wind stress regularly exceeds threshold, as discussed in NR15.

Fig. 6(a) and (b) demonstrate where the majority of dust lifting occurs in simulations CON1 and CTA1 by showing the change in surface dust after 20 years of each simulation. There is a distinctive difference between the CON1 and CTA1 pattern, with the latter having increased peak dust removal within a southern hemisphere 'zonal collar,' in particular over northern Hellas and to the north and NW of Argyre. This is to be expected, as the first 20 years of CTA1 includes three GDS (see Fig. 6), and enhanced lifting typically occurs in this region during the expansion phase of major dust storms as the Hadley circulation and associated near-surface wind stresses intensify. But regardless of whether the CTA are included or not, the majority of dust lifting clearly comes from a very restricted set of grid points, primarily on the slopes of northern Tharsis (mainly around Alba Patera and Olympus Mons) and the slopes of northern Hellas, as well as N/NW of Argyre in CTA1.

Significantly, the concentration of source regions and basic pattern of dust lifting do not change when the CTA forcing is included. In other words, rather than changing the pattern of dust lifting entirely, the bulk of dust lifting continues to come from the same small number of source regions. Hence it is only the CTA impact on wind stress magnitudes *in those regions* that significantly affects year-to-year variability. This closely ties the GDS predictions in the model to the CTA's influence *in a small number of areas* in which the wind stress threshold is regularly exceeded, rather than - as assumed in MS17 - to the difference in global mean wind stress. This will naturally produce some differences between our predictions and those made by MS17, although in both cases the majority of the signal comes from the same, high wind stress regions.

The restricted set of source regions found here is typical of simulations with unlimited surface dust cover, in which the primary sources for dust storms are simply the peak wind stress regions, regardless of the fact that such regions may experience far more dust removal than deposition and hence may have little to no surface dust availability in reality (NR15). By contrast, Martian dust storms are observed to originate from lifting over a far wider range of source regions (e.g. Strausberg et al., 2005; Szwast et al., 2006; Cantor 2007; Wang and Richardson, 2015). As shown in NR15, assuming finite surface dust availability (and a lower threshold) and allowing the surface dust distribution to rearrange itself self-consistently results in primary lifting regions that are far more evenly distributed over the Martian surface. With more widespread source regions, the results of MarsWRF dust cycle simulations with the CTA included will likely change and may provide a better match to observations, in terms of both the realism of onset locations (see Section 5.2) and the prediction of GDS and non-GDS years. This suggests that investigating the impact of CTA on such simulations should be a priority in the future.

4.2. Prevention or early decay of global storms due to insufficient dust

While difficult to assess from this work, it is possible that some predicted storms did not occur on Mars itself due to a lack of available dust. In particular, a GDS may move dust out of a certain region, with the region taking several years to recover the dust reserves it had before; conversely, a GDS may move dust into a region that is typically kept relatively dust-free by dust devil or small scale lifting processes (e.g. Szwast et al., 2006). In CTA1, the MY 17 storm is predicted but not observed, and follows a run of three years with predicted GDS (MYs 14–16), out of which we know that one (in MY 15) definitely occurred and do not have information for the others. It therefore seems plausible to suggest that a MY 17 GDS would have occurred if not for prior GDS activity that scoured its source regions of dust. However, we will never be able to determine this with any certainty.

As noted in Section 3.4, the global storms predicted in simulation CTA1 tend to last longer than those observed, with most not decaying until the end of the storm season regardless of their onset time. In addition, the T15 values for two of the predicted positive polarity GDS (in MYs 15 and 21) are larger than those of any GDS observed by spacecraft instruments to date (compare the T15 curves in Figs. 2 and 4). This may be a result of the small sample size available from observations, but it is also possible that these GDS are unrealistically large. Both issues may potentially be explained by the unlimited surface dust assumption, with storms found to shut down faster and hence end earlier in finite dust simulations (e.g. NR15) due to source dust running out during the storm's growth period.

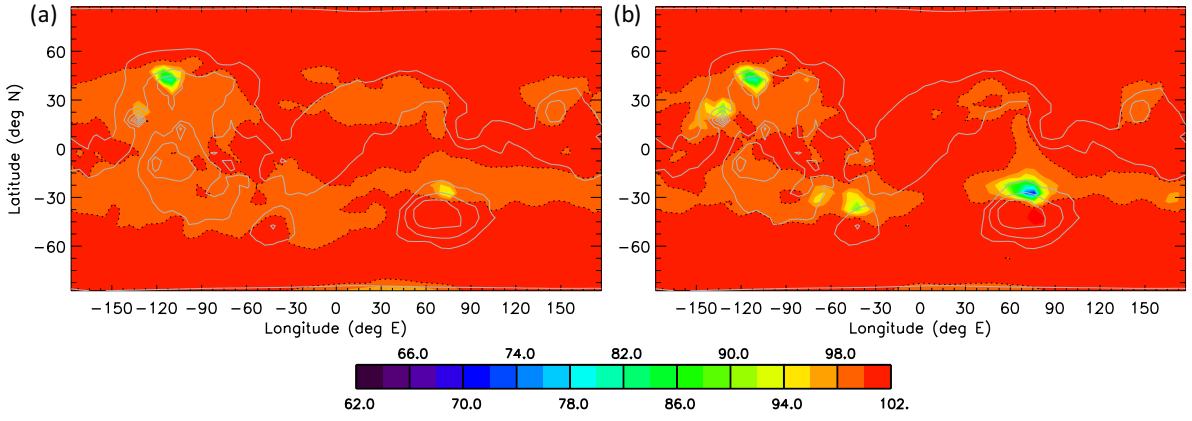

Fig. 6. Surface dust cover (in kg m^{-2}) after 20 years of simulations (a) CON1 and (b) CTA1. Note that the initial surface dust cover is 100 kg m^{-2} in these simulations, but this is effectively unlimited as no region is fully depleted by the end of the simulation.

5. The causality of different storm seasons in simulation CTA1

This section examines the effect of the CTA in several MYs of CTA1. It begins with a general look at typical storm seasons in different polarity years (Section 5.1), then examines the link between the CTA and changes in the circulation and dust lifting during onset of four positive polarity year storms (Section 5.2). Note that this analysis focuses on how the CTA forcing triggers enhanced lifting and major storm onset, but does not explore how the CTA may potentially affect the expansion phase or final size of storms. The problem is that, once dust loading has increased past the onset phase, it will dominate differences in the circulation over those produced by the CTA themselves, at which point a comparison between CTA1 and CON1 results is no longer useful in examining the CTA's influence alone.

For this reason, Section 5.2 focuses on the period of GDS onset in each MY, i.e., the period over which enhanced dust lifting in simulation CTA1 first produces a clear, continuous increase in dust opacities and T15 values relative to simulation CON1. In order to assess what role the CTA may play in the subsequent expansion of storms, idealized experiments would be required with identical dust loadings (or lifting rates) in the with- and without-CTA simulations to properly examine the CTA influence, which is beyond the scope of the present work.

Finally, Section 5.3 compares positive polarity MYs 1 and 27, which have very similar phases but different magnitudes of ***dL/dt***, and demonstrates that no storm is predicted in MY 27 due to the smaller magnitude of the CTA forcing compared to MY 1.

5.1. Polarity and storms

Breaking CTA1 down by polarity, the results shown in Table 2 may be summarized as follows:

5.1.1. Positive polarity years

For positive polarity years, CTA1 fails to predict a GDS in MY 10 and predicts a possible GDS in MY 33 (when no GDS occurred). However, it succeeds in correctly predicting a GDS in MYs −16, 15, 21 and 28, predicts a possible GDS in MYs 1 and 9, and correctly predicts no GDS in MY 27. The onset timing is also the same as observed in all but one of the GDS or possible GDS years, with some realism also noted in onset locations. The initiation of these positive polarity year GDS is explored in detail in Section 5.2.

The good performance for positive polarity storm seasons suggests that the model may be close to correctly capturing the interactions between the CTA and the rest of the forcings in this case. Note that tuning the lifting rate parameters for other combinations of wind stress threshold and c values has not been explored. In addition, some of the source regions that contribute dust to the modeled storms may be dust-free in reality, and/or other physics responsible for certain circulation elements (such as those introduced by coupling with the water cycle) may be missing, both of which may be affecting the dust cycle adversely even if the chosen threshold and c value are correct. For example, the GDS that occurred in MY28 was observed to grow primarily in Noachis, located roughly midway between the Argyre and Hellas basins (e.g. Wang and Richardson, 2015). However, there is no major lifting center here in simulation CTA1 (see Fig. 6b). This is because regions with the highest wind stresses - the slopes of Alba Patera, Tharsis, and Hellas - dominate in simulations in which surface dust cover is assumed to be unlimited, as is the case in CTA1. NR15 explored the impact of allowing surface dust to be exhausted in regions with more dust lifting than deposition. The primary lifting centers are very different in such simulations because the regions with the highest wind stresses (which dominate in unlimited dust simulations) are typically exhausted of dust. Hence NR15 found that Noachian storms often occurred in their finite surface dust simulations.

Finite surface dust simulations can also produce significant T15 increases after $L_s \sim 300^\circ$ (e.g. NR15, Fig. 10a), due to a primary dust source that is triggered late in the storm season. The presence of strong lifting at this time of year even in the absence of CTA forcing may be vital to predicting late GDS (such as that observed in MY 10) when the CTA are included. MY 10 is a year in which simulation CTA1 predicts no major storm activity, but the peak of the ***dL/dt*** waveform is very late compared to all other positive polarity storms (see Fig. S1 or SM17, Fig. 4), indicating the potential for enhancing any existing dust lifting that occurs late in the storm season.

All of this suggests that a better match to observed storm seasons may potentially be produced when the CTA are combined with finite dust availability.

5.1.2. Negative polarity years

Simulation CTA1 predicts no GDS in any negative polarity years, which is correct for MYs −8, 11, 24 and 31. However, this means that it also fails to predict the first GDS observed in MY 12 or the GDS observed in MY 25. (Note that the second GDS of MY 12 may have occurred due to atmospheric changes caused by the first, hence the focus here is on why the first GDS is not predicted in that year.)

Aside from MY 15, which was exceptional as it had the largest CTA amplitudes of the period MY -18 to 40 (see Fig. S1), the early MY 12 and equinoctial MY 25 GDS are the earliest observed GDS by at least 45° of L_s (see Table 2). A possible explanation for not predicting them is therefore the lack of coupling between the dust and water cycles in these simulations. A fully interactive water cycle that includes scavenging of dust and the radiative effects of ice clouds is known to alter the circulation, in particular by changing the distribution of surface water ice deposits and associated flows. This effect has been suggested as a possible mechanism for producing GDS with onset shortly after southern spring equinox (e.g. Kahre et al., 2012; Lee et al., 2018), as global storms had not otherwise been predicted this early in any dust cycle models (e.g. Basu et al., 2004, 2006; Kahre et al., 2006; Mulholland et al., 2013; NR15). If such a vital process for increasing wind stresses around equinox is missing, it is not surprising that the CTA alone may have trouble reproducing some GDS that occur early in the storm season. Future work should therefore explore which circulation elements are strengthened around equinox in coupled water simulations, and how these would interact with the CTA in years with negative, positive, and transitional polarities. Note that such coupled simulations typically allow the dust (and water ice) size distributions to evolve, affecting dust (and water ice) optical properties, hence also improve the realism of radiative transfer through the model atmosphere.

5.1.3. Transitional polarity years

Simulation CTA1 has only one incorrect prediction of a GDS (in MY 17) out of all seven transitional polarity storm seasons. The lack of an observed GDS in any transitional year seems to suggest that the CTA are the crucial factor in shifting the Martian circulation from having no GDS (as in simulation CON1) to producing them (in CTA1), by strongly perturbing the circulation and surface wind stress patterns. However, the phase with respect to perihelion in MY 17 (38.6°) is the highest of all the transitional years and the magnitude of ***dL/dt*** is quite large when the storm begins. As noted in Section 4.2, it is possible that this storm would have occurred on Mars but did not due to a lack of available surface dust, as MY 17 follows a period predicted to have three GDS in a row. Additional discussion of the anomalous transitional polarity phasing of the simulated MY17 GDS is provided in Shirley et al. (2019).

5.2. The cause of GDS onset in selected positive polarity MYs

This section examines how the CTA resulted in enhanced dust lifting, and storm onset very close to the observed GDS onset time, in four positive polarity years: MYs 9, 15, 21 and 28. An average over 8 years of simulation CON1, which had no CTA included and produced no regional or global dust storms, is used for comparison in all cases. The

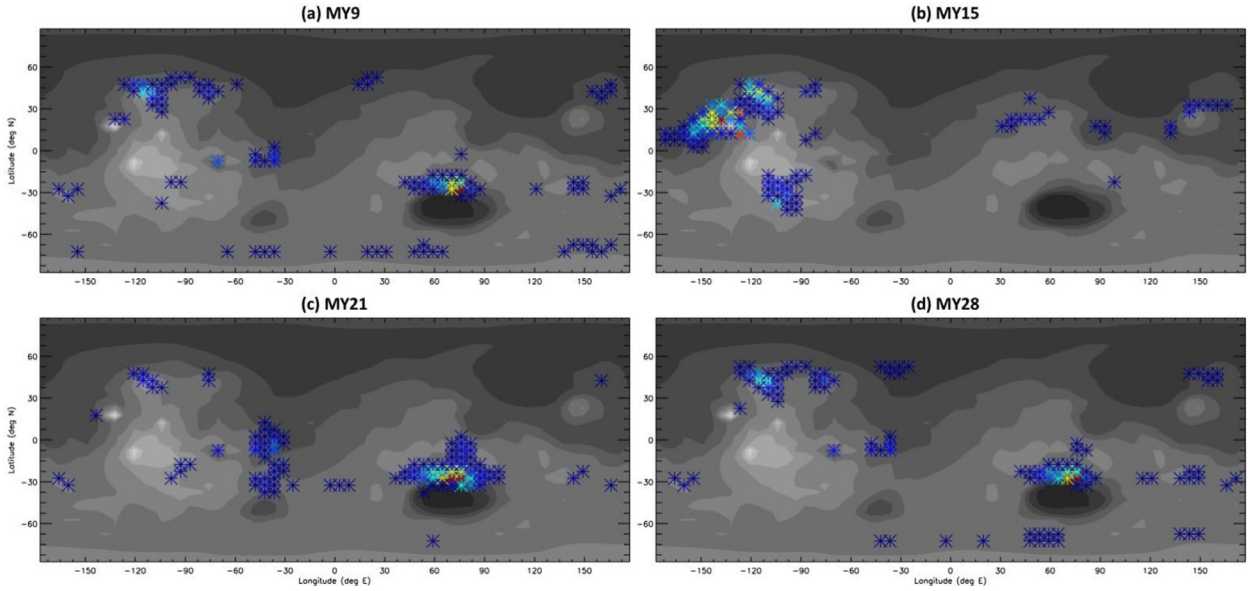

Fig. 7. The top 100 dust source grid points over 20 sols centered around the approximate onset time of the MY 9, 15, 21 and 28 positive polarity GDS, colorized according to the fraction of total dust mass contributed over the entire year. Deep red shows the top ranked grid points, while the dark blue points indicate the lowest contributions within the top 100.

analysis focuses on the period of dust storm onset, as once a dust storm is past this stage the effects of increased dust loading overwhelm the impact of the CTA themselves.

Fig. 7 shows the primary dust lifting regions during onset of the simulated MY 9 possible GDS and the MYs 15, 21 and 28 GDS. The modeled storms in MYs 9, 21 and 28 all begin within the same 10° L_s period surrounding perihelion and have very similar patterns of lifting during onset, despite the simulated storm in MY 21 ultimately growing to be much larger than those in MYs 9 and 28. Peak lifting (shown by red and yellow asterisks) occurs on the northern rim of the Hellas basin, with a secondary peak (shown by lighter blue asterisks) on the northern flank of Alba Patera and in Chryse Planitia. The real MY 28 GDS occurred at roughly the L_s predicted, and was observed to begin as a combination of a Chryse flushing storm and a Noachis storm. Lifting in the Noachis region does not occur in the simulated MY 28 GDS, but there is some top-100 grid point lifting in Noachis in MY 21, although it is relatively weak. The lack of Noachis lifting relative to Hellas in unlimited surface dust simulations was discussed in Section 5.1.1.

By contrast, the MY 15 storm - which begins ~40° of L_s earlier than the others, although still 20° of L_s after equinox - shows a distinctly different onset pattern, with peak lifting grid points (red asterisks) occurring on the western flanks of Olympus Mons but with very strong lifting also at grid points to the west of Tharsis and across western Tharsis and Alba Patera. Many of the top-100 source grid points are clustered here, although there is also a large lifting center consisting of grid points with moderately strong lifting rates (blue asterisks) in the Solis/Claritas region in the southern hemisphere, and other lifting clustered on the western side of Syrtis Major and on the flanks of Elysium Mons. Interestingly, although observations were not able to trace the onset of this storm closely, McKim (1999) suggests Solis as the likeliest of the proposed onset locations.

An obvious question is why wind stresses are enhanced in the regions described above in positive polarity years. To answer this, it is necessary to investigate how the CTA interact with the existing circulation in these regions at different times of sol. Sections 5.2.1 and 5.2.2 do so for the MY 15 (early) and MY 28 (perihelion) GDS, respectively.

5.2.1. A closer look at early GDS onset in positive polarity MY 15

Figs. 8–10 show a time sequence of the impact of the CTA on winds, wind stresses and dust lifting leading up to onset of the MY 15 GDS. Each figure shows wind vectors with and without the CTA, the CTA themselves, and the difference in wind vectors, wind stresses, and predicted dust lifting between simulations CTA1 and CON1, averaged over 20 sols at the same local times of sol. Three different 'snapshots' are chosen to show the variation across the surface when the time at −90°E is midnight (Fig. 8), noon (Fig. 9), and 6 p.m. (Fig. 10). (The snapshot for 6 a.m. at −90°E is not included as there was no enhancement in lifting then.) The local time at any longitude, λ °E, is given by the hour at −90°E plus 12 × (λ + 90)/180.

The solid black contours in Figs. 8–10d show where enhanced dust lifting exists in simulation CTA1, which naturally coincides with where the wind stress is strongly enhanced; slight offsets are due to the difference in wind stress being shown at one instant, while the difference in lifting shown is that summed over the previous half hour. Although these figures do not show all local times at all locations, a comparison with Fig. 7b indicates that they capture lifting from most of the key source regions for onset of the simulated MY 15 GDS. Peak enhancements in Figs. 8 and 9 occur to the west of Tharsis (corresponding to ~8 p.m. and 8 a.m. local time), in Claritas (~ −105°E, at ~11 p.m. and 11 a.m. local time), and on the flanks of Olympus Mons (at ~9 p.m. and 9 a.m. local time). In Fig. 9 the enhanced Claritas lifting is stronger and extends east into Solis (at ~noon local time), with lifting also on the flanks of Alba Patera (at ~11 a.m. and 1 p.m. local time). In Fig. 8, there is also enhanced lifting from Syrtis Major (corresponding to ~9 a.m. local time), while in Fig. 10 there is enhanced lifting to the west and east of Elysium Mons (corresponding to ~9–11 a.m. local time).

Before relating lifting to the pattern of near-surface winds, note that wind stress, not wind speed, is key for dust lifting (see Section 2.1.1). Hence dust lifting may occur where there are moderate rather than peak wind speeds, provided the near-surface air density is sufficiently larger in the former region. Focusing first on the lifting to the west of Tharsis in Figs. 8 and 9, the no-CTA wind vector plots (Figs. 8a and 9a) show that this occurs at times when there are already strong meridional winds here. These winds reverse direction between the evening and morning periods and are due to the interaction of the thermal tide with topography, and to the flow being channeled between the Tharsis and Elysium bulges. Crucially, as shown in Figs. 8b and 9b, the CTA in MY 15 at this season are such that they act to reinforce these meridional flows during both periods, since the CTA are largest and toward the

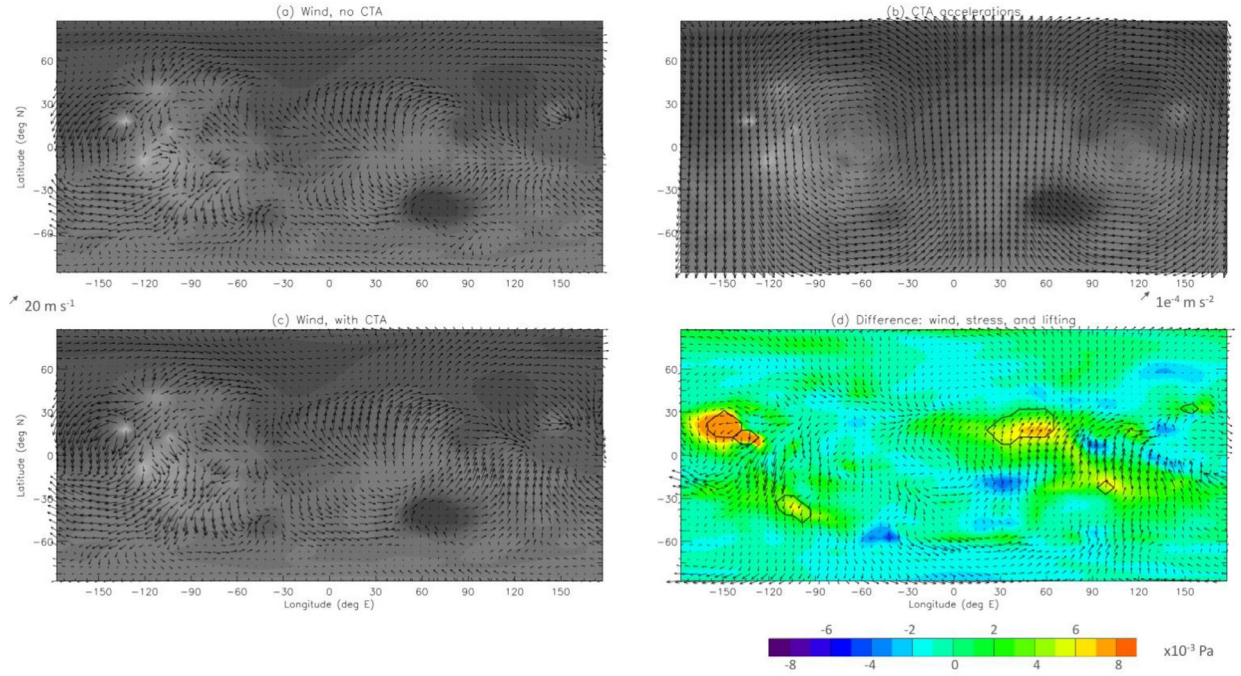

Fig. 8. Lowest level winds in (a) CON1 and (c) CTA1 MY 15, plotted over topography; (b) CTA vectors, plotted over topography; and (d) CTA1 MY 15 - CON1 difference in wind (vectors), wind stress (shaded contours), and dust lifting (line contours), for sols 387–406 (L_s ~ 188–200°) at midnight at −90°E. In (d), black (red) contour lines show where 1×10^{-6} kg m^{-2} h^{-1} more (less) dust was lifted in CTA1 MY 15 than CON1 over the preceding half hour (however, there are no red contour lines for this time of day).

south when the winds are northerlies (i.e. flow from north to south), and equally large but toward the north when the winds are southerlies. This coincidence of intrinsically strong winds and the strong, in-phase CTA accelerations results in the CTA1 winds (Figs. 8c and 9c) being extremely strong here and producing an enhancement in wind stress. Similarly, the southerly tidal winds north of Syrtis Major (centered at ~20°N and 45°E) in Fig. 8a are enhanced by the northward CTA at this time of the morning (Fig. 8b), producing an increase in wind strength, wind stress, and hence dust lifting here (Fig. 8c).

The enhanced lifting in Solis and Claritas around noon is due to the CTA adding a northward component to the existing tidally-produced southerly flow at this time, with both the flow and CTA forcing being reversed when there is Claritas enhancement around midnight. The enhancement on the flanks of Olympus Mons and Alba Patera in the late evening (Fig. 8d) and late morning (Fig. 9d) largely depends on the correspondence between the direction of nighttime downslope or daytime upslope flows and the CTA directions at the same time (Figs. 8b and 9b), although tidal effects are also present. Finally, the already-strong southerlies around Elysium Mons in the late evening (which are likely dominated by tides) are enhanced by the northward CTA in Fig. 10.

At the time of the MY 15 storm (L_s ~ 200° in the model), the

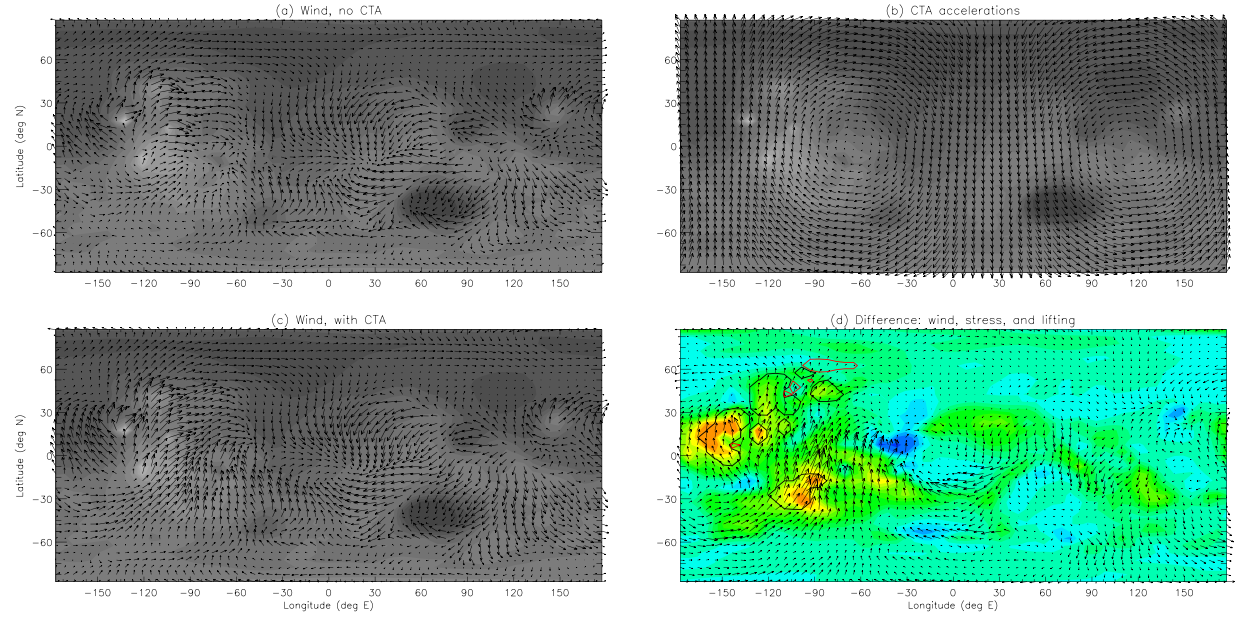

Fig. 9. As in Fig. 8 but for noon at −90°E.

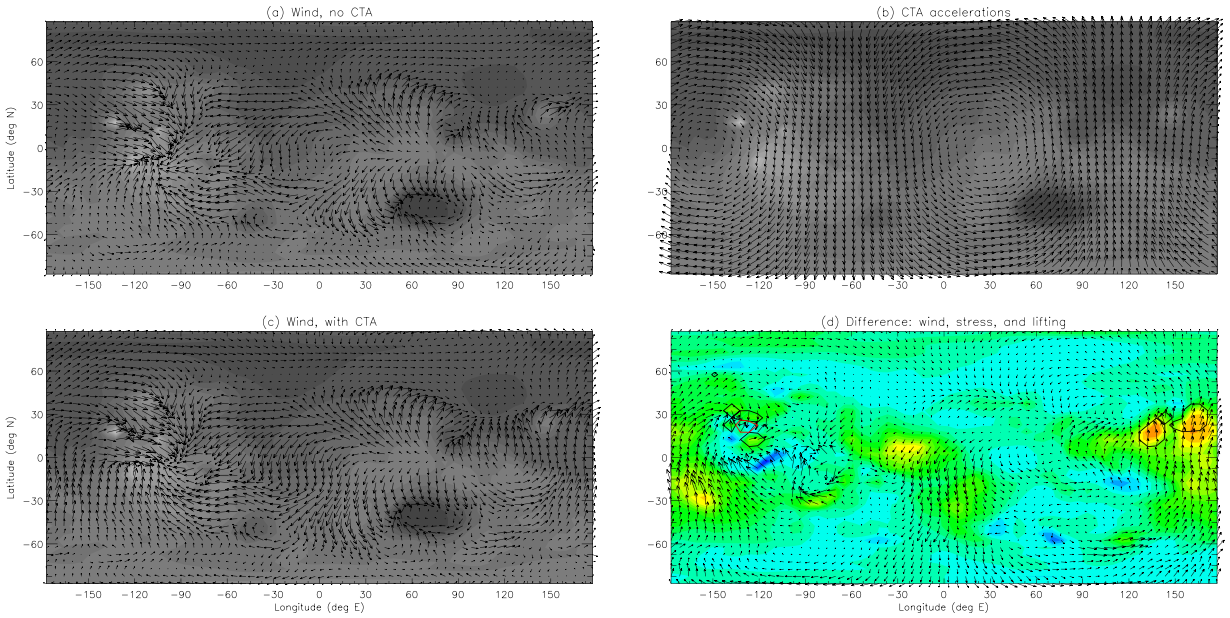

Fig. 10. As in Fig. 8 but for 6 p.m. at −90°E.

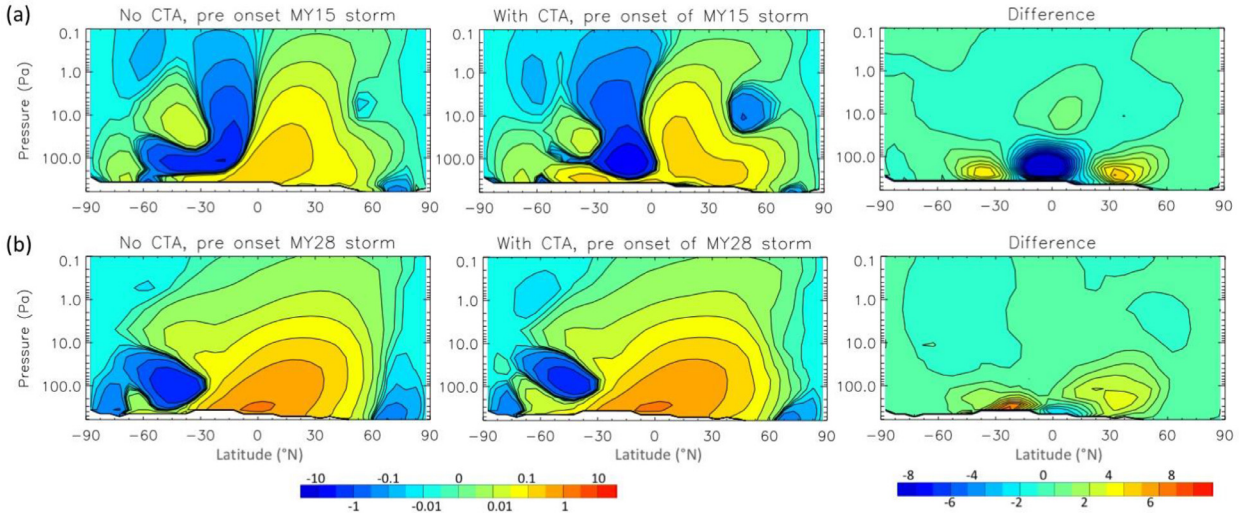

Fig. 11. Mass streamfunctions in the 20 sols leading up to onset of the (a) MY 15 and (b) MY 28 GDS in simulation CON1 (left column) and CTA1 (middle column), with the difference shown in the right column. Units are 1×10^8 kg/s. (Note that the 'Difference' plot contour intervals are one third of those used in Supplementary Material Figs. S2 and S4.).

circulation still consists of an 'equinoctial' double cell rising roughly at the equator that has just begun its transition to the stronger 'solsticial' single cell rising in the south; if anything, the CTA appear to slow this transition, as shown in Fig. 11a. Thus the production of a MY 15 GDS in simulation CTA1 is due to very specific constructive interactions between the CTA and existing flows in regions of already high wind stresses, rather than being a result of the CTA simply enhancing the Hadley circulation overall. These constructive interactions may have been enough to trigger an early GDS in MY 15, but not in other positive polarity MYs with similar waveforms / phases, due to the particularly large peak amplitude of ***dL/dt*** in MY 15 (see Supplementary Material Fig. S1 or Fig. 4 of SM17).

Overall, it appears that the development of the MY 15 and similar early (or late) positive polarity storms may rely on complex combinations of thermal tidal, topographic, and CTA forcing that result in enhanced lifting in very specific regions on Mars. Each pattern of enhancement may only occur for a specific range of CTA positive polarity phases and when the *dL/dt* amplitude is sufficiently strong. Due to the very specific locations of enhanced lifting, the occurrence and nature of such storms are more likely to be sensitive to surface dust availability and/or other physical processes that even subtly affect circulation patterns (such as water cycle coupling). They likely also rely on the amplitude of ***dL/dt*** being larger than normal at the time of GDS onset.

5.2.2. A closer look at perihelion GDS onset in positive polarity MY 28

Figs. 12–15 show the predicted wind, wind stress, and dust lifting changes due to the CTA over a 20-sol period containing onset of the MY 28 GDS at $L_s \sim 245°$, at four times of sol. Unlike the MY 15 onset period, the wind enhancements in regions of enhanced dust lifting now show far less correlation with the sense (or strength) of the CTA vectors at the same time of sol. This is especially true in the southern hemisphere, which is where the bulk of the storm lifting occurs (the northern hemisphere enhancements in lifting, shown by black contours, are roughly balanced by northern hemisphere decreases in lifting relative to CON1, shown by red contours). For example, dust lifting on the northern rim of the Hellas basin at ~11 a.m. (Fig. 12) and ~4 a.m.

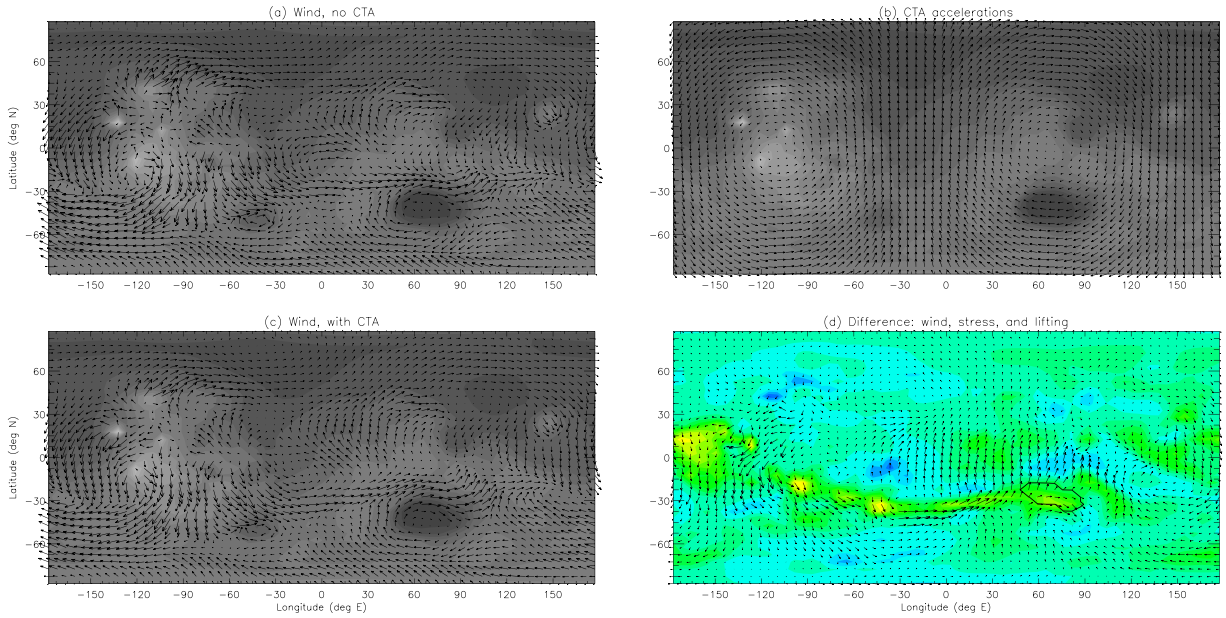

Fig. 12. As in Fig. 8 but for MY 28 sols 467–485 (L_s ~ 238–251°).

(Fig. 15) local time corresponds to periods when the CTA accelerations are in a completely different direction to the wind vector enhancements produced: in both cases there are enhanced westerly winds, whereas the CTA point respectively westward and northward at these times. The CTA are in the same direction as the enhanced wind vectors at ~10°S, −15°E in Fig. 15, but this is when the CTA magnitudes are weak, thus it seems unlikely that they are the direct cause.

This suggests that, unlike the results found for the earlier GDS in MY 15, the perihelion GDS are more likely triggered by the CTA having caused a general strengthening of the global Hadley circulation. At perihelion, the latter corresponds to a single, broad cell rising in the southern hemisphere (see Fig. 11b) with near-surface return flow from north to south across the equator, but note that this is only true in a zonal mean sense. Indeed, the tide so dominates the flows associated with this Hadley cell at some times of sol that it completely overwhelms and reverses them, resulting in the strong time of day dependence in e.g. Fig. 13a or c, where the typical Hadley-related (i.e. monsoonal) surface winds are only apparent in the 6 p.m. to 6 a.m. locations (90°E through 180°E to −90°E), but are flowing 'the wrong way' for the monsoonal flow over the other hemisphere of the planet. Overall, however, there are zonal mean northerlies across the equator, which are turned by Coriolis accelerations as they move further into the southern hemisphere to produce strong zonal mean westerlies in a band from ~15 to 45°S. The wind vector enhancements produced in simulation CTA1 reflect a strengthening of those westerlies, and this also explains the increased wind stresses and dust lifting that largely peaks in the same latitude range and is seen in all four Figures; additional wind stress and lifting enhancements are also produced in the 'western boundary current' region to the east of Tharsis, where the meridional return flow of the Hadley circulation is strongest (Fig. 15).

What is causing this overall enhancement in the Hadley circulation? Figs. 12–15 show that the CTA have their strongest southward

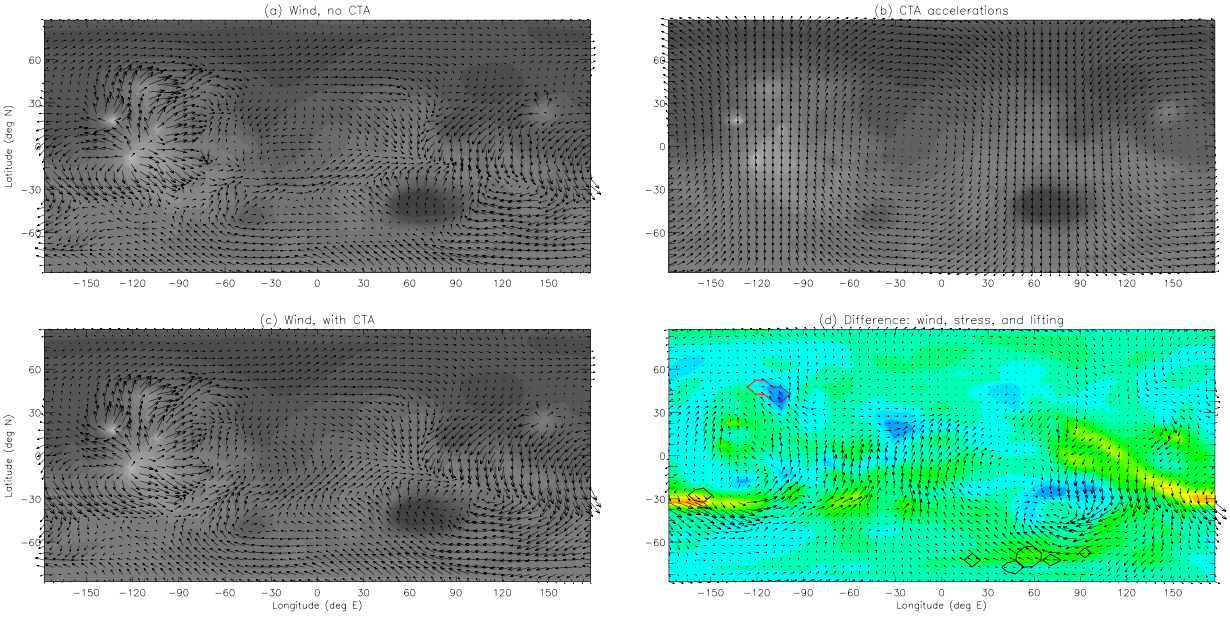

Fig. 13. As in Fig. 12 but for 6 a.m. at −90°E.

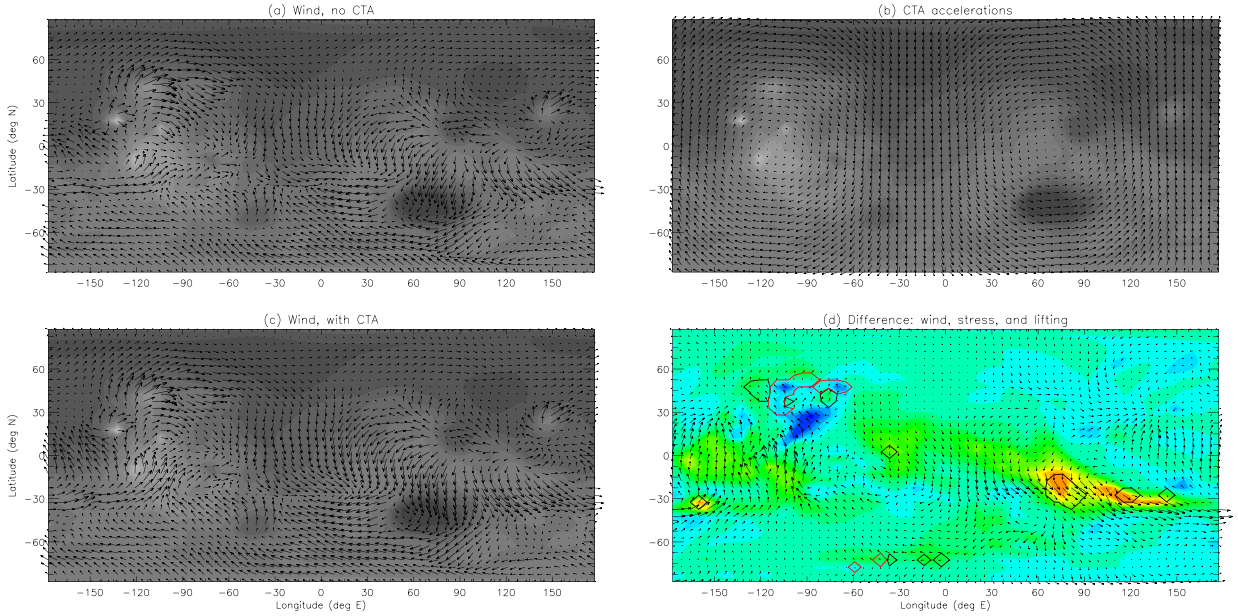

Fig. 14. As in Fig. 12 but for noon at −90°E.

magnitudes in the late afternoon (centered at ~4:30 p.m.) in MY 28 at this time of year. In other words, the CTA reinforce the near-surface branch of the Hadley circulation (which consists of zonal mean northerlies at low-to-mid-latitudes) during a time of strong heating and convection, i.e., at the time when the Hadley circulation is being strongly driven. This can be seen by looking at the wind vector differences (which are generally southward) at the longitudes corresponding to late afternoon local time in Figs. 12–15d (e.g. −130°E to about −70°E in Fig. 14d). Conversely, the CTA have their strongest northward magnitudes in the early morning (centered at ~4:30 a.m.), hence oppose the global mean Hadley circulation during periods of zero (or weak) solar heating. This can similarly be seen as generally northward wind vector differences at early morning longitudes in Figs. 12–15d. The slight phase lag between the time of day modulation of the Hadley cell by the thermal tide and when the CTA constructively interferes with both may be a result of the CTA acting uniformly in direction over the whole atmosphere, whereas the Hadley circulation and the thermal tides have significant changes in direction with altitude.

As suggested by MS17, the CTA in MY 28 therefore appear to reinforce the near-surface, large-scale meridional circulation during the afternoon (when the associated winds are generally stronger) and to weaken it overnight, producing a strengthening of the existing Hadley circulation at this time of year. That in turn produces greater wind stresses at locations where strong wind stress is associated with the Hadley circulation - most notably in western boundary current flows, in the ‘zonal collar’ of strong westerlies, and especially on the northern Hellas rim where those westerlies reinforce nighttime downslope flows into the basin. And in turn, this results in enhancements of dust lifting in those already high-wind-stress locations. This presumably extends to all other GDS occurring around perihelion in positive polarity years, since by definition the CTA phasing will be similar close to perihelion in all such years.

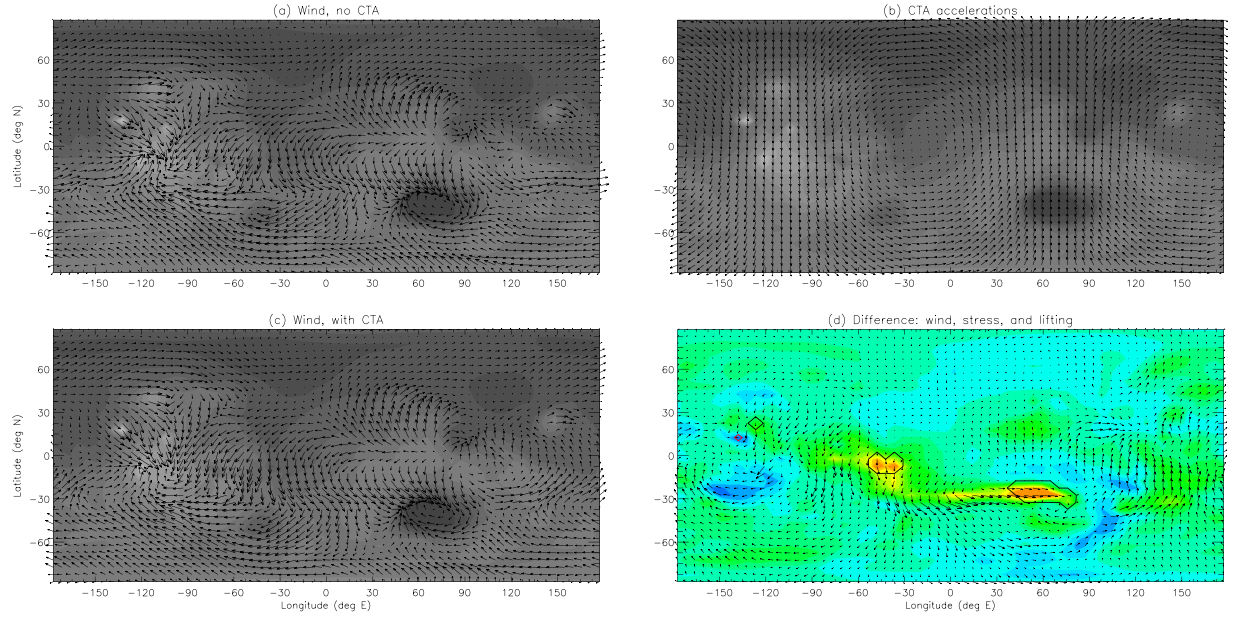

Fig. 15. As in Fig. 12 but for 6 p.m. at −90°E.

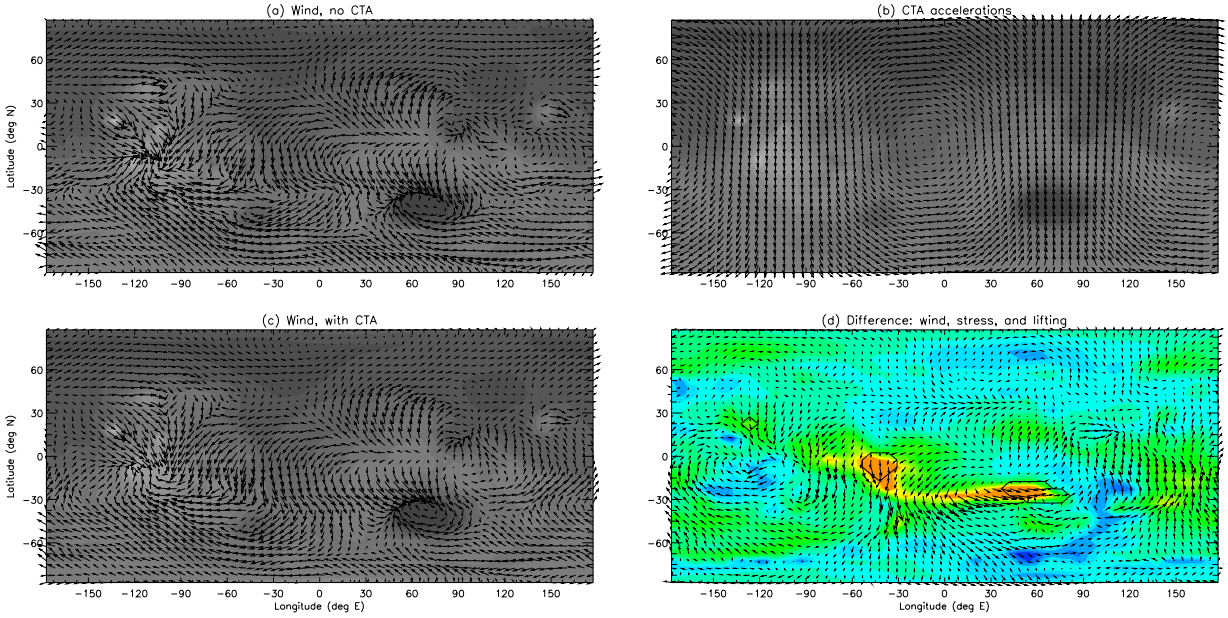

Fig. 16. As in Fig. 10 but pre-onset of the simulated MY 1 GDS (sols 467–471, L_s ~ 238–240°).

5.3. Difference between positive polarity MYs 1 and 27

MYs 1 and 27 have very similar phases with respect to perihelion (Table 2), but while a GDS was observed and a possible GDS is predicted by the model shortly before perihelion in MY 1, no major storm was observed or is predicted in MY 27. As shown in Fig. S1 (and Fig. 4 of SM17), the peak amplitude of $\boldsymbol{dL/dt}$, which occurs before perihelion, is about 80% larger in MY 1 than in MY 27. The result of this difference in forcing amplitude on winds, wind stresses, and dust lifting is shown by comparing Figs. 16 and 17, which show a snapshot of results at 6 p.m. at −90°E in MYs 1 and 27, respectively; snapshots at other times demonstrate similar differences between the two MYs, hence are not included.

The orientations of the CTA forcings at all locations are very similar between MYs 1 and 27 (Figs. 16 and 17b), as are those of the wind vector differences with the CTA forcings included (Figs. 16 and 7d). As in the ~perihelion GDS positive polarity years discussed in Section 5.2.2, the CTA appear to have strengthened the Hadley-related return flow, with enhanced northerlies to the east of Tharsis and westerlies in a southern zonal collar. However, the magnitude of the CTA forcings and consequent wind response is much smaller in MY 27, hence the wind stress threshold is not exceeded even with the CTA included, and there is no enhanced lifting in those regions. These results again demonstrate the importance of the amplitude as well as the phase of the CTA forcing.

6. Predicted GDS activity in upcoming Mars years based on infinite dust simulations

As shown in Fig. 4, simulation CTA1 predicts a GDS in the current storm season (MY 34) at L_s ~ 245° and a smaller GDS in MY35 at L_s ~ 260° (see also Supplementary Material Fig. S4 for the predicted

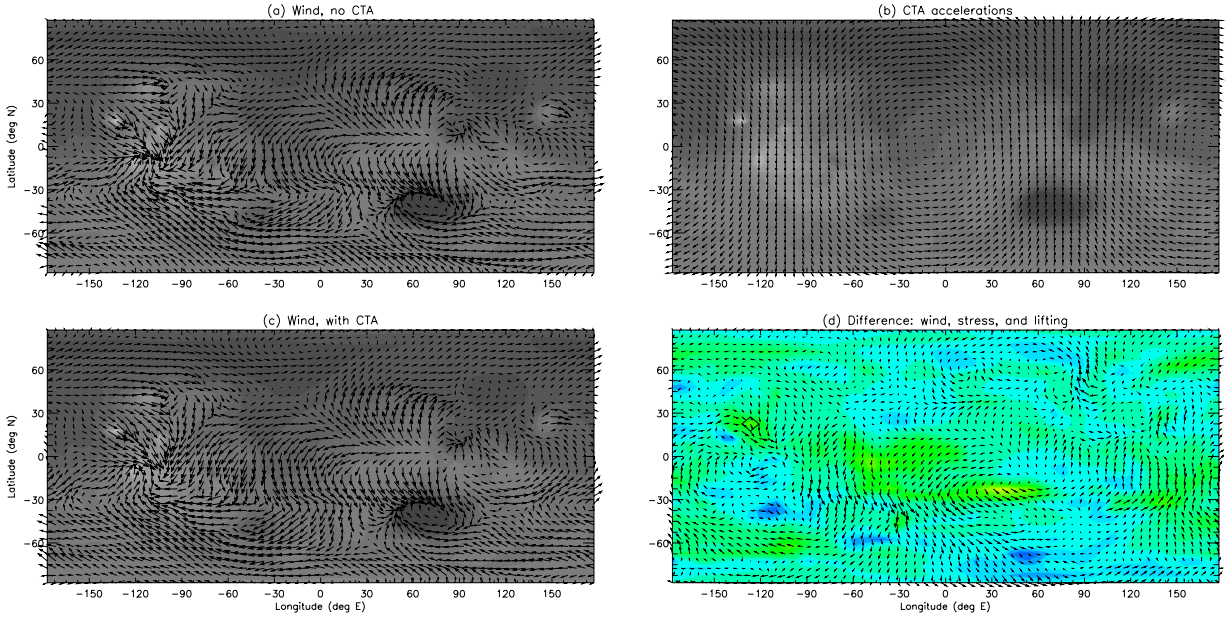

Fig. 17. As in Fig. 16 but for the same period in MY 27.

impact of these GDS on the circulation). These MYs correspond to Earth years 2018 and 2020, respectively. In particular, MY34 has the same polarity (positive) and a similar phase of ***dL/dt*** at perihelion to the GDS that occurred in MY21, which was predicted well by the model (see Table 2). By comparison, MY 35 is a transitional polarity year, in which no GDS have been observed, and simulation CTA1 incorrectly predicts such a GDS in MY 17, which has a rather similar ***dL/dt*** waveform to that of MY 35 (see Fig. S1 or SM17, Figs. 2 and 7). Thus the prediction of a GDS in MY 35 should likely be taken with more caution. Finally, if the GDS in MY 17 was prevented by scouring of source regions during prior year positive polarity GDS (Section 4.2), then a GDS in MY 35 may become even less likely if one occurs in MY 34.

No GDS are predicted in MYs 36–38. However, MYs 37 and 38 are negative polarity years, for which the model does not predict two early GDS (in MYs 12 and 25) that were observed, possibly due to the lack of water cycle coupling. The ***dL/dt*** waveform for MY 38 in particular is very similar to that for MY 25, in which a large equinoctial GDS was observed, thus it seems likely that - if the CTA are controlling GDS activity - a large GDS may be expected to occur around equinox in MY 38. The phase of the ***dL/dt*** waveform in MY 37 is between that of MYs 25 and 24 (in which no GDS was predicted or observed), and MY 37 has a similar amplitude at perihelion to that in MY 24, thus it is difficult to make any estimate for MY 37.

7. Conclusions

A postulated orbit-spin coupling on Mars produces additional 'coupling term accelerations' (CTA) that impact the circulation, and hence surface wind stresses, provided that the 'coupling efficiency coefficient', *c*, is sufficiently large. Because the CTA vary on cycles that are asynchronous with the seasonal cycle of solar forcing, they could potentially dominate year-to-year variations in dust lifting during the dust storm season. This paper presents the first GCM simulations of fully interactive dust cycle modeling with the CTA included, enabling the onset time and location(s), size, and duration of dust storms to be spontaneously and self-consistently predicted.

The results show that including the CTA is sufficient to introduce interannual variability of GDS occurrence and timing into a simulation that would otherwise be very repeatable from year to year. Further, the simulated storm seasons are strongly controlled by the phasing and amplitude of the orbit-spin coupling; results are very robust to small changes in initial conditions, and simulations with modified lifting parameters resemble each other in terms of the onset time and relative size of storms produced in the same MY.

Including the CTA also dramatically improves the model's skill at predicting the observed type of storm season, with the simulation that most closely matches observations (CTA1) clearly wrong in only 4 out of 22 well-observed years. In one of those years (MY17) a GDS is predicted when one was not observed, and in the other three years (MYs 10, 12 and 25) a GDS is not predicted when one was observed.

No GDS are observed in transitional polarity years (in which the CTA amplitude near perihelion is relatively small), which supports the argument that the CTA contribute to producing interannual variability in GDS on Mars. A simulated GDS in MY 17 is incorrectly predicted in a transitional year, although a real storm on Mars in this year may possibly have been prevented by scouring of dust from source regions during a potential series of large GDS in prior years.

Most storm seasons are predicted correctly in positive polarity years, when the CTA amplitudes are large and the sign of the ***dL/dt*** waveform is positive around perihelion. The timing of GDS is also largely captured correctly. For storms with onset near perihelion (e.g. MYs 9, 21 and 28), the CTA's constructive interference with ~daytime tidal flows appears to reinforce the solsticial near-surface Hadley circulation flows, enhancing lifting within a band of strong southern hemisphere westerlies and in strong cross-equatorial flows. In the present model, lifting during storm onset peaks in northern Hellas, due to constructive interference with nighttime downslope flows into the Hellas basin, and in Chryse. In finite surface dust simulations, it is plausible that the primary source regions would shift from northern Hellas to other, less rapidly dust-depleted locations within the zonal collar of enhanced westerlies, such as the Noachis lifting predicted by NR15. This might produce similar storms with slightly different source locations, may better match the observed MY 28 storm which began as a combination of lifting in Noachis and Chryse, and may explain why the model is so successful at predicting positive polarity ~perihelion storms, despite the likely unrealistic assumption that surface dust is available everywhere.

Conversely, an early positive polarity year storm in MY 15 appears to occur mostly due to the especially large magnitude CTA during that storm season, constructively interfering with existing strong tidal-topographic circulations at certain locations and local times of sol. In such cases, storm onset may be more sensitive to specific locations having surface dust present. A more realistic surface dust map would provide a more realistic distribution of primary dust sources (no longer mapping simply to locations of peak wind stress), and may change the onset timing, regions, and occurrence of predicted GDS, although strong predicted lifting in Solis appears consistent with observations. A more realistic finite dust map may potentially be needed to produce the late-season MY 10 GDS, by producing primary dust sources at the time of year when the MY 10 ***dL/dt*** amplitude peaks, with a ~perihelion storm in MY 10 not possible due to the low amplitude of ***dL/dt*** earlier in the storm season.

MYs 1 and 27 have similar ***dL/dt*** waveform shapes, but the amplitude is much larger in MY 1 (when major storms are both observed and predicted) than in MY 27 (when no major storm is observed or predicted). This is due to the enhanced wind stresses being much smaller in MY 27 hence insufficient to raise dust, despite having a very similar pattern, and emphasizes the importance of the amplitude, rather than just the sign and phase, of the CTA forcing.

None of the GDS observed in negative polarity years are predicted by the model. Ignoring the second GDS observed in MY 12 (as it may have been triggered by the first), the non-prediction of the L_s~204° GDS in MY 12 and equinoctial GDS in MY 25 may be due to some vital circulation element operating early in the storm season that could be enhanced by the CTA but is missing from the model. Given that early storm season GDS have typically not been predicted in Mars models unless they include a coupled water cycle (Kahre et al., 2012; Lee et al., 2018), this suggests that future work should involve examining the CTA impact in different polarity years when running the coupled dust-water package within MarsWRF (Lee et al., 2018).

Looking to the future, a GDS is predicted in MY 34 (2018) and a smaller GDS in MY 35 (2020), both near perihelion. The shape of the ***dL/dt*** waveform in MY 34 is very similar to that of MY21, in which a GDS was correctly predicted by the model. By comparison, MY 35 is a transitional polarity year, in which no GDS have been observed, and simulation CTA1 incorrectly predicts such a GDS in MY 17, which has a rather similar ***dL/dt*** waveform to that of MY 35 and may have been prevented by dust source depletion in a previous GDS year. Thus the prediction of a GDS in MY 35 should be treated with caution, especially if a large GDS occurs in MY 34. No GDS are predicted in MYs 36–38; however, MYs 37 and 38 are negative polarity years, for which the model does not predicts the two GDS that were observed. The ***dL/dt*** waveform for MY38 is very similar to that for MY 25, in which a large GDS was observed, thus it seems likely that a large GDS will occur around equinox in MY 38 (2026) if the CTA do control GDS activity.

Acknowledgments and data

The authors gratefully acknowledge the support of NASA's Solar System Workings Program (Proposal # 14-SSW14_2-0354). This investigation has benefited from discussions with members of the Mars Climate Sounder Science Team at JPL. A portion of the research was

carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. We are also grateful for the constructive criticism and valuable suggestions made by reviewers Michael Smith and Mike Wolff.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.icarus.2018.07.023.

References

Basu, S., Richardson, M.I., Wilson, R.J., 2004. Simulation of the Martian dust cycle with the GFDL Mars GCM. J. Geophys. Res. 109, E11006,. https://doi.org/10.1029/2004JE002243.

Basu, S., Wilson, R.J., Richardson, M.I., Ingersoll, A.P., 2006. Simulation of spontaneous and variable global dust storms with the GFDL Mars GCM. J. Geophys. Res. 111, E09004,. https://doi.org/10.1029/2005JE002660.

Briegleb, B.P., 1992. Delta-Eddington approximation for solar radiation in the NCAR community climate model. J. Geophys. Res. 97, 7603–7612.

Cantor, B.A., 2007. MOC observations of the 2001 Mars planet-encircling dust storm. Icarus 186, 60–96. https://doi.org/10.1016/j.icarus.2006.08.019.

Clancy, R.T., Lee, S.W., 1991. A new look at dust and clouds in the Mars atmosphere - analysis of emission-phase-function sequences from global Viking IRTM observations. Icarus 93 (1), 135–158.

Clancy, R.T., Lee, S.W., Gladstone, G.R., McMillan, W.W., Rousch, T., 1995. A new model for Mars atmospheric dust based upon analysis of ultraviolet through infrared observations from Mariner 9, Viking, and Phobos. J. Geophys. Res. 100 (E3), 5251–5264.

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

Garvin, J.B., Frawley, J.J., Abshire, J.B., 1999. Vertical roughness of Mars from the Mars Orbiter Laser Altimeter. Geophys. Res. Lett. 26, 381–384.

Haberle, R.M., Leovy, C.B., Pollack, J.B., 1982. Some effects of global dust storms on the atmospheric circulation of Mars. Icarus 50, 322.

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

Hong, S.-Y., Pan, H.-L., 1996. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. Mon. Weather Rev. 124, 2322–2339.

Hourdin, F., 1992. A new representation of the absorption by the CO2 15-lm band for a Martian general circulation model. J. Geophys. Res. 97, 18319–18335.

Jimenez, P.A., Dudhia, J., 2012. Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. J. App. Met. Clim. 51, 300–316.

Kahre, M.A., Murphy, J.R., Haberle, R.M., 2006. Modeling the Martian dust cycle and surface dust reservoirs with the NASA Ames general circulation model. J. Geophys. Res. 111, E06008,. https://doi.org/10.1029/2005JE002588.

Kahre, M.A., Hollingsworth, J.L., Haberle, R.M., 2012. Simulating Mars' dust cycle with a Mars general circulation model: effects of water ice cloud formation on dust lifting strength and seasonality. In: Comparative Climatology of Terrestrial Planets Meeting, , 8062 LPI Contribution No. 1675.

Kawamura, R., Study of Sand Movement by Wind Translated (1965) *as University of California Hydraulics Engineering Laboratory Report HEL 2–8 Berkeley*, 1951.

Kok, J.F., Parteli, E.J.R., Michaels, T.I., Karam, D.B., 2012. The physics of wind-blown sand and dust. Rep. Prog. Phys. 75. https://doi.org/10.1088/0034-4885/75/10/106901.

Lee, C., Richardson, M.I., Newman, C.E., Mischna, M.A., 2018. The sensitivity of solsticial pauses to atmospheric ice and dust in the MarsWRF General Circulation Model. Icarus 23–34. https://doi.org/10.1016/j.icarus.2018.03.019.

Liu, J., Richardson, M.I., Wilson, R.J., 2003. An assessment of the global, seasonal, and interannual spacecraft record of Martian climate in the thermal infrared. J. Geophys. Res. 108 (E8), 5089. https://doi.org/10.1029/2002JE001921.

McKim, R., Telescopic Martian Dust Storms, 1999. A Narrative and Catalogue. Mem. Br. Astron. Assoc. 44 ISBN 0-902749-13-7.

Mischna, M.A., Shirley, J.H., 2017. Numerical Modeling of orbit-spin coupling accelerations in a Mars general circulation model: implications for global dust storm activity. Planet. Space Sci. 141, 45–72. https://doi.org/10.1016/j.pss.2017.04.003.

Montabone, L., Forget, F., Millour, E., Wilson, R.J., Lewis, S.R., Cantor, B., Kass, D., Kleinböhl, A., Lemmon, M.T., Smith, M.D., Wolff, M.J., 2015. Eight-year climatology of dust optical depth on Mars. Icarus,. https://doi.org/10.1016/j.icarus.2014.12.034.

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

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, 344–358. https://doi.org/10.1016/j.icarus.2012.12.003.

Murphy, J.R., Toon, O.B., Haberle, R.M., Pollack, J.B., 1990. Numerical simulations of the decay of Martian global dust storms. J. Geophys. Res. 95 (B9), 14629–14648. https://doi.org/10.1029/JB095iB09p14629.

Newman, C.E., Lewis, S.R., P.L., Read, Forget, F., 2002a. Modeling the Martian dust cycle. 1: Representations of dust transport processes. J. Geophys. Res. (Planets) 107 (E12), 5123,. https://doi.org/10.1029/2002JE001920.

Newman, C.E., Lewis, S.R., Read, P.L., Forget, F., 2002b. Modeling the Martian dust cycle. 2: multi-annual radiatively active dust transport simulations. J. Geophys. Res. (Planets) 107 (E12), 5124,. https://doi.org/10.1029/2002JE001910.

Newman, C.E., Richardson, M.I., 2015. The impact of surface dust source exhaustion on the Martian dust cycle, dust storms and interannual variability, as simulated by the MarsWRF general circulation model. Icarus 257, 47–87. https://doi.org/10.1016/j.icarus.2015.03.030.

Pankine, A.A., Ingersoll, A.P., 2002. Interannual variability of Martian global dust storms simulations with a low-order model of the general circulation. Icarus 155, 299–323. https://doi.org/10.1006/icar.2001.6757.

Pankine, A.A., Ingersoll, A.P., 2004. Interannual variability of Mars global dust storms: an example of self-organized criticality? Icarus 170, 514–518. https://doi.org/10.1016/j.icarus.2004.04.006.

Powers, J.G., et al., 2017. The weather research and forecasting model: overview, system efforts, and future directions. Bull. Am. Meteor. Soc. 98, 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1.

Renno, N.O., Burkett, M.L., Larkin, M.P., 1998. A simple thermodynamical theory for dust devils. J. Atmos. Sci. 55, 3244–3252.

Richardson, M.I., Toigo, A.D., Newman, C.E., 2007. PlanetWRF: a general purpose, local to global numerical model for planetary atmospheric and climate dynamics. J. Geophys. Res. (Planets) 112, E09001,. https://doi.org/10.1029/2006JE002825.

Shirley, J.H., 2015. Solar system dynamics and global-scale dust storms on Mars. Icarus 252, 128–144. https://doi.org/10.1016/j.icarus.2014.09.038.

Shirley, J.H., 2017. Orbit-spin coupling and the circulation of the Martian atmosphere. Planet. Space Sci. 141, 1–16. https://doi.org/10.1016/j.pss.2017.04.006.

Shirley, J.H., Mischna, M.A., 2017. Orbit-spin coupling and the interannual variability of global-scale dust storm occurrence on Mars. Planet. Space Sci. 139, 37–50. https://doi.org/10.1016/j.pss.2017.01.001.

Shirley, J.H., Newman, C.E., Richardson, M.I., Mischna, M.A., 2019. Replication of the historic record of Martian global dust storm occurrence in an atmospheric general circulation model. Icarus 317, 197–208. https://doi.org/10.1016/j.icarus.2018.07.024.

Strausberg, M.J., Wang, H., Richardson, M.I., Ewald, S.P., Toigo, A.D., 2005. Observations of the initiation and evolution of the 2001 Mars global dust storm. J. Geophys. Res. 110, E02006. https://doi.org/10.1029/ 2004JE002361.

Szwast, M.A., Richardson, M.I., Vasavada, A.R., 2006. Surface dust redistribution on Mars as observed by the Mars Global Surveyor and Viking orbiters. J. Geophys. Res. 111, E11008. https://doi.org/10.1029/ 2005JE002485.

Toigo, A.D., Richardson, M.I., 2002. A mesoscale model for the Martian atmosphere. J. Geophys. Res. 107 (E7), 5049. https://doi.org/10.1029/ 2000JE001489.

Toigo, A.D., Lee, C., Newman, C.E., Richardson, M.I., 2012. The impact of resolution on the dynamics of the martian global atmosphere: Varying resolution studies with the MarsWRF GCM. Icarus 221 (1), 276–288. https://doi.org/10.1016/j.icarus.2012.07.020.

Wang, H., Richardson, M.I., 2015. The origin, evolution, and trajectory of large dust storms on Mars during Mars years 24–30 (1999–2011). Icarus. http://dx.doi.org/10.1016/j.icarus.2013.10.033.

White, B.R., 1979. Soil transport by wind on Mars. J. Geophys. Res. 84, 4643–4651.

Wilson, R.J., Hamilton, K.P., 1996. Comprehensive model simulation of thermal tides in the Martian atmosphere. J. Atmos. Sci. 53, 1290–1326.