

Aeronomy of terrestrial upper atmospheres

DAVID E. SISKIND AND STEPHEN W. BOUGHER

As one moves upward in altitude in a planetary atmosphere, several important changes in composition and structure are apparent. Most notably, as a consequence of hydrostatic equilibrium, the gas density decreases, i.e. the air becomes "thinner". The decrease in density is exponential and governed by a scale height which typically varies in the range of about 5–50 km. Concomitant with this density decrease, the atmosphere becomes increasingly transparent to shorter wavelengths in the solar (or stellar, for exoplanets) spectrum. These shorter wavelengths, typically in the mid, far, and eventually, extreme ultraviolet (MUV, FUV, and EUV respectively), can first dissociate and then at higher altitudes, ionize, various gases in the atmosphere and this alters the composition of the atmosphere. Furthermore, with decreasing density, the frequency of collisions between atmospheric molecules decreases to the point where bulk motions such as turbulence are no longer able to mix the atmosphere. Instead, molecular diffusion becomes the more rapid process and this also leads to a composition change whereby the lighter constituents, typically atomic species such as atomic oxygen, diffuse upwards more rapidly than their heavier counterparts such as O₂, N₂, or CO₂. The region where the atmosphere is well mixed is known as the homosphere; the region where diffusive separation dominates is known as the heterosphere. Although this transition takes place over a range of altitudes, it is common to define some reference boundary altitude known as the homopause to divide the two regimes.

A second transition occurs in the thermal structure. The increased exposure of the atmosphere to energetic UV radiation and the greater dominance of atomic species which are typically inefficient infrared radiators means that the temperature increases markedly with increasing altitude. The altitude regime where the temperature exhibits a large positive temperature gradient is known as the thermosphere. Because that portion of the solar UV spectrum which forms the thermosphere is more variable than the longer wavelengths which heat lower altitudes,

thermospheres respond much more strongly to solar variability than atmospheres at lower altitudes. While the thermosphere and heterosphere are closely related and generally overlap in altitude, the physical processes which govern their variability are not precisely identical. In this chapter we will discuss both “spheres”, while lumping the two together under the more general label of “upper atmosphere”. The lower atmosphere (or often further divided into the lower and middle atmosphere) consists of the troposphere, stratosphere, and mesosphere. A classic reference which summarizes this is Chamberlain and Hunten (1987). The middle atmosphere, and its possible variability with changing climate, is discussed in Ch. 16 of Vol. III.

The study of the upper atmosphere is important for two reasons. First, it is where satellites orbit. Our ability to track and forecast satellite trajectories and quantify satellite lifetimes depends critically on an understanding of how the density of the thermosphere varies (Emmert and Picone, 2010; Pilinski *et al.*, 2013). Further, and of relevance for this chapter, planetary upper atmospheres co-exist with a weakly ionized plasma known as the ionosphere. As described in Ch. 13 of Vol. III, the ionization fraction varies from one part in a million at the homopause to about 1% at the base of the exosphere (defined loosely as where the molecular mean free path exceeds a scale height, about 600 km altitude; cf. Eq. (7.4)). That chapter also discusses the nomenclature for the various ionospheric regions (e.g., D, E, and F) and makes the point that among terrestrial planets, the Earth’s ionosphere is unique. This is because at its peak, it is dominated by an atomic ion (O^+) which recombines fairly slowly (an F_2 ionosphere). By contrast, the Martian and Venusian ionospheres are dominated by molecular ions which recombine rapidly and thus are of the chemically controlled F_1 type. This means that the terrestrial ionosphere will present dynamical variability that is not seen on the other planets.

Our emphasis is on the inner terrestrial planets, with specific emphasis on Earth and Mars. Chapter 8 of this volume provides an overview of thermospheres and ionospheres of the giant planets. Both Earth and Mars are being studied by NASA remote sensing missions. In the case of Earth, this includes the ongoing data from the NASA Thermosphere Ionosphere Mesosphere Energetics and Dynamics (TIMED) mission and upcoming data from the ICON (Ionospheric Connection) and GOLD (Global Observations of Limb and Disk) missions. For Mars, the MAVEN (Mars Atmosphere Volatiles Experiment) mission is focusing on the aeronomy of the Martian upper atmosphere and ionosphere. In the sections below, we present an overview of the basic structure of planetary upper atmospheres and then focus on ways that variations in these neutral atmospheres can influence planetary ionospheres.

9.1 Global mean upper-atmospheric structure

9.1.1 Composition

As noted above, and as discussed in standard textbooks (e.g., Chamberlain and Hunten, 1987; Banks and Kockarts, 1973) the two distinguishing characteristics of planetary upper atmospheres are the transition to diffusive equilibrium and a positive temperature gradient. While the altitudes marking these transitions are physically near each other, they should not be confused as being identical since, to first order, they are governed by different physical processes. First, we consider the homopause. At altitudes above the homopause, the composition varies as a function of altitude due to molecular diffusion. Since molecular diffusion coefficients (D) vary inversely as molecular mass, the molecular diffusion velocities are greater for the lighter constituents and smaller for heavier constituents. Furthermore, they vary inversely as the total density (i.e. diffusion of a gas is more rapid if collisions are less frequent), thus D increases with altitude.

By contrast, in the lower atmosphere, collisions are sufficiently rapid that bulk atmospheric motions dominate. These bulk motions have historically been characterized by an eddy diffusion coefficient known as K (Colegrove *et al.*, 1965, 1966) and the homopause historically has been defined as that altitude where D equals K . The term “turbopause” is often used synonymously for this layer and expresses the concept that turbulence, with associated constituent mixing, effectively ceases at this altitude (Hall *et al.*, 2008). Physically, this turbulence has been linked to the breaking of small-scale gravity waves (Garcia and Solomon, 1985) which are typically unresolved in global models. Since the resultant diffusion is primarily in the vertical, K is typically expressed as either K_z , $K(z)$ or K_{zz} (the latter because middle atmosphere waves are often parametrized as a horizontal eddy diffusion, K_{yy} ; Garcia (1991). As we will discuss in this chapter, for upper-atmospheric applications, this view of atmospheric mixing as being solely due to turbulent diffusion is too simplistic. As theoretical models have become more sophisticated in resolving atmospheric wave motions, it has become recognized that there is a hierarchy of atmospheric chemical and dynamical phenomena that can produce mixing. For example, Fuller-Rowell (1998) showed that the large-scale height–latitude circulation of the thermosphere at solstice effectively increased the constituent mixing by bring up molecule-rich air from lower altitudes. A model which did not capture this circulation-induced mixing (for example, a purely one-dimensional globally averaged model) would necessarily require a larger value of K_{zz} to properly simulate the resultant vertical constituent profiles. Nonetheless, the concept of eddy diffusion and a turbopause remains quite useful as a tool to understand the variability of planetary upper-atmospheric structure.

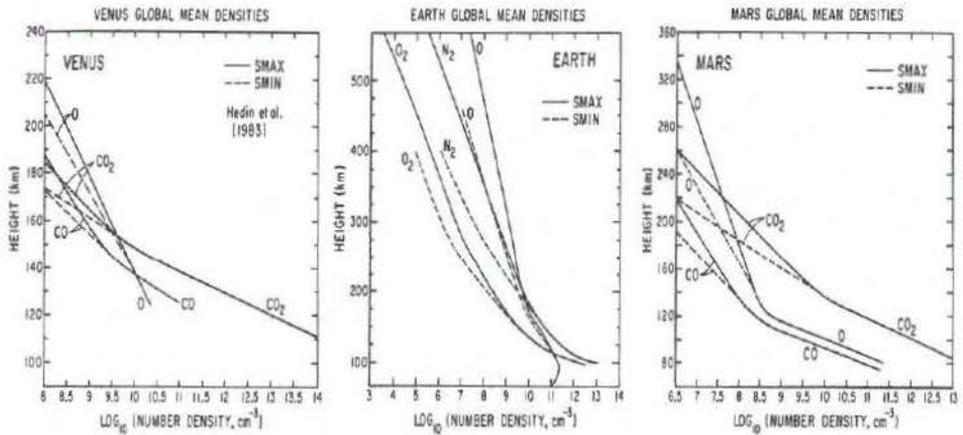


Fig. 9.1 Compositions of the upper atmospheres of Venus, Earth, and Mars. (From Bougher and Roble, 1991.)

Above the homopause or turbopause, the relative abundance of the lighter atomic constituents, such as atomic oxygen, increase at the expense of molecular constituents. Altitude profiles of upper-atmospheric constituents for the Earth, Venus, and Mars are given in Ch. 13 of Vol. III; a similar figure is given here as Fig. 9.1. In all three planets, the relative abundance of atomic oxygen increases with increasing altitude; however, this transition occurs much more rapidly for the Earth. General turbopause altitudes for Venus are in the range 130–136 km (von Zahn *et al.*, 1980), for Mars, 120–130 km (Nier and McElroy, 1977), and for the Earth about 100–110 km (Colegrove *et al.*, 1965, 1966). The difference between the Earth and the other two terrestrial planets is probably linked to more vigorous mixing from the lower atmospheres of Mars and Venus, hence higher K values. However, in the case of Venus, it was hypothesized by von Zahn *et al.* (1980) that the global circulation might provide a contribution to vertical mixing that could mimic high K much as we discussed above in the terrestrial case described by Fuller-Rowell (1998).

It should also be noted that strictly speaking the designation of a single altitude (or narrow range of altitudes) as the homopause is an oversimplification for two reasons. First, since the transition to diffusive equilibrium depends upon the mean molecular mass of each constituent, we expect the $K = D$ criteria to be different for different constituents. Garcia *et al.* (2014) recently discussed the case of CO_2 on Earth. Since CO_2 is heavier than the background atmosphere, it has a smaller vertical scale height (defined for the i th constituent as $H = kT/m_i g$ where k is Boltzmann's constant, m_i is the mean molecular mass, and g is gravity) and also experiences a net downward diffusion velocity (by contrast, atomic species such as atomic hydrogen and helium, which are lighter than the background atmosphere

experience a net upward diffusion velocity) (cf., Garcia *et al.*, 2014, their equation (5)). The net result is that the CO₂ density decreases more rapidly with altitude than does the background atmosphere and thus CO₂ departs from a well-mixed condition at a relatively low altitude (~ 0.01 hPa or 80 km) compared with the conventional turbopause at 100 km. Second, there is a large variability in K such that the $K = D$ criteria can vary significantly with latitude and season (Smith, 2012).

The sensitivity of the upper-atmospheric composition to variability in eddy diffusion was first considered by Colegrove *et al.* (1965, 1966). They showed the consequences of varying K_{zz} by a factor of 4 on the calculated O, O₂, and N₂ profiles of the atmosphere. They showed that for higher values of K_{zz} , more atomic oxygen is transported downwards, out of the thermosphere to where chemical recombination can more rapidly occur. Through diffusive equilibrium, this depletion gets transmitted up to higher altitudes so that the net effect of increasing K_{zz} is to decrease the atomic oxygen in favor of the molecular constituents. This is replicated in Fig. 9.2, which shows two K_{zz} profiles (top panel) and the calculated neutral constituents (middle panel). The model we use is the global averaged thermosphere–ionosphere model first introduced by Roble (1987). The case with the increased K_{zz} profile displays consistently lower atomic oxygen from 100–250 km. O₂ is seen to vary inversely.

These neutral atmospheric changes from varying K_{zz} can then be transmitted to the ionosphere through ion–neutral chemistry. The importance of varying recombination rates on the ionosphere is discussed in Ch. 13 of Vol. III. Briefly, because the radiative recombination of O⁺ with free electrons is very slow, the rate limiting pathway for O⁺ recombination is the reaction with either O₂ or N₂, followed by dissociative recombination. The dissociative recombination of NO⁺ or O₂⁺ is approximately 10⁶ times faster than the radiative recombination of O⁺. The bottom panel of Fig. 9.2 shows that for higher K_{zz} , the O₂⁺ and NO⁺ is increased and the O⁺ is decreased. Thus Fig. 9.2 demonstrates why we expect the electron density near the peak ionosphere to correlate with the O/N₂ ratio and to anticorrelate with eddy diffusion at the base of the thermosphere.

The above demonstration is important for two reasons. First, K_{zz} continues to be a favorite free parameter of thermosphere–ionosphere modelers as a mechanism to parametrize uncertain dynamical effects emanating from the lower atmosphere on thermosphere neutral constituents. Qian *et al.* (2009) postulated a seasonal variation of a factor of 5 in K_{zz} to explain variations in the neutral density sampled by satellites at 400 km. Much as we have done here, they showed how the O/N₂ ratio varied inversely with their assumed K_{zz} and argued that the resultant variations agreed with observations from the TIMED Global UV Imager (GUVI) instrument. Subsequently, Qian *et al.* (2013) showed that variations in the neutral

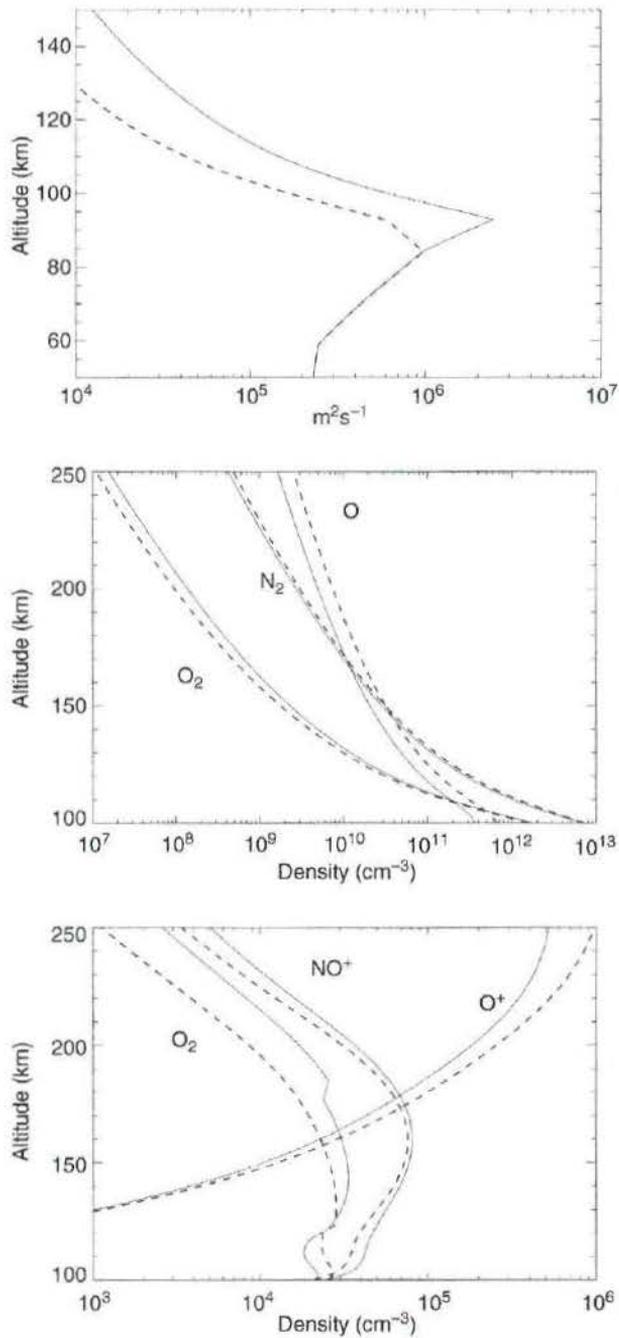


Fig. 9.2 Effect of varying K_{zz} on neutral thermospheric and ionospheric constituent profiles. The top panel shows the two K_{zz} profiles. The center panel shows the resultant neutral constituent profiles where the dashed curves are associated with the dashed K_{zz} profile. The bottom panel shows the resultant ionospheric profiles where using the dashed K_{zz} profiles leads to the largest value of O^+ at 250 km.

constituents led to variations in the ionospheric F_2 peak densities and ultimately to improved agreement with radio-occultation measurements of the ionospheric electron-density profile. Although they applied this K_{zz} change to a 3D model and we used a simple 1D globally averaged model here, the mechanism is the same. Ultimately however, K_{zz} is merely a proxy for complex dynamical effects. In the next section we give an example of one such process to illustrate how the inclusion of a more complete representation of dynamics in three dimensional general circulation models (GCMs) can obviate the need for large values of assumed K_{zz} in models.

Second, our example illustrates the important link between vertical motion of constituents and the ionosphere. This motion can be either upwards or downwards. While increasing K_{zz} decreases O by transporting it down to the lower atmosphere, increases in upward bulk motion (vertical wind) will have similar consequences because it will transport molecule-rich air up to higher (or down to lower) altitudes and locally decrease (or increase) the O/N_2 ratio. This phenomenon is well known in cases where there is added heat deposited in the lower thermosphere. For example, during strong geomagnetic activity, the increased heating from auroral particles and electric fields acts to cause upwelling of N_2 -rich air that will change the composition of the ionosphere (Hays *et al.*, 1973; Burns *et al.*, 1989). More recent work has emphasized that both upwelling and downwelling can occur with concomitant changes in the O/N_2 ratio and F region electron density that vary together (Fuller-Rowell *et al.*, 1996; Immel *et al.*, 2001; Crowley *et al.*, 2006). Pross (2012) reviewed the long history of research into the response of thermospheric and ionospheric composition to solar and auroral energy input.

For Mars, the analog to the terrestrial O/N_2 ratio is the O/CO_2 ratio. Like on Earth, ionospheric chemistry is strongly impacted by the O abundance. Specifically, the conversion of CO_2^+ (from CO_2 ionization) to O_2^+ (primary ion) via atomic O is a very fast reaction. The measured O_2^+/CO_2^+ ratio at the dayside F_1 -ion peak can be used as a diagnostic of the local O/CO_2 ratio itself (see Hanson *et al.*, 1977). Also, like on Earth, the O/CO_2 ratio above the homopause is sensitive to the magnitude of eddy diffusion.

Unfortunately, Martian atomic oxygen abundances are not presently well measured. The limited data we have dates back to the Viking 1 and 2 Landers (e.g., Nier and McElroy, 1977). The two sets of density profiles, corresponding to solar minimum near aphelion conditions, reveal CO_2 to be the major species, followed by N_2 and CO. Atomic O could not even be measured; instead, ionospheric model calculations (to match the Viking Lander 1 ion composition measurements) have estimated that the O density exceeds the CO_2 density at ~ 200 km (e.g., Hanson *et al.*, 1977). Likewise, the O/CO_2 mixing ratio at ~ 130 km (the altitude of the

dayside F_1 -ion peak) was estimated to be 1.2%. There have also been some uses of Martian dayglow emission to retrieve neutral densities from both the Mariner and Mars Express missions (e.g., Strickland *et al.*, 1972, 1973; Stewart *et al.*, 1992; Chaufray *et al.*, 2009; Gronoff *et al.*, 2012). Extracted O/CO₂ ratios range from ~0.6% to 2.0% at a similar F_1 -peak altitude, reflecting a mixture of solar cycle and local time variations. Atomic O should also serve as a tracer of the Martian global thermospheric circulation for which larger O/CO₂ ratios should appear on the nightside than the dayside. Finally, variations in the global distribution of the O/CO₂ ratio should accompany solar-flare events, similar to the O/N₂ variations seen in the Earth's upper atmosphere. Acquiring comprehensive O/CO₂ data from both airglow and mass spectrometry is a key focus of the MAVEN mission.

As summarized by Nair *et al.* (1994), the data that we do have on Mars' thermospheric composition described above have led to K_{zz} estimates which are quite high, well in excess of $10^7 \text{ cm}^2 \text{ s}^{-1}$. As compared with Fig. 9.2, this is over a factor of 10 greater than typical global mean values assumed for the Earth and is linked to the correspondingly higher homopause on Mars relative to the Earth. This factor of 10 is consistent with recent work on gravity waves on Mars (Barnes, 1990; Fritts *et al.*, 2006) which suggests about a $10 \times$ greater flux of gravity waves relative to Earth, ultimately driven by the more prominent topography on Mars.

9.1.2 Temperature

The second transition process which distinguishes planetary upper atmospheres is the steep increase in temperature with increasing altitude (and which leads to the term thermosphere). This arises because the peak EUV and FUV absorption (i.e. the heat sources) occurs at altitudes above the infrared cooling which balances this heating globally. Thus heat is transported downward via thermal conduction. The altitude at which the entire conducted heat is radiated away is defined as the mesopause, where $\frac{\delta T}{\delta z} = 0$. On all three of the terrestrial planets, the mechanism for IR cooling at the mesopause is radiation by CO₂ at 15 μm . This emission is enhanced by the collision of atomic oxygen with CO₂ which excites the $\nu_2 = 1$ bending mode of CO₂. This process proceeds with different efficiencies on the Earth, Venus, and Mars and this impacts the altitude where the $\frac{\delta T}{\delta z} = 0$, is satisfied. This is another area where the O/CO₂ ratio and its variability are believed to be of importance on planets such as Venus and Mars. The magnitude of this IR cooling at the base of the Martian thermosphere is throttled by this O abundance, and both should vary with the solar cycle (see Bougher *et al.*, 1999, 2000). Such variability can serve to limit the solar cycle variation in dayside thermospheric temperatures.

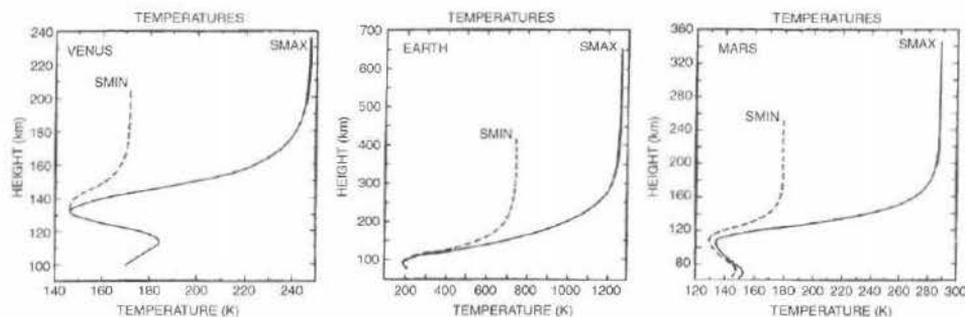


Fig. 9.3 Three planet global mean temperature profiles for solar minimum (SMIN) and maximum (SMAX) conditions. (From Bougher and Roble, 1991.)

Figure 9.3 illustrates the globally averaged temperature profiles from Venus, the Earth, and Mars. Some important differences between the three planets are apparent. First the mesopauses of the three planets are at different altitudes, lowest on the Earth, highest on Venus, and intermediate on Mars. This reflects the efficiency of the CO_2 cooling. On Earth, CO_2 is a minor constituent and the thermospheric heat is not fully radiated away until transported down below 100 km. Second, the relative paucity of CO_2 in the terrestrial thermosphere means that the Earth's thermosphere gets much hotter than those of either Venus or Mars. Finally, because of the inefficiency of cooling in the terrestrial thermosphere, it is more sensitive to variations in the input heat source from the Sun. Thus the solar-cycle variation in the terrestrial thermosphere is larger than on Venus and Mars (Bougher *et al.*, 1999, 2000, 2009) and this has been confirmed by densities derived from orbital drag measurements from the Mars Global Surveyor (MGS) spacecraft (Forbes *et al.*, 2008) as well as earlier such data from Venus spacecraft (Kasprzak *et al.*, 1997).

There is one complication in the terrestrial thermospheric energy budget which is not present on Mars and Venus, namely the existence of an additional major cooling term. This is the emission from vibrationally excited NO at $5.3 \mu\text{m}$. Although, as noted in Ch. 13 of Vol. III, the triple bond of molecular nitrogen is strong, energetic electron impact from both photo and auroral electrons can dissociate N_2 and create atomic nitrogen (Siskind *et al.*, 1989a, b). This atomic nitrogen is rapidly oxidized in the lower thermosphere to make nitric oxide (NO). The key characteristic of NO that is relevant for the thermal budget is that there is a pronounced solar-cycle variation in its abundance (Barth *et al.*, 1988). Thus its role in the thermospheric heat budget maximizes at solar maximum. Roble (1987) quantified the relative roles of NO and CO_2 in the energy budget. Figure 8 of that paper is reproduced here as Fig. 9.4. Note that the CO_2 cooling curve changes little from solar maximum to solar minimum whereas at solar maximum, the NO profile is greatly

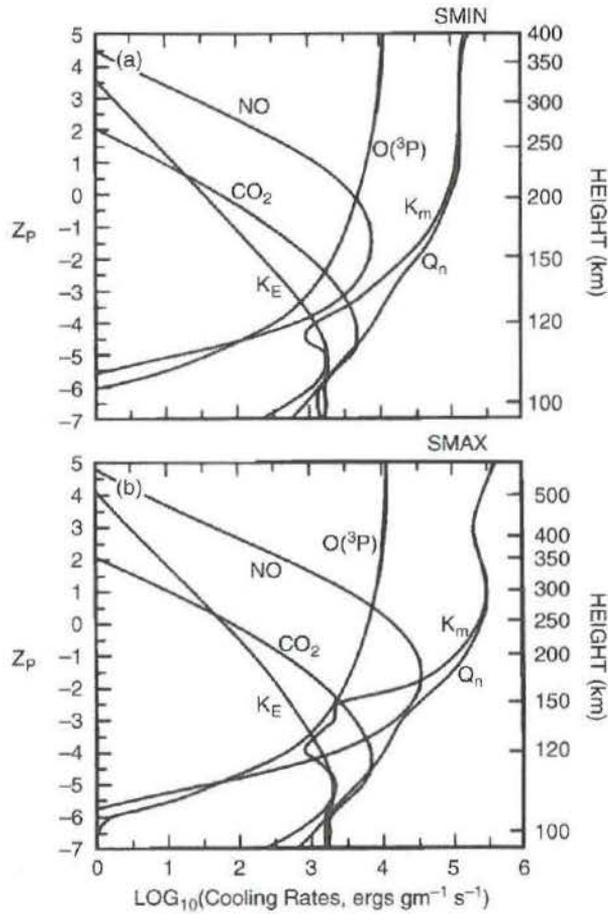


Fig. 9.4 Calculated \log_{10} neutral gas heating and cooling rate profiles ($\text{ergs gm}^{-1} \text{ s}^{-1}$) for (a) solar minimum and (b) solar maximum conditions. Q_n is the total neutral heating rate, K_m is the cooling rate by downward molecular thermal conduction, K_E is for eddy thermal conduction; NO is radiative cooling from the $5.3 \mu\text{m}$ emission from nitric oxide, CO_2 is radiative cooling from $15 \mu\text{m}$ emission from CO_2 and $O(^3P)$ is cooling from the fine structure of atomic oxygen. (From Roble, 1987.)

enhanced. Observations of the infrared emission from these two cooling terms by the Sounding of the Atmosphere by Broadband Emission Radiometry (SABER) instrument on TIMED (Mlynczak *et al.*, 2010) have confirmed this theory.

The varying significance of NO and CO_2 cooling in the terrestrial thermosphere has important consequences for one variation in the Earth's thermosphere that is without any analog on Mars or Venus, namely anthropogenic global change. Since the Roble and Dickinson (1989) paper, it has been recognized the upper atmosphere

should cool and thus become less dense at a given altitude in response to increases in the CO₂ abundance. Evidence for density decreases has emerged from a study of 40 years of orbit data (Emmert *et al.*, 2008). It is also apparent that there is a solar-cycle dependence to this trend with the largest density decreases at solar minimum when the CO₂ cooling term is most dominant. It also has been proposed that this should affect the ionosphere, most evidently as lowering of the F₂ layer height (Qian *et al.*, 2009, and references therein). As with the density trend, this should be most apparent at solar minimum. Chapter 14 of Vol. III presents an overview of a wide range of scenarios for long term geospace climate change, including, among others, solar Maunder minimum conditions and geomagnetic field reversals. The conclusion is that the recent finding of neutral density decreases is more robust than the purported changes in the F₂ layer height due to uncertainties in the interpretation of ionosonde data.

9.2 How do neutral dynamics affect planetary ionospheres?

Having considered the global mean structure and variability of terrestrial thermospheres and ionospheres, we now review in greater detail the dynamics by which neutral thermospheres will affect planetary ionospheres. The dynamics of the upper atmosphere have been discussed in Ch. 15 of Vol. III (Walterscheid, 2010). Here we present recent observations and models of coupling processes between the neutral and ionized fractions of planetary upper atmosphere. Of particular interest is the coupling with the lower atmosphere. As we will discuss, many of the perturbations which are manifested in ionospheric variability can be tracked back to meteorological disturbances in the troposphere and middle atmosphere. Thus the differences in the lower and middle atmospheres of the terrestrial planets have ionospheric consequences.

Three differences, relative to the Earth, that we encounter in studying Mars are (1) the absence of a significant dipole magnetic field on Mars, (2) the large orbital eccentricity that provides a $\pm 22\%$ variation in the net solar flux received at the planet during the year, and (3) the unique role of dust. The electrodynamic consequences of the first factor are discussed in Sect. 9.2.1 below. Of the latter two effects, they are coupled in that dust heating is an important atmospheric forcing function and the variable heating throughout the year affects this forcing such that the height of a dayside reference pressure level rises/falls by 10–15 km over the course of the year. This seasonal expansion and contraction of the lower atmosphere impacts the upper atmosphere as well. For example, since the dayside F₁ (primary ion peak) occurs at approximately a constant pressure level, its altitude also varies by 10–15 km in step with the lower/middle-atmospheric forcing (see Stewart, 1987; Zhang and Busse, 1990). In general, the photochemical control of

the Martian ionosphere (below 200 km altitude) means that perturbations are often manifested as changes in the height of the electron density peak and we will see other examples of this below.

Concerning dynamics, we consider both the background zonal mean wind and the waves which represent the variability about this mean state. The background mean wind can affect ionospheric plasma through direct perturbations to the O/N₂ (or O/CO₂) ratio, through ion–neutral coupling and associated electrodynamic effects, and by changing the background stability of the atmosphere. Of wave perturbations, following Ch. 15 of Vol. III, we consider three classes: rotational low-frequency planetary waves, divergent high-frequency gravity waves, and tides, which combine aspects of both planetary and gravity waves. Wave phenomena can impact the ionosphere in one of two ways. First, wave dissipation is a means for transferring angular momentum from one region of the atmosphere to another. This momentum deposition can change the circulation field and impact neutral constituent transport. As we will discuss below, what is considered as K_{zz} variability in a one-dimensional global mean framework above, can often be resolved as a specific wave mode in a three-dimensional framework. An example of the second wave effect, from non-dissipating waves, is those that result from oscillations of the ion–neutral dynamo. We consider each dynamical category in turn below.

Finally, we stress that the discussion below is merely to highlight some similarities and differences between the Earth and Mars. It is by no means fully comprehensive, particularly for the Earth. A comprehensive modern discussion of the Earth's ionosphere is given by Kelley (2009). For Mars, there are recent reviews of the Martian ionosphere by Withers (2009), Haider *et al.* (2011), and Bougher *et al.* (2014).

9.2.1 Mean winds and electrodynamics

As we have noted above, in any comparative analysis of terrestrial planetary upper atmospheres and the coupling between the ionic and neutral constituents, the fact that the Earth has a large intrinsic magnetic field, and Mars (and Venus) do not, is of primary importance. One interesting consequence is that, as observed from deep space, the Earth's ionosphere reveals structure that is directly the result of its magnetic field. Observations of the far-ultraviolet airglow from the Apollo 16 astronauts on the Moon (Carruthers and Page, 1976) show this nicely (Fig. 9.5). One sees two bands of airglow on the nightside that are roughly symmetric about the equator. This is the so-called Appleton anomaly, or the Equatorial Ionospheric Anomaly (EIA). It is an electrodynamic effect that is due ultimately to E–W (zonal) neutral winds in the E-region (110 km) which through collisions with the ions,

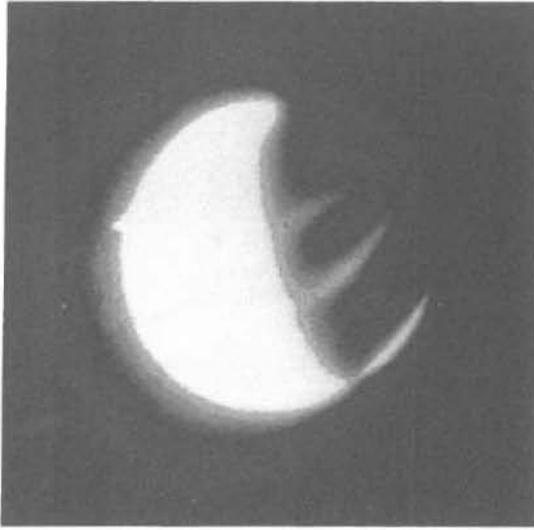


Fig. 9.5 FUV airglow of the Earth as observed with the Apollo 16 lunar camera. (From Carruthers and Page, 1976.)

create an E–W electric field. This is known as a dynamo electric field and it leads to a vertical ion drift according to

$$\mathbf{v} = \frac{\mathbf{E} \times \mathbf{B}}{B^2}. \quad (9.1)$$

As outlined by Fig. 2 of Immel *et al.* (2006), plasma gets lifted upwards over the equator and then, at F-region altitudes, diffuses poleward and downward along magnetic field lines. Thus the peak electron densities are typically found on either side of the equator, even though the photoionization source peaks at the equator. As we discuss in Sect. 9.2.2, the variability of the E-region dynamo caused by oscillations in neutral tidal winds has been of great recent interest.

Mars has no analog to the above ionospheric structure. However, it is untrue to say that there are no magnetic field effects on the Martian ionosphere. Rather, complex crustal fields, a remnant of the formerly strong Mars intrinsic magnetic field, were observed by the MGS (Acuña *et al.*, 1998). Field strengths at ionospheric altitudes can exceed 1000 nT (several percent of Earth’s field strength) which is sufficient to counter the solar-wind pressure. These regions, where both ends of the crustal field lines that thread the ionosphere intersect with the surface, rather than being open to space have thus been termed “mini-magnetospheres” (Mitchell *et al.*, 2001). Electron density enhancements have been indicated both above these crustal field regions (Nielsen *et al.*, 2007; Duru *et al.*, 2008) and within the mini-magnetospheres (Krymskii *et al.*, 2009). An example of one such enhancement is

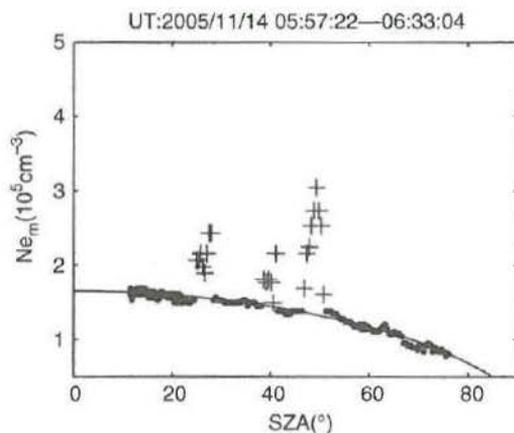


Fig. 9.6 Electron densities from the Mars Express as function of solar zenith angle (SZA). (From Nielsen *et al.*, 2007.)

given in Fig. 9.6. Explanations for these enhancements have focused on elevated electron temperatures which would reduce the rate of dissociative recombination (Schunk and Nagy, 2009). In turn, the causes of these electron temperature enhancements have been attributed to a two-stream plasma instability above the fields and to trapping of hot photo-electrons within the mini-magnetospheres.

9.2.2 Tides

Tides are the response to periodic thermal or mechanical forcing of the atmosphere, typically on diurnal time scales. Tides grow in amplitude as they propagate upwards until by thermospheric altitudes, they are the dominant wave mode in the neutral wind. They are labeled by their direction of propagation (eastward or westward), their period (i.e. diurnal, semi-diurnal), and their wavenumber. There are two categories of tides: migrating or non-migrating. Migrating tides are forced by atmospheric heating that follows the diurnal rising and setting of the Sun and propagate westward. Mathematically, tides are characterized by their zonal wavenumber (e.g., $s = \dots - 2, -1, 0, 1, 2 \dots$) and temporal harmonic ($n = 0, 1, 2$), where diurnal tides are $n = 1$, semi-diurnal are $n = 2$, etc. Westward-propagating tides have $s > 0$, eastward-propagating tides, $s < 0$, and zonally symmetric tides have $s = 0$. Migrating tides are therefore defined as $s = n$ and are independent of longitude while non-migrating tides have $s \neq n$ and capture the longitudinal dependence (Bougher *et al.*, 2004). As discussed by Forbes *et al.* (2002), when observed by a satellite in Sun-synchronous orbit, non-migrating tides appear as stationary waves varying in longitude with a zonal wavenumber of $|s - n|$.

Table 9.1 Summary of primary tidal forcings for Earth and Mars.

Tidal component	Earth	Mars
Migrating diurnal	NIR abs. of H ₂ O	VIS and IR abs. of CO ₂
Migrating semi-diurnal	Ozone heating	Dust heating
Non-migrating	Latent heat release from tropical convection	Topographic modulation of near surface heating

As first discussed by Lindzen (1970) and noted by Moudden and Forbes (2010), the similar rotation rates of the Earth and Mars implies that the mathematical characteristics of the tidal modes for the Earth and Mars are similar. However, the excitation of these tides differs greatly between both planets and results from unique characteristics of each planet's atmosphere and surface. For both planets, absorption by IR active gases drives the migrating tides; on the Earth, this is H₂O and stratospheric ozone; on Mars, it is CO₂. Additionally, absorption by dust in the Martian atmosphere contributes significantly to the excitation of the semi-diurnal tide. An additional source for the diurnal tide in the thermospheres of both planets arises from absorption of solar EUV at altitudes above about 120 km. Regarding non-migrating tides, on the Earth, they are excited by the latent heat release from tropical convection; the general concentration of this convection in three rain forests across the tropics (Amazon, African, SE Asian) leads to a wavenumber 3 pattern. On Mars, non-migrating tides are forced by the modulation of the daily cycle of solar surface heating by the very large surface pressure variations forced from the prominent variations in Martian topography (Zurek, 1976). Table 9.1 summarizes these differences. In general, due to these large topographic variations, the manifestation of these tides in the Martian upper atmosphere is greater on Mars than the Earth.

On Mars, tidal variability has shown up as an obvious feature in accelerometer data from the Mars Global Surveyor orbiter (Keating *et al.*, 1998; Withers *et al.*, 2003; Haider *et al.*, 2011 (their Fig. 7)) and are attributed to the topographic modulation of solar thermal tides excited near the surface. Figure 9.7 shows an example of these waves. The amplitudes of these waves appear to maximize at about 130 km (just above the mean homopause) and decay with increasing altitude up to approximately 160 km (presumably due to dissipation). These altitudes are near the peak of the Martian ionosphere and the neutral thermospheric waves have a corresponding impact on the ionospheric structure: e.g., the height of the main ionospheric peak varies in step with these wave features (Bougher *et al.*, 2001, 2004; Withers, 2009). Figure 9.8 shows an example of these longitude-fixed features in the MGS data analyzed by Bougher *et al.* (2004). A spectral fit to these data showed the presence of a wave 3 variation, consistent with a semi-diurnal wave frequency

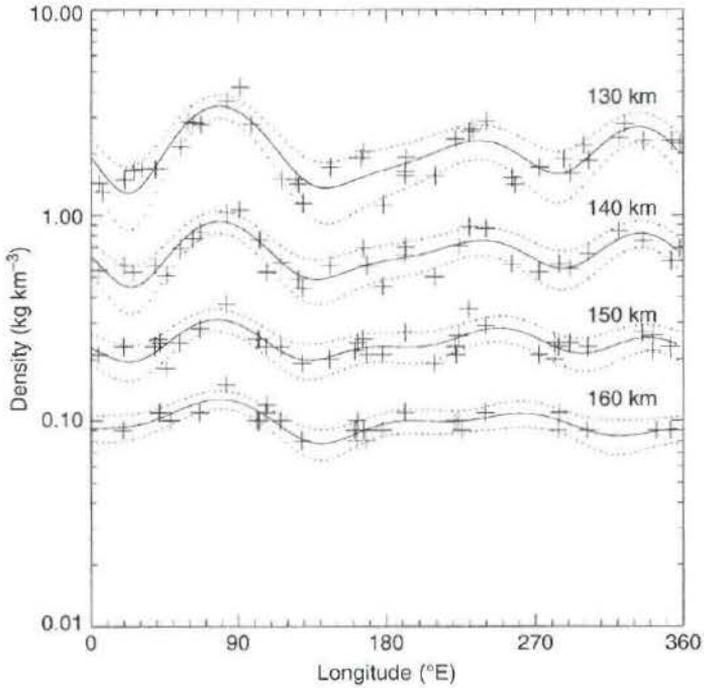


Fig. 9.7 Mass density measurements at 130, 140, 150, and 160 km altitude obtained between 10° – 20° N during MGS aerobraking. Model fits to data (waves 1–3) from each altitude are plotted as solid lines and $1-\sigma$ uncertainties about each fit as dotted lines. All data are taken from a local time of 15 h. Measurements at each altitude were taken over about 1 week. All statistically significant peaks and troughs appear fixed in longitude. (From Withers *et al.*, 2003.)

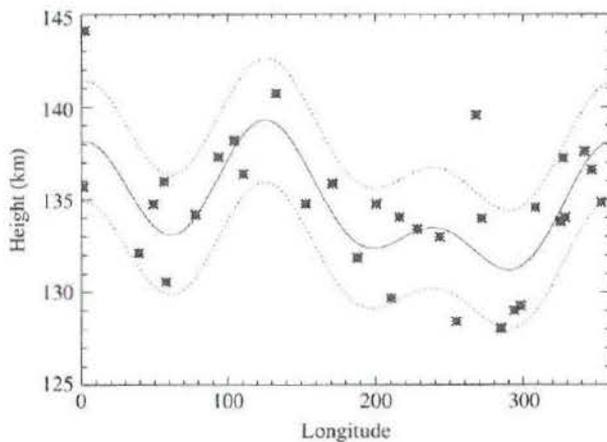


Fig. 9.8 Heights of primary electron-density peak showing a wave pattern. (From Bougher *et al.*, 2004.)

indicating non-migrating (i.e. $s \neq n$) tidal forcing. These altitude changes were considered notable because such variability was unexpected given that the peak of the Martian ionosphere is in photochemical equilibrium and the background solar flux, latitude, and local time had all remained constant.

On the Earth, motivated by the observations by Immel *et al.* (2006) of longitudinal variations in ionospheric airglow, there has been an explosion of interest in interactions between tides and the ionosphere. Indeed, this interest was a key motivating factor behind the selection of the Ionospheric Connections (ICON) Explorer by NASA for a 2017 launch. Immel *et al.* showed that there was a distinctive wave 4 pattern in the brightness pattern of the F-region airglow and identified the non-migrating diurnal eastward-propagating wave 3 (DE3) tide as the cause (note, as discussed above, a diurnal tide ($n = 1$) with $s = -3$ (eastward-propagating), will be observed as a wave 4 pattern in a fixed local time satellite measurement). Their key result is shown as Fig. 9.9. They proposed that tidal oscillations in the neutral wind at E-region altitudes will modulate the zonal electric field which in turn, modifies the $E \times B$ vertical ion drift given above in Eq. (9.1). Because this redistributed plasma produces airglow through O^+ chemistry, any variations in the source of this plasma will be reflected as airglow variability. These results have sparked many studies on the coupling between tropospheric tides and the ionosphere. Hagan *et al.* (2009) looked at the effects of tidal dissipation on the E-region winds. Other studies that identified other waves which could contribute the patterns found by Immel *et al.* include Pedatella *et al.* (2012) and Liu and Richmond (2013).

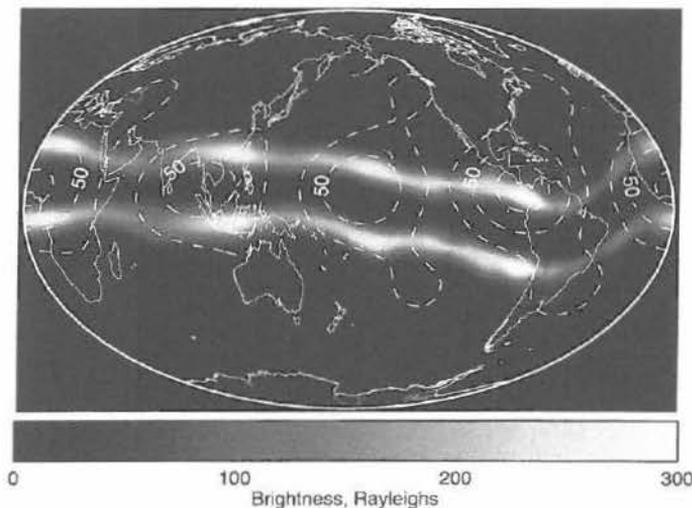


Fig. 9.9 Earth's FUV airglow showing a wave pattern. (From Immel *et al.*, 2006.)

Global whole-atmosphere models have had partial success in capturing this feature. For example, Jin *et al.* (2011) demonstrated the quantitative link between the wavenumber spectrum of tropical rainfall and that of key ionospheric parameters such as peak F_2 density and vertical $E \times B$ drifts.

Recent work has highlighted a second mechanism by which tides can influence the ionosphere, through direct modulation of the vertical transport of atomic oxygen. Siskind *et al.* (2014) used the NCAR Thermosphere Ionosphere Electrodynamics GCM (TIEGCM) and compared calculated electron densities for a case where the TIEGCM was forced at the boundary by an idealized representation of migrating tides versus a case where it was forced by realistic meteorological variability (winds, temperatures, and geopotential heights) from the NOGAPS-ALPHA high-altitude weather model (Navy Operational Global Atmospheric Prediction System – Advanced Level Physics High Altitude; Eckermann *et al.*, 2009). Figure 9.10 shows spectra of the calculated equatorial vertical wind from the TIEGCM near 115 km for the case with the NOGAPS-ALPHA boundary forcing and compares with the standard NCAR formulation that solely uses forcing from migrating tides. The figure shows that the NOGAPS/TIEGCM combination displays a rich spectrum of diurnal, semi-diurnal, and terdiurnal components which

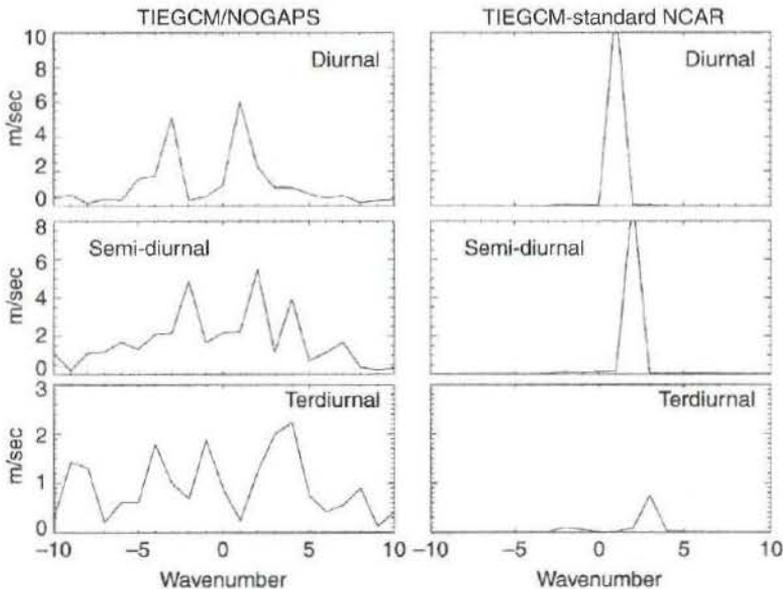


Fig. 9.10 Calculated spectrum of the lower-thermospheric vertical wind, monthly averaged (March) at the equator near 115 km. The calculation on the left uses a model forced by realistic meteorological conditions from the NOGAPS-ALPHA assimilation system. The calculation on the right assumes only forcing from migrating diurnal and semi-diurnal tides.

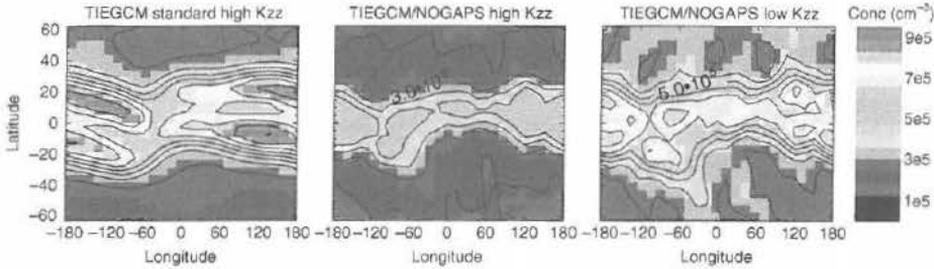


Fig. 9.11 Calculated averaged peak F_2 electron density for March, 1300 local for three TIEGCM simulations. Left column uses standard $K_{zz} = 125 \text{ m}^2 \text{ s}^{-1}$ with vertical winds from the right column of Fig. 9.7. Middle uses NOGAPS-ALPHA vertical winds. Rightmost field is with NOGAPS-ALPHA vertical winds and K_{zz} divided by 5. A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

when taken together drive a more vigorous upwelling than that obtained with solely migrating tides. Siskind *et al.* (2014) found that the calculated ionosphere using the NOGAPS-ALPHA forcing showed surprisingly low electron densities unless the assumed K_{zz} was divided by 5 from the standard NCAR input. Figure 9.11 shows a sample of their results, comparing the monthly averaged model output for the standard NCAR case (migrating tides + high K_{zz}), the NOGAPS-ALPHA run with high K_{zz} , and the NOGAPS-ALPHA run with lower K_{zz} . Much as we demonstrated for the one-dimensional case in Fig. 9.2, Siskind *et al.* (2014) traced the effect of lowering K_{zz} on atomic oxygen and showed that the case with the lower K_{zz} gave atomic oxygen mixing ratios near the homopause which were in better agreement with SABER observations. This illustrates two points. First, that models which incorporate a greater representation of lower atmospheric dynamics should use a lower value of K_{zz} than models which parametrize a portion of these dynamics and, second, that an accurate representation of the neutral atmosphere at the homopause is important for an accurate calculation of the ionosphere.

9.2.3 Planetary waves

Planetary waves can be classified as either traveling or stationary. Traveling waves are normal modes of the atmosphere. Typical periods for the Earth are centered near 2, 5, 10, and 16 days. Stationary waves are forced in the troposphere by topographic variations either from land–sea contrasts or orography (mountain ranges). In both cases, when these waves dissipate they can transfer momentum to the background flow. As noted in Ch. 15 of Vol. III and by Liu *et al.* (2010), while the middle atmosphere supports a rich variety of planetary waves, it is typically believed that the stationary waves either dissipate or are absorbed by a critical level (where the

background wind equals the phase speed of the wave) before reaching thermospheric altitudes. Cahoy *et al.* (2006) came to an analogous conclusion for Mars, i.e. thermal tides dominate the variability above 75 km altitude.

Given the limitations on stationary-wave propagation into the thermosphere, it was therefore somewhat of a surprise that pronounced changes in tropical ionospheric vertical drifts were observed in response to the sudden stratospheric warmings (SSWs) (which are essentially triggered by stationary planetary waves in the stratosphere) in January 2008 (Chau *et al.*, 2009) and 2009 (Goncharenko *et al.*, 2010). Figure 9.12 shows the salient result: the normal diurnal variation of the upward ion drift is significantly perturbed during SSW events. Since it is unclear how stationary waves in the polar middle atmosphere could affect the tropical ionosphere, attention has focused on interactions between these waves and tides, which can then propagate into the thermosphere. Through the dynamo mechanism outlined above, tidal wind variability could then couple to the ionosphere. There have been several model simulations of the effects of SSWs on the thermosphere and ionosphere; not all of them are consistent. Pedatella *et al.* (2014a,b) analyzed the results of four different so-called whole-atmosphere model simulations of the dramatic 2009 SSW. They concluded that whereas the diurnal tide variability was similar in all four models, the semi-diurnal tide differed amongst the four. Ultimately, they tracked back many of the differences to differing approaches towards gravity-wave drag parametrizations. Gravity-wave drag indirectly affects the upper atmosphere by controlling the mesospheric wind field through which upward-propagating planetary waves and tides propagate.

Concerning traveling planetary waves, an example of traveling planetary wave effects on the ionosphere that has recently received new attention was given by Chen (1992). Reproduced as Fig. 9.13, he documented the occurrence of a pronounced two-day oscillation in the peak electron density at the crest of the equatorial anomaly region. Also note from the figure that during the period of the two-day oscillations (March 6–16), that there is an overall net decrease in the electron density (given by foF₂, the maximum ordinary-mode radio frequency which can be reflected by the F₂ layer decreases from about 14 MHz prior to the event to an average of about 11–12 MHz during the event). The westward-traveling two-day wave (more properly, quasi-two-day wave since the period can vary, hereinafter Q2DW) has long been of interest to middle-atmospheric scientists and is considered to be an atmospheric normal mode that is excited by baroclinic instabilities in the mesospheric summer jets. Recent interest has been stimulated by the suggestions that it is implicated in facilitating the rapid transport of space shuttle exhaust (Nicejewski *et al.*, 2011) and that it contributes to the variability of Polar Mesospheric Clouds (Siskind and McCormack, 2014). Coordinated observations of the Q2DW in the neutral lower-thermospheric temperatures and in total electron

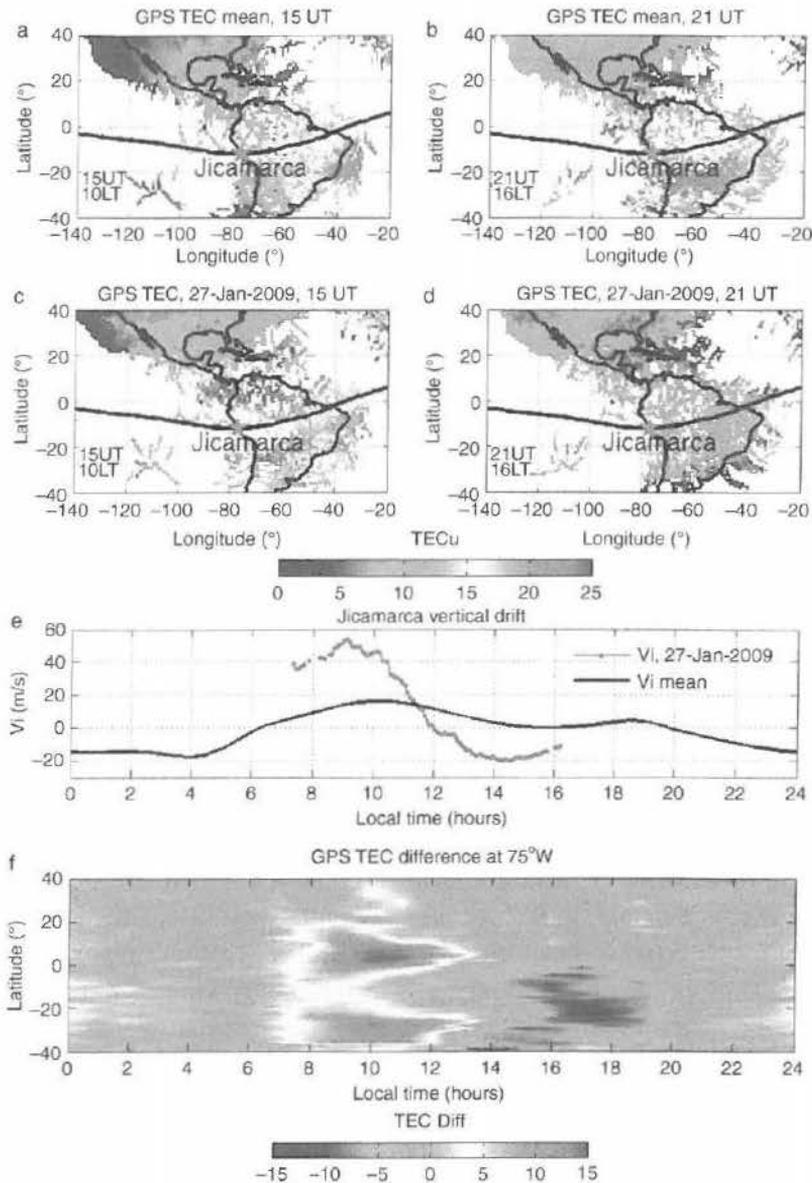


Fig. 9.12 Perturbations to the ionosphere, both total electron content (TEC) and vertical ion drift from the sudden stratospheric warming (SSW) of January 2009. The top row shows typical morning (15 UT = 10 local time at 75°W) and afternoon (21 UT) TEC fields over South America. The second row shows these fields after the SSW with a notable enhancement of TEC in the morning. The third panel shows the difference in the vertical ion drift as measured from Jicamarca Peru. The bottom panel shows difference fields between the SSW perturbation and the mean case as a function of local time, emphasizing the morning TEC enhancement and the afternoon depletion. (From Goncharenko *et al.*, 2010.) A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

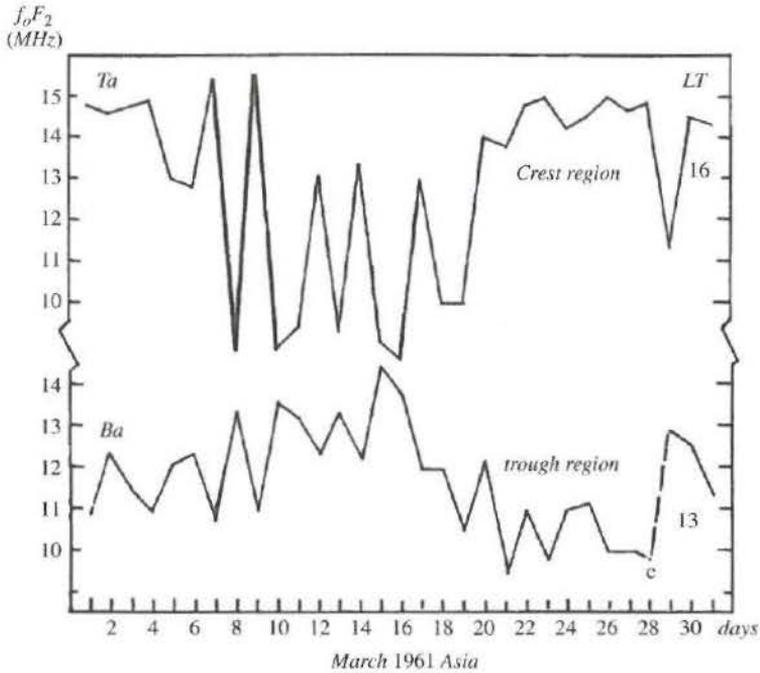


Fig. 9.13 Ionosonde data illustrating the oscillation of the peak F_2 electron density variations with a quasi-two-day wave. (From Chen, 1992.)

content (TEC) were presented by Chang *et al.* (2011). Yue *et al.* (2012) modeled the coupling between the neutral Q2DW and the ionosphere through modulation of the E-region dynamo. In this manner, the Q2DW effects on the ionosphere act similarly to the tidal modulations described above. Most recently Yue and Wang (2014) have proposed that the dissipation of the Q2DW in the lower thermosphere acts as a momentum source and drives a secondary circulation which changes the ratio of atomic oxygen to molecular nitrogen which, as we have seen, is a primary way the terrestrial neutral atmosphere modulates the terrestrial ionosphere.

9.2.4 Gravity waves and instabilities

Because of their small scale, gravity-wave effects on the thermosphere and ionosphere of both the Earth and Mars are the least well understood of the dynamical phenomena discussed here. Fundamental questions about their sources and consequences remain. Owing to the larger topographic variability coupled with faster middle-atmospheric wind speed, gravity-wave amplitudes and effects (such as turbulent mixing) appear to be much larger on Mars than on Earth (Fritts *et al.*, 2006). Accelerometers on MGS and Mars Odyssey revealed significant small-scale

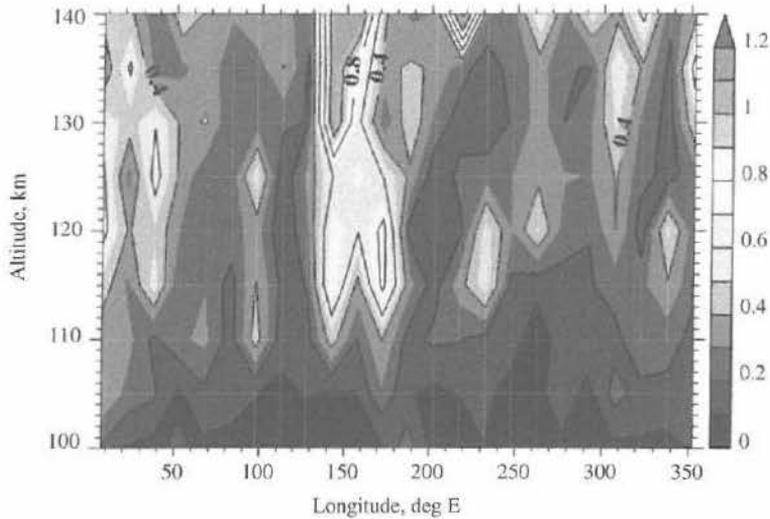


Fig. 9.14 Density variance data from the Mars Odyssey accelerometer in 15° and 5-km bins as function of longitude over about 127 orbits. (From Fritts *et al.*, 2006.) A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

fluctuations that can be attributed to gravity waves. A sample of these data is shown as Fig. 9.14, taken from Fritts *et al.* (2006, their Fig. 10). Fritts *et al.* note the localization in longitude and also that the altitudinal growth in variance is less than ρ^{-1} , which suggests dissipation, mixing, and momentum deposition. The altitudes shown correspond to the peak of the Martian F_1 ionosphere; unfortunately, there are not the data to link these data to ionospheric variability. Clearly this is an area for future research.

For Earth, there have been numerous studies linking thermospheric gravity waves both to thermospheric circulation and ionospheric instabilities. For example, gravity waves have been linked to the phenomenon of Equatorial Spread F (ESF) (also termed Convective Equatorial Ionospheric Storms; Kelley (2009)). ESF is understood as a Rayleigh Taylor instability whereby bubbles of depleted plasma can develop after sunset and rise to high altitudes. Kelly *et al.* (1981) suggested that gravity waves might be the triggering mechanism although this is unconfirmed (Woodman, 2009). ESF is almost undoubtedly sensitive to the background neutral wind. This was first shown in the two-dimensional study of Zalesk *et al.* (1982) and more recently in the three-dimensional study of Huba *et al.* (2009). Traveling Ionospheric Disturbances (TIDs) are seen as fluctuations in the F-layer electron density which are also likely linked to gravity waves. Here, however, since these disturbances propagate away from the auroral zones, it is typically believed that auroral heating is a primary source rather than the lower atmosphere, although disturbances

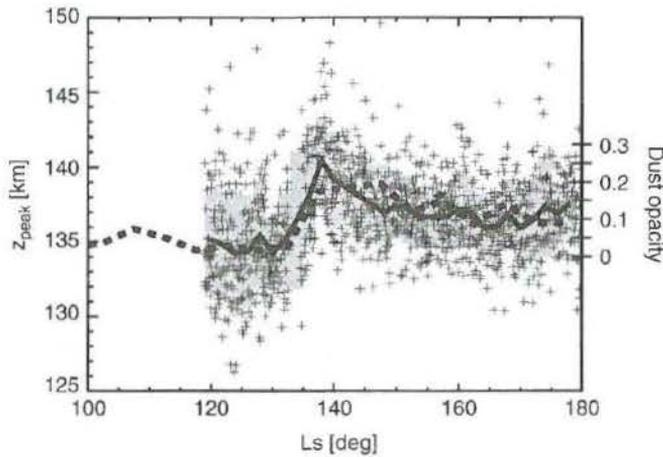


Fig. 9.15 Martian ionospheric peak altitudes measured at latitudes 62° – 80° N by the MGS radio occultation instrument during a dust storm. Data are shown as crosses, the solid line is a 2° average, and the gray shading represents the 1σ variation. The dashed curve is the average dust opacity referenced to the right-hand axis. (From Withers and Pratt, 2013.)

recorded during geomagnetically quiet times may have a source from tropospheric convection (Vadas and Liu, 2009) or even ocean waves (Djuth *et al.*, 2010).

9.2.5 Dust storms on Mars

Dust storms provide an episodic forcing on the Martian upper atmosphere and ionosphere that has no clear terrestrial analog. Dust storms can be regional or global and are generally unpredictable, although there is a “dust storm season” which covers the second half of the Martian year (Zurek *et al.*, 1992; Kass *et al.*, 2013). During the first MGS aerobraking phase in 1997–1998 there was a notable regional dust storm which was associated with a factor of three enhancement in thermospheric densities (Keating *et al.*, 1998; Bougher *et al.*, 1999). This corresponds to a 10 km change in the height of a reference upper-atmosphere pressure surface (see Keating *et al.*, 1998). For subsequent dust-storm events (e.g., in July 2005) there was no upper-atmospheric sampling with accelerometers; however there were radio occultation data that observed changes in the F_1 ion peak height (Withers and Pratt, 2013). As we have discussed, for the photochemically controlled Mars F_1 ionosphere, perturbations in the neutral atmosphere are registered as changes in the peak height. Figure 9.15 shows that for the 2005 event, the dayside heights rose by 5 km.

9.3 Summary and outlook

As we have noted above, our understanding of the relative roles of neutral atmospheric effects on the ionospheres of both Earth and Mars are expected to

significantly improve in the 2015–2020 time frame. On Earth, the ICON mission will make simultaneous observations of both E- and F-region neutral wind variability and the F-region electric fields and particles to quantify the link between lower-atmospheric wave forcing and equatorial ionospheric variability. GOLD hopes to be able to image the variability of thermospheric and ionospheric weather from geostationary orbit. Both ICON and GOLD are slated for launch in 2017. On Mars, the MAVEN mission's overall goal is to understand the processes that have shaped the evolution of the Martian atmosphere, in particular its escape. As part of fulfilling that overall goal, we will have comprehensive global data on the composition of the neutral thermosphere and the ionosphere for the first time. We will finally have sufficient data on atomic oxygen from which to study the global variability of the O/CO₂ ratio. Thus we expect to observe much of the same dynamical variability that we see on Earth.