The lower thermosphere during the northern hemisphere winter of 2009: A modeling study using high-altitude data assimilation products in WACCM-X

Fabrizio Sassi,1 Han-Li Liu,2 J. Ma,3 and Rolando R. Garcia4

Received 8 April 2013; revised 24 June 2013; accepted 5 July 2013; published 29 August 2013.

[1] We present numerical simulations using the Whole Atmosphere Community Climate Model, extended version, constrained below 90 km by a combination of NASA’s Modern Era Retrospective Analysis for Research and Applications and the U.S. Navy’s Operational Global Atmospheric Prediction System - Advanced Level Physics High Altitude assimilation products. The period examined is January and February 2009, when a large stratospheric warming occurred on 24 January 2009, with anomalous circulation persisting for several weeks after the event. In this study, we focus on the dynamical response of the lower thermosphere up to 200 km. We find evidence of migrating and nonmigrating tides, Rossby and Rossby-gravity modes, and Kelvin waves, whose amplitudes appear to be modulated at the times leading and following the stratospheric warming. While the Rossby, Rossby-gravity, and Kelvin modes are rapidly dissipated in the lower thermosphere (above 110 km), the tides maintain substantial amplitude throughout the thermosphere, but their vertical structure becomes external above about 120–150 km. Most waves identified in the simulations decrease in amplitude in the thermosphere, indicating remote forcing from below and strong dissipation by molecular diffusion at high altitudes; however, the amplitude of the migrating DW1 tide increases in the thermosphere suggesting in situ forcing. We show that the amplitude of the tides (such as the DW1) changes as the background wind alters the vorticity in the tropics, which broadens or narrows the tropical waveguide. Our results also suggest that fast Rossby normal modes (periods ≤ 10 days) are excited by instability of the zonal-mean wind distribution following the stratospheric warming.


1. Introduction

[2] During Northern Hemisphere winter, the climatology of the stratosphere and lower mesosphere is characterized by strong eastward zonal winds. This circulation is profoundly affected by the occasional amplification of slowly westward propagating, planetary-scale Rossby waves, which dissipate and transfer westward momentum to the zonal circulation, resulting in a deceleration of the zonal-mean zonal flow. On occasion, a full reversal of the zonal-mean zonal

wind at 60 N and 10 hPa (~30 km) occurs; such an event is known as a major sudden stratospheric warming (SSW) [see Andrews et al., 1987]. Major SSWs occur on average about once every 2–3 years [Charlton and Polvani, 2007], although many general circulation models generate SSWs much less frequently [Charlton et al., 2007]. Charlton and Polvani [2007] distinguish between splitting SSWs, in which the polar vortex is divided into two separate entities, and displacement SSWs (the more frequent case), wherein the vortex is displaced from its normal circumpolar position. An example of a splitting vortex event is the 2009 major SSW, which is characterized by a wave-2 deformation and split. In addition to major SSWs, minor stratospheric warming events, characterized by a weakening of the stratospheric circulation at 10 hPa and often a wind reversal in the upper stratosphere and lower mesosphere, occur much more frequently than the SSWs, almost every winter.

[3] The occurrence of an SSW is known to produce significant changes in the dynamical and thermal structure of the upper mesosphere and lower thermosphere (UMLT) [Liu and Roble, 2002; Siskind et al., 2007, 2010], as well as in its composition [Randall et al., 2006]. This influence may
not be limited to the neutral behavior of the UMLT but may extend into the ionosphere, as shown by the correlation between the enhancement of the vertical drift velocity of the plasma and the meteorological behavior the stratosphere [Chau et al., 2009]. Various observational [Goncharenko et al., 2010; Pedatella and Forbes, 2010] and modeling studies [Liu et al., 2010a; Fuller-Rowell et al., 2010] have suggested that a possible mechanism for coupling the stratosphere with the thermosphere/ionosphere is the amplification of planetary-scale waves in the UMLT, where they can interact with tropical tides.

[4] The major SSW on 24 January 2009 is one of the largest splitting SSWs on record [Manney et al., 2009; Harada et al., 2010]. The thermal perturbation over the Arctic polar cap at 10 hPa at the peak of the event exceeded 70 K. In the lower stratosphere, the perturbation to the background atmosphere lasted several weeks after the onset of the SSW [e.g., Manney et al., 2009; Sassi et al., 2012]. Because of the sheer magnitude of this event, numerous studies have been conducted to document its impact on the UMLT [e.g., Kurihara et al., 2010; Yuan et al., 2012]. Furthermore, Wang et al. [2011] showed that the amplitude of the semidiurnal migrating tide (SW2) was notably reduced during the SSW. Although these studies suggest a causal connection between the lower and the upper atmosphere during a SSW, a satisfactory mechanism to explain the connection is lacking to date.

[5] As noted above, large-scale, quasi-stationary waves are amplified in the stratosphere during a SSW. Under favorable circumstances, this behavior extends to the UMLT [Liu and Roble, 2002]. Moreover, traveling, planetary-scale Rossby waves are also present with substantial amplitude in the UMLT [Sassi et al., 2012, and references therein]. The fastest of these waves are less likely to encounter critical layers at lower altitudes and can propagate into the lower thermosphere, where they could potentially influence the E-region [Meyer, 1999]. In fact, Fraser [1977] showed that an oscillation in ionospheric scatter of period 5–6 days appears to be coherent with similar temperature oscillations in the stratosphere, in both winter and summer. While it is not possible to evince the zonal behavior from local measurements, as in Fraser’s [1977] analysis, a 5 day oscillation is consistent with the theoretical period of the gravest Rossby normal modes of wave number 1 (1,1) and of wave number 2 (2,1). As shown in Sassi et al. [2012], both these modes are known to appear with significant amplitude during Northern Hemisphere (NH) winters in the U.S. Navy’s Operational Global Atmospheric Prediction System - Advanced Level Physics High Altitude (NOGAPS-ALPHA) data assimilation products.

[6] Another mechanism that might influence tidal behavior in the thermosphere is changes to the forcing in the lower atmosphere. Recently, Goncharenko et al. [2012] showed that, starting at times immediately following the SSW of 24 January 2009, zonal-mean tropical ozone in the upper stratosphere increased for a prolonged period. This increase follows from the strengthening of the zonal-mean meridional circulation during a SSW, which results in enhanced upwelling and adiabatic cooling in the Tropics, and downwelling and warming at high latitudes [see, e.g., Garcia, 1987]. Ozone is in photochemical equilibrium in the upper stratosphere, such that colder temperatures lead to an increase in ozone, and vice versa. Assuming that the increase in ozone results in a commensurate increase of tidal forcing, one expects tidal forcing to be enhanced for an extended period, with a resulting increase in the amplitude of the tide in the thermosphere following the SSW [see also Liu, 2013].

[7] In this study, we analyze in detail the response of the lower thermosphere to the SSW of January 2009 in a simulation made with the Whole Atmosphere Community Climate Model, eXtended version (WACCM-X), wherein meteorological fields below 90 km are constrained by data assimilation products. We refer to this configuration as specified dynamics WACCM-X, or SD-WACCM-X. This configuration is different from the WAM model utilized by Wang et al. [2011] in that there is no data ingestion during the simulation. Instead, the atmospheric specifications are used to constrain the model meteorology by adding tendency terms to the momentum and thermodynamic equations that bring the resultant winds and temperature very close to those in the data assimilation products.

[8] The paper is organized as follows: In section 2, we describe the global model; section 3 describes the data assimilation products; section 4 shows a validation of the global simulation; section 5 presents the results of a spectral analysis of waves throughout the UMLT; section 6 discusses the spatial structures of nontidal waves, and section 7 examines the tides; section 8 illustrates the temporal behavior and compares to Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) observations; section 9 attempts to relate the temporal behavior of tides with changes of the background vorticity; a summary and conclusions are presented in section 10.

2. SD-WACCM-X

[9] The WACCM is a build option of the National Center for Atmospheric Research (NCAR) Community Earth System Model version 1 (CESM1) and can be used in place of the standard atmospheric model. In its standard configuration, WACCM has 66 vertical levels from the ground to 5.9 × 10^{-6} hPa (~140 km geometric height). The vertical resolution is identical to the standard atmospheric model up to 100 hPa, but it varies above that level [see Garcia et al., 2007]. The model can be run at a variety of horizontal resolutions for different applications: the standard horizontal resolution of 1.9° × 2.5° (longitude by latitude) is used here. A discussion of the model climate can be found in (Marsh et al., Climate change from 1850–2005 simulated in CESM1 (WACCM). J. Climate, 2013 doi:10.1175/JCLI-D-12-00558.1).

[10] WACCM can be configured into an extended version (WACCM-X) that incorporates all the features mentioned above with the top boundary located in the upper thermosphere (2.5 × 10^{-9} hPa, or ~500 km altitude). A detailed description of WACCM-X is given in Liu et al. [2010b]. In WACCM-X, the ion drag and Joule heating are calculated according to Dickinson et al. [1981] and Roble et al. [1982], respectively. Major species diffusion is included using the formulation of Dickinson et al. [1981]. It should be noted that WACCM-X does not include a treatment of plasma transport. With the vertical extension of the model and the additional physics of the thermosphere, the time step for integration of the physics is 300 s.
[11] WACCM-X can be used to simulate specific events by constraining the model using data analysis products or observations. A detailed illustration of this modeling configuration can be found in Marsh [2011]. We refer to this particular model configuration as specified dynamics WACCM-X (SD-WACCM-X). An SD-WACCM-X integration begins from a suitable set of initial conditions, as in the regular configuration, but moves forward in time with the solution relaxed at each model time step towards a set of time-varying specified dynamical fields over part of the model domain. The specified dynamical fields are obtained from a linear combination of the Modern Era Retrospective Analysis for Research and Applications (MERRA) and NOGAPS-ALPHA DAS products, which both produce atmospheric specification every 6 h. This relaxation procedure is carried out by applying a tendency to the zonal wind, the meridional wind, and the temperature that is proportional to the difference between the modeled and the specified fields. The relaxation time scale is 10 model time steps, which was chosen because it yields model dynamical fields that match very closely the specified fields between the ground and 80 km; between 80 and 90 km, the relaxation becomes progressively weaker until, above 90 km, the model is unconstrained.

3. Data Assimilation Products

[12] Here, we review briefly the data assimilation products starting with the NOGAPS-ALPHA. More complete details of NOGAPS-ALPHA are provided in Eckermann et al. [2009] and Hoppe et al. [2008], and references therein. Just like the operational NOGAPS at the Fleet Numerical Meteorology and Oceanography Center, NOGAPS-ALPHA runs in a continuous 6 hourly forecast assimilation update cycle. However, it extends to much higher altitudes than NOGAPS (up to about 92 km) and includes assimilation of temperature, water vapor and ozone from the Microwave Limb Sounder on NASA’s Aura satellite and temperature profiles from the SABER instrument on NASA’s Thermosphere Ionosphere Mesosphere Energetics and Dynamics (TIMED) satellite. Numerous studies have been conducted that validate the NOGAPS-ALPHA data products. Stevens et al. [2010] show how well the NOGAPS-ALPHA analysis captures the diurnal variations in the MLT; Eckermann et al. [2009] show that the analysis captures the basic characteristics of the quasi-five day wave and migrating diurnal tides as a function of latitude and height; McCormack et al. [2010] discuss the interaction between the quasi-two day wave (QTDW) and the diurnal tide; Sassi et al. [2012] evaluate the transient planetary-scale behavior in recent NH winters. As with other 3DVAR DAS, NOGAPS-ALPHA and MERRA can be affected by noise generated by unbalanced initial conditions [e.g., Sankey et al., 2007; Nezlin et al., 2009], which may ultimately affect the tidal amplitudes. It should be borne in mind that a comparison of the diurnal tide DW1 in MERRA [Sakazaki et al., 2012] against SABER (MERRA does not assimilate SABER data) has revealed, that although somewhat smaller than in observations, the DW1 amplitude is in good agreement with SABER observations in the stratosphere and the lower mesosphere. NOGAPS-ALPHA instead assimilates SABER data, and its tidal amplitudes as well as behavior associated with the 5 day wave have been validated in terms of the effect on polar mesospheric clouds occurrence and distribution [Stevens et al., 2010], and the distribution of water vapor [Eckermann et al., 2009].

[13] SD-WACCM-X uses NASA/MERRA [Rienecker et al., 2011] data assimilation products between the ground and ~50 km [Marsh, 2011]. In this study, we show results that implement a hybrid data set configuration that combines the NOGAPS-ALPHA and MERRA data assimilation products; in doing so, we exploit advantages from both data sets, keeping the energy and near-surface budget constraint in the regular model configuration of SD-WACCM, which uses the MERRA meteorological fields, but switching to the high-level atmospheric specifications from NOGAPS-ALPHA in the stratosphere. While NOGAPS-ALPHA specifications are valid throughout the entire vertical domain, from the ground to 92 km, the standard output from NOGAPS-ALPHA does not include some necessary fields that characterize the surface meteorology and are required in the SD configuration. These include surface sensible heat flux, snow depth, soil moisture, zonal and meridional surface stresses, and surface temperature. These fields are necessary to determine the surface energy budget and constrain the lower atmospheric state to the observed behavior. Similar fields are also calculated by NOGAPS-ALPHA but were not made available in the standard output. Since re-running the entire NOGAPS-ALPHA system is not feasible, we opted for a hybrid system that preserves the energy constraints in the lower atmosphere obtained from MERRA but extends to the upper atmosphere using NOGAPS-ALPHA. More specifically, the hybrid data set uses the MERRA atmospheric specifications below 30 km, merges linearly between 30 and 50 km MERRA and NOGAPS-ALPHA, and uses only NOGAPS-ALPHA atmospheric specifications above 50 km up to 92 km. The resulting configuration has 104 levels from the ground to 3.3 × 10⁻⁹ hPa.

[14] The simulation that is the focus of the present study covers January and February of 2009. Unless explicitly stated otherwise, we analyze hourly output of temperature during this period with the output interpolated to geopotential height levels.

4. Validation of the Hybrid Atmospheric Specifications

[15] In this section, we validate the new hybrid data set, comparing it to MERRA and NOGAPS-ALPHA. Figure 1 compares the zonal-mean zonal wind on 16 January 2009 (top row) and on 24 January 2009 (bottom row) obtained from the MERRA products (left panels) to the NOGAPS-ALPHA products (middle panels). At times prior to the SSW (16 January), both data assimilation products reveal an eastward zonal circulation in the NH (winter) that maximizes at about 1 hPa (~50 km); however, only NOGAPS-ALPHA illustrates the zonal circulation above the stratopause, with the zonal-mean jet tilting equatorward in the mesosphere and decreasing in strength toward the UMLT. The NH jet configuration on 16 January is representative of quiet wintertime conditions. During the following few days, a large wave-2 event leads to the reversal of the zonal circulation in the winter stratosphere. At the peak of the SSW (24 January), the zonal-mean eastward wind in the stratosphere is replaced by westward winds at latitudes poleward of 30°N. Only the NOGAPS-ALPHA data products
are capable of illustrating the vertical structure of the circulation in the mesosphere.

[16] It is clear from Figure 1 that in the common domain (below 50 km), MERRA and NOGAPS-ALPHA produce atmospheric specifications that are very similar, while NOGAPS-ALPHA is capable of defining the atmospheric behavior up to the UMLT. The result of the linear combination of MERRA and NOGAPS-ALPHA is shown on the right-hand side panels of Figure 1.

[17] For verification purposes, a SD-WACCM-X simulation using the hybrid data set is compared to a simulation that instead implements only MERRA. Figure 2 compares the evolution of the zonal-mean zonal wind in the lower mesosphere (65 km; lower panels), and in the upper mesosphere (78 km; top panels): one simulation is constrained by MERRA alone (SD-WACCM-X/MERRA; left-hand side panels), and the second simulation is constrained by the hybrid data set (SD-WACCM-X/HYBRID; right-hand side panels).

[18] As expected, since MERRA atmospheric specifications are used in both data sets up to 50 km, the two model simulations are identical in the stratosphere (not shown), while they start to diverge in the mesosphere, at 65 and 78 km, where SD-WACCM-X/MERRA is unconstrained. SD-WACCM-X/MERRA displays many similarities to SD-WACCM-X/HYBRID in the winter (NH) lower mesosphere (Figure 2, lower panels) during January 2009; note the westward winds that replace the eastward flow during the onset of the SSW. However, at times following the SSW, the polar night jet in the mesosphere does not recover as strongly during February as in SD-WACCM-X/HYBRID. The restoration of a strong and isolated polar night vortex in the mesosphere is crucial for obtaining the mesospheric descent of air at high latitude that brings NOx downward to affect critical stratospheric ozone chemistry during the winter of 2009 [e.g., Manney et al., 2009; Randall et al., 2009; Salmi et al., 2011]. Larger differences are evident in the upper mesosphere (Figure 2, top panels), especially at high latitudes, where SD-WACCM-X/HYBRID shows a much stronger polar vortex in the NH compared to SD-WACCM-X/MERRA. It is likely that the differences between the two simulations illustrated by Figure 2 are associated with limitations of the gravity wave drag scheme that was initially tuned [e.g., Garcia et al., 2007] to reproduce the mean climate, but not specifically the mesoscale gravity wave field during the focus period of this study.

5. Spectral Behavior

[19] Hereafter we concentrate on the analysis of the SD-WACCM-X/HYBRID simulation during the focus period of January and February 2009 in the NH (winter). In order to illustrate the character of the spectrum of waves present in the UMLT, we now study with Figure 3 the eastward- and westward-propagating RMS frequency spectra. Results are shown for the Tropics, subtropics, and middle latitudes, at altitudes between 90 and 120 km, where the amplitude of the waves is large. Notwithstanding the increasing prominence of tidal variability at the higher levels, all the spectra have a typical geophysical “red” character, with generally larger amplitude at low frequencies. The statistical significance of peaks that rise above the red spectrum is evaluated using the methodology described in Sassi et al. [2012]. In brief, power corresponding to a first order auto-regressive spectrum (AR1) is estimated from the auto-correlation and the background white-noise variance of the eddy component of the model’s temperature field [see Wilks, 2006]. If no a priori hypothesis exists for the location of a spectral peak, statistical significance is achieved when a peak exceeds the
estimated AR1 spectrum by an amount proportional to a critical value obtained from the χ² probability distribution that corresponds to the estimated degrees of freedom and the probability, α*, of the null hypothesis. With an a priori hypothesis for the existence of variability at a given spectral location (e.g., migrating/nonmigrating tides, normal modes, etc.), the same level of statistical significance is achieved at a greater value for α* and thus at a lower threshold of the χ² critical value. It is important to bear in mind that the a priori hypothesis is used only in the spectral intervals where enhanced variability is expected from theoretical considerations; outside those intervals, the significance threshold is calculated without the a priori hypothesis. For more details, see the Appendix in Sassi et al. [2012], Wilks [2006], and Fisher [1929]. Figure 3 shows the threshold of significance at the 90% confidence level without an a priori hypothesis (red dashed line), and with an a priori hypothesis (red dotted line) for some of the specific waves that we discuss below.

(A centered, five-point running average has been applied to the raw periodograms to yield the spectra shown in Figure 3.)

The westward spectrum of wave-1 at middle northern latitudes (40°N and 92 km; Figure 3a) has a prominent peak around 0.2 cpd (5 days). Note that the RMS amplitude associated with this spectral location exceeds the threshold indicated by the red dashed line even without an a priori hypothesis. Following Sassi et al. [2012], we associate this spectral peak with the presence of a Doppler-shifted gravest symmetric wave number 1 Rossby normal mode (1,1), the so-called 5 day wave. Similarly, the spectral peak around 1 cpd is associated with the migrating diurnal tide at wave-1 (DW1); RMS amplitude at this spectral location also exceeds the 90% threshold without an a priori hypothesis. A much smaller spectral peak at 2 cpd, which we associate with the semidiurnal, westward-propagating tide (SW1), is also statistically significant without invoking the a priori hypothesis.

[21] Figure 3b illustrates the spectral behavior of westward propagating wave-2 at 40°N and 104 km. A spectral peak is found at frequencies near 0.2 cpd, but it is below the significance level without an a priori hypothesis. If we assume that this peak results from the presence of the gravest symmetric Rossby normal mode of wave number 2 (2,1) [see Salby, 1981; Sassi et al., 2012], use of the a priori hypothesis in the spectral range 0.10–0.25 cpd indicates that the peak is in fact statistically significant at the 90% confidence level (red dotted line). The spectral peak at 2 cpd, on the other hand, is statistically significant without invoking an a priori hypothesis and corresponds to the semidiurnal westward-propagating tides (SW2). Figure 3c shows the westward spectrum of wave-3 at 60°S and 100 km. A spectral peak at 0.5 cpd (2 days) is noticeable and is statistically significant without an a priori hypothesis. In the next section, we examine the spatial structure of this peak and whether it corresponds to the structure of a QTGW.

[22] Considering next the spectral behavior of temperature at 1 N (the model grid point nearest the equator), Figure 3d shows the westward spectrum of wave-1. The migrating diurnal (DW1) and the nonmigrating semidiurnal (SW1)
westward tides emerge clearly as statistically significant peaks over the background spectrum. Similarly, the eastward-propagating wave-1 near the equator (Figure 3e) shows one prominent peak that exceeds the threshold (dashed line) without the need to invoke an a priori hypothesis at the semidiurnal (SE1) frequency. A region of relatively high spectral amplitude at around 0.3 cpd suggests the presence of ultra-fast Kelvin waves (UFKWs) [Salby et al., 1984; Forbes, 2000], and in the next section, we investigate this possibility. These spectral peaks become statistically significant if we use the a priori hypothesis between 0.3 and 0.4 cpd (red dotted line).

Figure 3. Frequency spectra of temperature between 1 January 2009 and 28 February 2009. The black lines are the amplitude spectra; the red dashed lines show the 90% confidence levels with respect to an estimated auto-regressive spectrum without an a priori hypothesis; the red dotted lines show the 90% confidence levels with an a priori hypothesis. Relevant modes are identified with the appropriate acronym. See text for details. (a) Wave-1 westward spectrum at 40 N and 92 km; (b) wave-2 westward at 40 N and 104 km; (c) wave-3 westward at 60 S and 100 km; (d) wave-1 westward at 1 N and 120 km; (e) wave-1 eastward at 1 N and 100 km; (f) wave-2 westward at 30 N and 120 km; (g) wave-2 eastward at 1 N and 120 km; (h) wave-3 eastward at 1 N and 120 km.

[23] The westward spectrum of wave-2 is shown at 30 N in Figure 3f: the migrating SW2 tide emerges prominently at this latitude as the only statistically significant spectral peak. The spectra of eastward propagating waves 2 and 3 at the equator (Figures 3g and 3h) instead show prominent peaks at the diurnal and semidiurnal frequencies. Note the prominent peak at DE3 (Figure 3h), which is due presumably to the wave-3 nonmigrating tide excited by tropospheric deep...
convective heating. The DE3 tide is believed to play a crucial role in the coupling between the thermosphere and the ionospheric plasma [Hagan et al., 2007, 2009]. We note that nonmigrating tides also emerged prominently in calculations made with the NCAR/TIE-GCM, when that model is forced at the lower boundary with NOGAPS-ALPHA data products [Siskind and Drob, 2012]; this confirms that higher harmonics of the tides are present in the NOGAPS-ALPHA assimilation products. Note also the large amplitude of the nonmigrating SE2 peak at 112 km (Figure 3g). The presence of large SE2 and DE3 at the same time and location is a point also the large amplitude of the harmonics of the tides are present in the NOGAPS-ALPHA assimilation products. Note also the large amplitude of the nonmigrating SE2 peak at 112 km (Figure 3g). The presence of large SE2 and DE3 at the same time and location is a potential source of ambiguity when fields are plotted as a function of local time; only a spectral analysis can discriminate between these waves.

6. Structure of Nontidal Waves

[24] Figure 3 illustrates enhancements of spectral amplitude at certain frequencies that are statistically significant with respect to an estimated background red spectrum. While the occurrence of tidal variability (migrating/nonmigrating) at harmonics of the diurnal frequency is to be expected [Hagan and Forbes, 2002] and will be illustrated in the next section, the presence of spectral peaks at other frequencies is also thought to be important [Meyer, 1999; Liu et al., 2010a]. In this section, we investigate the coherent spatial structure of some of these frequencies to determine the nature of the nontidal waves identified above.

[25] Figure 4a shows the amplitude (black contours) and phase (red contours) of the band-passed variability in the spectral range 0.3–0.4 cpd (eastward) at wave number 1 obtained from the squared coherence method of Hayashi [1971]. This spectral band corresponds to the peaks of enhanced variability marked UFKW in Figure 3e. The amplitude is largest at the equator, decreasing almost symmetrically on either side toward the poles. It peaks at 110 km just below 1 K and decreases rapidly above; by 150 km altitude, the amplitude has already decreased below 0.5 K. The amplitude shown in Figure 4a is substantially smaller than the “low forcing” scenario that Chang et al. [2010] have calculated from a TIME-GCM simulation. It should be remembered, however, that the amplitude obtained from the spectral analysis presented here illustrates the mean amplitude during the focus period of January–February 2009; we will show later that the daily values can be substantially larger (see Figure 6). Also shown in Figure 4a are the regions where the variability is coherent (shading), with a confidence of at least 90% [Julian, 1975], relative to the base point (denoted by the green X), which in this case is located at the equator and 100 km. In this case, as in all following calculations, the base point is chosen to be nearly coincident with the largest amplitude. The corresponding phase illustrates a symmetric structure between 100 and 150 km; the implied vertical wavelength in the altitude range 80–130 km is about 50 km. The phase structure and the spatial distribution of the amplitude are consistent with that of a Kelvin wave. Using the Kelvin wave dispersion relationship, the vertical wavelength implies a zonal phase velocity of ~160 m/s for this wave. The spectral characteristics of the thermospheric UFKW wave in SD-WACCM-X may be compared with Salby et al.’s [1984] analysis of satellite observations of Kelvin waves in the mesosphere. These authors showed that Kelvin waves in the upper mesosphere have characteristic phase velocities of ~115–135 m s⁻¹, with a vertical wavelength of about 40 km. However, we expect that UFKW wave variance that reaches the lower thermosphere will be dominated by waves capable of surviving dissipation and encounters with critical levels at lower altitudes; thus, we expect the Kelvin wave spectrum in the lower thermosphere to peak at higher frequency (and hence higher phase velocity and longer vertical wavelength) than the waves observed by [Salby et al., 1984] in the mesosphere. In fact, our UFKW wave is consistent with the thermospheric estimates of Forbes [2000].

[26] Figure 4b illustrates the coherent behavior of the band-passed variability in the range 0.45–0.6 cpd (westward) at wave number 3. This corresponds to the peaks marked QTDW in Figure 3c; in this case, the base point is at 60 S and 100 km. The amplitude is largest in the SH (summer) with coherent behavior that extends down to the stratosphere (not shown) and across the equator into the NH in the UMLT; the phase structure is complex but shows a general antisymmetric behavior about the equator in the altitude region where the amplitude is the largest. All these properties are consistent with the expected characteristics of the QTDW, the Rossby-gravity normal mode of wave number 3 [Salby, 1981; McCormack et al., 2009, 2010].

[27] Figure 4c illustrates the coherent behavior of the band-passed variability between 0.12 and 0.2 cpd (westward) at wave number 1. This band corresponds to the spectral peaks marked (1,1) in Figure 3a. The base point for the calculation of the coherence squared is at 40 N and 90 km. The amplitude shows peak values at middle to high latitudes of both hemispheres. The phase indicates a structure without nodes and overall approximately constant phase in latitude, and very long vertical wavelength (~60–80 km) in the mesosphere and lower thermosphere. The amplitude and phase plots up to about 130 km are consistent with the gravest symmetric Rossby normal mode of wave number 1, also known as the 5 day wave [see Salby, 1981]. While this mode is relatively fast and thus is little affected by lower atmospheric winds, observations suggest that the 5 day wave can be Doppler shifted to other frequencies [Lieberman et al., 2003; Liu et al., 2004; Riggin et al., 2006; Sassi et al., 2012]. The structure above 130 km is latitudinally broad with increasing amplitude in the thermosphere, but it is not coherent with the behavior at lower latitudes. Moreover, the spectral amplitude in the thermosphere is not statistically significant with respect to the local background spectrum (not shown).

[28] Figure 4d illustrates the coherent behavior of the band-passed variability in the range 0.10–0.25 cpd (westward) for wave number 2. This frequency band corresponds to the spectral peaks marked (2,1) in Figure 3b. The base point is at 40 N and 104 km. Once again, the amplitude peaks are located at middle to high latitudes in both hemispheres, with the largest amplitude (~1.25 K) in the NH at about 110 km. The phase shows a uniform structure with no nodes in latitude, consistent with the gravest symmetric Rossby normal mode of wave number 2 [see Salby, 1981; Sassi et al., 2012]. Note that the coherent behavior extends throughout the whole atmosphere in the NH, and it reaches the low latitudes to cross over in the SH between 100 and 120 km altitude. Globally coherent behavior, such as that
illustrated in Figure 4d, would be consistent with Fraser’s [1977] finding of 5 day periodicity in ionospheric scatter.

7. Structure of the Tides

In this section, we illustrate the behavior of some of the most prominent tides identified earlier (Figure 3); a full assessment of the tides (diurnal and semidiurnal, migrating and nonmigrating) produced by the SD-WACCM-X simulation is beyond the scope of the present study.

[30] Figures 5a and 5b show the amplitude (black contours) and phase (red contours) of the band-passed behavior for some of the westward propagating tides: the diurnal wave-1 migrating tide (DW1; Figure 5a), and the semidiurnal wave-2 migrating tide (QTD2; Figure 5b).

---

Figure 4. Amplitude (black contours) and phase (red contours) of the band-passed behavior of (a) the ultra-fast Kelvin wave [frequency range: 0.3–0.4 cpd eastward], (b) the quasi-two day wave [frequency range: 0.4–0.6 cpd westward], (c) the Rossby normal mode (1,1) [frequency range: 0.12–0.20 cpd westward], and (d) the Rossby normal mode (2,1) [frequency range: 0.1–0.25 cpd westward]. Contour interval for the amplitude (black) is 0.1 K in Figures 4a and 4b; 0.25 K in Figures 4c and 4d. The phase contours show the zero line and ±30°, ±60°, and ±90° as solid (positive) and dashed (negative) lines. Shading shows the locations where the band-passed structure is coherent with the base point with a confidence of at least 90%. The base points are shown by the green X symbols.
wave-2 migrating tide (SW2; Figure 5b). The amplitude of DW1 (Figure 5a) increases with altitude toward the summer SH to values in excess of 16 K at 200 km altitude. The increasing amplitude of DW1 in the summer thermosphere is consistent with in situ forcing by extreme ultraviolet (EUV) solar heating [e.g., Forbes, 1995; Solomon and Qian, 2005]. The phase structure (red contours) in the UMLT shows an equatorially trapped wave; in the thermosphere, there is little phase variation in height or latitude. This phase behavior is consistent with a latitudinally narrow and vertically propagating DW1 tide in the UMLT, which becomes a broad external (nonpropagating) wave in the thermosphere [see Chapman and Lindzen, 1970]. The change in modal structure is likely due to a combination of strong molecular dissipation that removes the propagating component of the tide above ~120 km, and the latitudinally broad EUV heating in the thermosphere that projects preferentially on modes with negative equivalent depth, as

Figure 5. As in Figure 4, but for some of the tidal modes: (a) westward wave-1 diurnal, (b) westward wave-2 semidiurnal, (c) eastward wave-2 diurnal, and (d) eastward wave-2 semidiurnal. The band-pass frequency range for the diurnal tides is 0.9 – 1.1 cpd westward or eastward, as appropriate; the band-pass frequency range for the semidiurnal tides is 1.9 – 2.1 cpd westward or eastward. Contour interval is 1 K in Figure 5a, and 0.5 K in Figures 5b, 5c, and 5d.
is the case in the stratosphere with forcing due to UV absorption by ozone.

[31] The amplitude of the SW2 tide (Figure 5b) is less than 1 K below 100 km but increases toward the winter NH, reaching an amplitude in excess of 4 K at 60 N and 200 km. SW2 is different from DW1 because it has a prominent amplitude peak (4 K) at middle to high latitudes in the winter hemisphere, while the amplitude in the summer SH is less than half of that in the winter hemisphere. The phase of SW2 is, by and large, antisymmetric about the equator, with constant phase lines in the thermosphere, illustrating once again an external mode above 120 km.

[32] The amplitude of DE2 (Figure 5c) has a maximum of about 3.5 K at about 120 km at the equator with a broadly equatorially symmetric phase structure (red contours) about the equator, without nodal points in latitude. The amplitude is slightly skewed toward the winter NH at higher altitudes. It should be noted that the amplitude and phase structure of

Figure 6. Temporal behavior of the amplitude of (a) the migrating diurnal tide at 100 km, (b) the migrating semidiurnal tide at 100 km, (c), the migrating diurnal tide at 120 km, (d) the migrating semidiurnal tide at 116 km, (e) the amplitude of the ultra-fast Kelvin mode at 100 km and (f) at 80 km, and (g) the Rossby mode (2,1) at 100 km. Contour interval is 2 K in each panel. Shading is used for emphasis to highlight the regions where the amplitude is greater than 6 K. The vertical dashed line denotes the SSW. The frequency bands used to synthesize the wave amplitude are identical to those in Figures 4 and 5.
DE3 (not shown) are very similar to DE2. SE2 (Figure 5d) has two amplitude maxima in the NH and SH around 120 km in the subtropics, and an equatorial maximum of smaller magnitude between 110 and 140 km. While the amplitude keeps increasing with altitude in the NH, the amplitude in the summer SH decreases above about 150 km. The phase is antisymmetric about the equator in the UMLT with a propagating character below 120 km, and external mode behavior above.

[33] The migrating (DW1 and SW2) and nonmigrating tides (DE2 and SE2) show some common aspects and some differences. A property that is common to these waves is the global character of the coherence, as is expected from oscillations that are forced by solar heating, which is itself globally coherent. A distinguishing feature is the increasing amplitude of a nonpropagating DW1 in the thermosphere, which is presumably the result of prominent, westward-propagating EUV heating in the thermosphere, as noted above. All the tides are also vertically propagating up to about 120–140 km and become external modes above this altitude.

8. Temporal Behavior

[34] Figure 6 shows the temporal evolution of the amplitude of some of the waves described in Figures 4 and 5 (shading in all panels of Figure 6 highlights the locations where the amplitude exceeds 6 K). At 100 km, the migrating diurnal tide (DW1, Figure 6a) has the largest amplitude at the equator, with a pronounced minimum between day 20 and day 30, and an overall peak amplitude of 12 K, which is attained around day 50. The migrating semidiurnal tide at the same altitude (SW2; Figure 6b) has a node at the equator (cf. Figure 5b) with amplitude maxima in the subtropics; most prominently in the northern subtropics, the amplitude of SW2 has a minimum between day 20 and day 25 and peaks in early February in the SH. It is interesting to compare the amplitude of these two tides between 100 km and 120 km (Figures 6c and 6d). The overall evolution of the amplitude of the two migrating tides is similar at 100 km; it is characterized by a minimum near the time of the SSW (indicated by the dashed vertical line), and larger amplitude at following times. At 120 km, the amplitude is not obviously influenced by the occurrence of the SSW, especially in the case of DW1, but both waves display pronounced amplitude vacillations.

[35] The amplitude of the UFKW (Figure 6e) at 100 km peaks at the equator and decreases very rapidly towards midlatitudes. The UFKW shows at least three amplification events: one in early January, a second one around the end of January, and a final event in the second half of February, around day 48. The forcing for the UFKW has been shown to be associated with deep tropical convection [Holton, 1972, 1973; Salby and Garcia, 1987; Garcia and Salby, 1987; Forbes, 2000], and thus the amplification events in Figure 6e may reflect behavior that extends to lower levels; in fact, the amplitude of the UFKW at 80 km (Figure 6f) shows amplification at times leading those at 100 km, suggestive of vertical propagation. (The vertical group velocity of a wave-1 Kelvin wave with a vertical wavelength of about 50 km is ~0.2 m s⁻¹ which results in a propagation time of 1.5–2 days between 80 and 100 km. This is consistent with the delay of the amplitude peaks at the higher level compared to the lower one.) However, if the amplitude at 100 km were dictated only by lower level forcing and vertical propagation, then it should scale according to the density stratification between the two levels. However, while the amplitude at 80 km near the end of January is larger than in February, the opposite is the case at 100 km. Thus, while it is likely that the UFKW has its source in the lower atmosphere, some other processes (such as dissipation) contribute to modulating its amplitude as it travels throughout the UMLT.

[36] We investigate next the temporal behavior of the Rossby normal modes (1,1) (2,1), which show substantial and coherent amplitude over a deep atmospheric layer (Figures 4c and 4d). Figures 6g and 6h show the amplitude of these waves at 100 km during January and February. The Rossby (1,1) wave (Figure 6g) shows amplification right at the time of the SSW, and decreasing amplitude during February 2009. The Rossby (2,1) wave (Figure 6h) shows similar amplification at the time of the SSW reaching larger amplitude than (1,1); several episodes of amplification in the NH continue during the focus period but with decreasing amplitude in February. Note the hemispheric asymmetry between the two Rossby modes, such that the amplitude of the (1,1) mode is largest in the SH while the (2,1) mode is largest in the NH. Recently, [Chandran et al., 2013] have shown that an amplification of wave-1 around 6.5 days occurs in the upper mesosphere at times following the SSW of January 2012. They argue, based upon the reversal of the background potential vorticity gradient and the behavior of the wave activity flux around 6.5 days, that the amplification is due to instability of the zonal-mean zonal wind. Meyer and Forbes [1997] also showed an amplification of a Rossby mode at period of 6.5 day and wave number 1, which they also associated with instability of the mesospheric zonal-mean wind during August and September of 1993. Thus, it is likely that the amplification of the Rossby modes during January and February 2009 that we document in the present simulation is likewise associated with instabilities of the zonal-mean wind.

[37] The temporal evolution illustrated in Figure 6 describes the variability of several specific waves during the NH winter of 2009, illustrating behavior at high latitudes (Rossby modes) and at low latitudes (equatorially trapped modes). We show in Figure 7 the corresponding variability obtained from SABER temperature observations at ~100 km, mapped synoptically following the method of Salby [1982a, 1982b] with specific modifications for SABER sampling [see Garcia et al., 2005], and spectrally synthesized to match closely the frequency intervals used for the WACCM results of Figure 6. It should be borne in mind that the TIMED satellite must yaw approximately every 61 days, so that the SABER detector never faces the sun. The observations shown in Figure 7 were made during the northward-looking yaw period that began on 12 January 2009. The southernmost latitude observed by SABER during this period is ~52 S; however, local time is poorly sampled at this latitude over a single yaw period, such that results can be unreliable [Salby, 1982a; Sassi and Salby, 1998]. For this reason, we show SABER observations to 48 S only.

[38] As was the case in WACCM, the (1,1) and (2,1) modes observed by SABER (Figures 7a and 7b) show...
amplification at the time of the SSW (black dashed line), such that the overall temporal evolution in the WACCM simulation (Figures 6g and 6h) corresponds closely to the satellite observations. The amplitude of the (2,1) mode is matched relatively well by the WACCM simulation, while the amplitude of the (1,1) mode is smaller in the model (6 K vs. 9 K).

Figures 7c and 7d illustrate SABER observations of two equatorially trapped waves, the UFKW and the DW1. The UFKW (Figure 7d) shows amplification right after the SSW, minimum amplitude in early February, and a final amplification in the second half of February 2009. Not only is the general character of the temporal evolution matched closely in the model simulation (cf., Figure 6e), but so is the actual wave amplitude. Finally, the DW1 tide (Figure 7c) in SABER observations shows a prominent amplitude minimum at the time of the SSW followed by an amplification beginning at the end of January. The maximum amplitude is reached near the end of February at about 16 K. The temporal evolution is matched remarkably well by the model results, although the peak amplitude is smaller in WACCM (~12 K). This is likely due to the fact that the vertical structure of the tide (whose vertical wavelength is ~20–25 km in the upper mesosphere; cf. Figure 5a) is sampled only by few grid points in the current model grid (Δz ~ 3.5 km); tests with doubled vertical resolution yield a somewhat larger amplitude (H.-L. Liu, 2013, personal communication).

9. Temporal Behavior of the Migrating Tides

Figures 6 and 7 suggest a temporal association between the behavior of equatorially trapped waves in the UMLT and the high-latitude behavior. Based on analysis of other winters [Meyer and Forbes, 1997; Chandran et al., 2013], we have argued above that the amplification of the Rossby modes at the time of the SSW is likely due to baroclinic instability of the background zonal-mean wind. As regards the UFKW, the SSW also marks the beginning of a period of rapid amplification, but this appears to be related to the behavior at lower levels (Figures 6e and 6f). However, the sharp amplitude decrease of the tides at the time of the SSW remains to be explained.

Classical tidal theory assumes constant winds [Chapman and Lindzen, 1970]. This approximation explains well the gross tidal features because tides are fast propagating waves (c ~ 460 ms⁻¹ at the equator), which are not affected directly by typical wind changes in the Earth’s atmosphere. On the other hand, winds could affect the tides indirectly by changing the absolute zonal-mean vorticity (f’ = f − Ω/C₀) of the background atmosphere and thus the effective planetary rotation frequency, which ultimately determines the meridional width of the tropical waveguide for these waves. This hypothesis is due to McLandress [2002], who showed in a numerical model that the seasonal variation of wind shear in the upper atmosphere appears to explain the observed semianual variation of tidal amplitude in the UMLT.
westward meridional shear (i.e., eastward meridional shear in the NH (i.e., gradient of the zonally averaged zonal wind. From (1), an
show instead how
in the subtropics of each hemisphere. In Figure 8, we
average of the amplitude of the migrating tides DW1
period of the model simulation and compare to the tropical
mesosphere (70–90 km) and SW2 (blue solid curve), in the upper
R

\[ R = \frac{f - \pi_v}{f}, \]  

(1)

where \( f \) is the Coriolis parameter, and \( \pi_v \) is the meridional gradient of the zonally averaged zonal wind. From (1), an
eastward meridional shear in the NH (i.e., \( \pi_v > 0 \)) reduces
\( R \), which broadens the tropical waveguide; conversely, a
westward meridional shear (i.e., \( \pi_v > 0 \)) in the NH increases
\( R \), which narrows the tropical waveguide; see McLandress

[42] Here, we follow McLandress [2002] and calculate the
ratio of total vorticity and planetary vorticity:

Figure 8. Temporal evolution during January and February
2009 of the vorticity ratio \( R \) in the upper mesosphere between
70 and 90 km. Shading shows the min-max range of \( R \) on
each day between 70 and 90 km. The red solid thick line shows the
daily mean of the tropical average of \( R \) for each
day. The black (blue) solid thick line shows the corresponding
amplitude of DW1 (SW2), at 100 km, synthesized over
the same frequency band as in Figure 6, and averaged
between 30 S and 30 N. The tidal amplitudes are shown on
the right-hand side ordinate. The vertical dashed line
between 30 S and 30 N in both hemispheres to obtain an
increase of some of these modes peaks in the lower thermosphere
amplitude of the DW1 tide (black solid line in Figure 8). The
wave amplitudes decrease to a minimum around 27 January
2009 for DW1 and 20 January for SW2. Afterwards, the
amplitude of DW1 begins to increase throughout the remainder
of the focus period as \( R \) decreases. The amplitude of SW2
(blue solid line in Figure 8) shows similar behavior, except
that the minimum amplitude precedes the SSW by few days,
and the amplitude in February undergoes vacillations that are
absent in \( R \).

[45] By and large, the temporal evolution of \( R \) in the upper
mesosphere is opposite to that of the amplitude of DW1: as \( R 
\)
increases (decreases), the amplitude of DW1 decreases
(increases). Day-to-day changes in tidal amplitude do not
follow changes of \( R \), which suggests that the day-to-day
variability of DW1 is not influenced by changes of the back-
ground vorticity. However, the low-frequency, intraseasonal
behavior is similar to the seasonal variations of the tides
illustrated by McLandress [2002], where an increase of \( R 
\)
resulted in lower tidal amplitude, and vice versa.

10. Summary and Conclusions

[46] We have presented the results of a “whole atmo-
sphere” simulation, from the ground to the upper thermo-
sphere, where the meteorology below 90 km is constrained
by data assimilation products. The focus period is the NH
winter of 2009, specifically the months of January and
February, which show a large and long lasting SSW. The
goal of the study is to document the changes to the dynamics
of the UMLT during this focus period and its relation to the
dynamical behavior occurring in the stratosphere.

[47] The data products used to constrain WACCM from
the ground to 90 km are obtained by a combination of
NASA/MERRA and Navy/NOGAPS-ALPHA data assimila-
tion products. We have shown that in the common domain
(below 50 km), the two data products are essentially indistin-
guishable. Between 50 and 90 km, we use only the Navy’s
data products to constrain the model meteorology. Above
90 km, WACCM is free running.

[48] A close inspection of the spectra of the planetary
scales at different latitudes and altitudes reveals a rich spec-
trum of waves, from tidal (migrating and nonmigrating)
waves to Rossby and Rossby-gravity normal modes, and
UFKWs. All these modes become external (nonpropagating)
above about 120 km because of molecular viscosity. For the
same reason, the spatial structures reveal that the amplitude
of some of these modes peaks in the lower thermosphere
and decreases rapidly above 120–150 km due to dissipation.

[49] The temporal behavior of tides near 100 km shows
reduced amplitude at times before and during the SSW. We
have presented circumstantial evidence that, at least for the
migrating diurnal tide (DW1), this behavior is related to
changes of the zonal-mean vorticity in the Tropics, which
may affect the effective “\( e \)-folding” width of these waves.
This mechanism was suggested by McLandress [2002] to
explain the seasonal behavior of DW1 in the upper
mesosphere. For the first time, we have applied here the same
reasoning to explain its intraseasonal behavior. We find that
the behavior described by McLandress [2002] on seasonal
time scales holds also on intraseasonal scales in the UMLT.

[50] It remains to be explained how temporal changes of
vorticity can affect the diurnal migrating tide. The work of

8966
McLandress [2002] and our own results here support the relationship illustrated above; whether that happens because the forcing projects onto modes of different meridional width as the background winds change, or because the vertical propagation itself is affected (for example via changes of the vertical group velocity) cannot be addressed in the context of the present model simulation and requires further investigation.

It should be borne in mind that our results do not preclude the possibility that other processes might also contribute to modulating or amplifying the tides, in particular, those processes involving enhanced forcing of tides in the lower atmosphere [Goncharenko et al., 2012], or the presence of resonant triads in the thermosphere [Wang et al., 2011]. In addition, the results of this study do not rule out other mechanisms of interaction between the tides and other normal modes (like the QTDW), such as phase locking [Walterscheid and Vincent, 1996; McCormack et al., 2010]. It will be in the goal of a future study to determine their relative importance.

The simulation analyzed in this study is an example of state of the art modeling from the ground to the upper atmosphere that exploits for the first time observations and data assimilation products up to the lower thermosphere. This simulation does not include an ionosphere, and thus the conclusions we draw apply strictly to the neutral thermosphere. Future applications of this methodology will need to include the ionosphere in order to identify how the changes of the neutral dynamics impact the plasma in the E and F regions. Although spatially and temporally varying wind shear has been recognized for a long time to have potentially large effects on the vertical propagation of tides [Chapman and Lindzen, 1970; Forbes, 2000], our result show that further investigations need to be carried out to better understand how the vertical propagation of equatorially trapped waves is coupled to the background meteorology.

Acknowledgments. FS and HL were partially supported by NASA grant NNX12AT21G. RG was partially supported by NASA grant X09AJ83G. FS acknowledges also the support of 6.1 funding from the Office of Naval Research under the ISES (Integrating Sun and Earth Systems) project. The National Center for Atmospheric Research is sponsored by the U.S. National Science Foundation. This work was supported in part by a grant of computer time from the DOD High Performance Computing Modernization Program at the US Navy DOD Supercomputing Resource Center (NAVO).

References


SASSI ET AL.: THE LOWER THERMOSPHERE AND THE 2009 SSW


