Solar Forcing of the Thermosphere and Ionosphere from below: Coupling via Neutral Wave Dynamics

Dave Fritts and Ruth Lieberman, NorthWest Research Assoc., and Bela Fejer, Utah State Univ.

Executive Summary

Neutral atmosphere – ionosphere coupling represents a key link in the response of the terrestrial environment to solar forcing. Neutral motions, primarily waves arising at lower altitudes or in the auroral zone, enter the thermosphere and ionosphere, where they achieve large amplitudes and play major roles in accounting for the neutral and plasma structure and variability. Key questions include:

1. How does the global spectrum of thermospheric waves evolve with altitude, time, and solar forcing?
2. How do waves contribute to the mean structure, dynamics, and electrodynamics of the thermosphere and ionosphere?
3. How do waves drive thermospheric and ionospheric variability?

Understanding these coupling processes in the terrestrial environment is essential to improved Space Weather prediction, mitigating communication and navigation impacts of ionospheric scintillations and plasma bubbles, predicting orbits and orbital decay for spacecraft and space debris, and gaining insights into similar coupled dynamics in other planetary atmospheres.

Such a mission requires global measurements and could be addressed, depending on instrumentation, with one or several satellite platforms. The instrument capabilities exist and such a mission could be performed within the scope of an EX or MIDEX Explorer budget.

1. Introduction

Observations, modeling, and theoretical studies over ~50 years have suggested that neutral atmosphere influences account for many of the most prominent features of the ionosphere and its spatial and temporal variability. These include the E and F region dynamos, large-scale polarization electric fields, currents, plasma drifts, and the equatorial ionospheric anomaly (e.g., Crain et al. 1993; Heelis 2004; England et al. 2006; Immel et al. 2006). Neutral atmosphere influences have also been implicated in smaller-scale dynamics, i.e., traveling ionospheric disturbances, equatorial plasma bubble seeding, and mid-latitude spread F, both under geomagnetically quiet and active conditions (McClure et al. 1998; Nicolls et al. 2004). Indeed, recent studies have provided compelling evidence that many of these ionospheric dynamics are influenced directly by neutral dynamics and “weather” arising in the lower atmosphere.

These ionospheric influences accompany various wave motions that arise due to solar energy inputs in the lower and middle atmosphere (~0-10 km and ~10-100 km, respectively) and in the thermosphere. Waves represent indirect rather than direct solar forcing of the thermosphere and ionosphere (TI) at higher altitudes. They induce very large neutral wind and density variations extending well into the TI due to the approximately exponential growth of wave amplitudes with decreasing mean density. These induce, in turn, large plasma density perturbations, plasma drifts, currents, and polarization electric fields that account for many of the major features of the ionosphere. They also appear to contribute to plasma instabilities and their finite-amplitude manifestations that can have significant and adverse effects on human activities. Indeed, these indirect effects of solar forcing may contribute enormous neutral and plasma variability in the TI specifically because their sources are at lower altitudes, where neutral densities are as much as $10^{10}$ larger than in the F layer. This allows waves having apparently insignificant amplitudes at their source altitudes to nevertheless achieve large amplitudes and effects in the rarified TI, especially under quiet geomagnetic conditions (e.g., Fejer and Schierless 2001).
The neutral waves that account for the transport of solar energy from lower altitudes into the TI include solar tides, planetary waves (PWs), and internal gravity waves (GWs). Migrating and non-migrating solar tides are global-scale waves that are excited by the diurnal cycle of solar radiation absorption in the troposphere and stratosphere, the longitudinal variations of this absorption accompanying the distributions of deep tropical convection and stratospheric ozone, and in situ forcing in the thermosphere (e.g., Hagan and Forbes 2002; Oberheide and Forbes 2008a). Nonlinear interactions among the tides, PWs, and GWs where wave amplitudes are large contribute additional modes and variability (e.g., Teitelbaum and Vial 1991; Ortland and Alexander 2006; Fritts et al. 2010). These tidal forcing mechanisms yield a complex superposition of tidal modes that are apparent in the mesosphere and lower thermosphere (MLT) and which also extend into the E and F regions (e.g., Forbes et al. 2009; Hagan et al. 2009; Oberheide et al. 2009; Talaat and Lieberman 2010). Tides in the MLT and above also appear to provide the major modulation of GW propagation and fluxes of energy and momentum into the TI, the dominant forcing of the E and F region dynamos, large perturbations of minor species and plasma, and the pre-conditioning of the ionosphere for various plasma instabilities (e.g., Fritts and Vincent 1987; Immel et al. 2006; Häusler et al. 2007, 2010; Lühr et al. 2007, 2008; Kudeki et al. 2007; Vadas 2007; Fritts and Vadas 2008; Oberheide and Forbes 2008b).

PWs arise in the lower and middle atmosphere through a variety of processes. Large-scale topography and land-sea temperature contrasts excite global-scale, quasi-stationary PWs that typically extend well into the stratosphere. Occasionally these lead to strong PW/mean-flow interactions during winter known as sudden stratospheric warmings (SSWs), primarily in the northern hemisphere, which have influences extending into the MLT and TI (e.g., Liu and Roble 2002; Goncharenko and Zhang 2008; Fejer et al. 2010). Other PWs arise due to barotropic and/or baroclinic instabilities of the large-scale circulation and thermal structure and exhibit spatial structures and periods resembling those expected for the various normal modes of the atmosphere. These motions are also seen to impose variability in the TI and modulation of plasma dynamics, though the mechanisms by which this coupling occurs remain unknown at present (e.g., Forbes and Leveroni 1992; Pancheva et al. 2006, 2008; Fejer et al. 2010).

GWs have many sources in the lower and middle atmosphere. These include topography, convection, jet streams and frontal systems, nonlinear interactions and instability processes, and local body forces arising due to GW dissipation and momentum deposition. Their importance in the MLT and TI derives from their amplitude growth with altitude, their strong interactions with, and modulation by, mean winds, tides, and PWs, and their significant transport of energy and momentum (e.g., Fritts and Alexander 2003; Vadas 2007; Fritts and Vadas 2008; Fritts et al. 2008). Their large vertical group velocities also allow them to achieve high altitudes despite their relatively smaller spatial scales. GW effects in the TI appear to include large neutral density and wind perturbations, corresponding large plasma densities, gradients, drifts, and polarization electric fields, significant neutral accelerations accompanying their momentum transport and deposition, and potential seeding of plasma dynamics and instabilities, especially equatorial plasma bubbles that may penetrate to altitudes of ~1000 km or higher.

Indirect solar forcing is recognized to be critical in defining TI structure, energetics, and variability, but our understanding of these dynamics and their effects is very primitive at present. New global observations are paramount to advance the understanding of TI dynamics and their implications for human activities. The remainder of this white paper provides examples of neutral wave influences in the TI and a brief overview of satellite mission requirements that would enable a quantitative understanding of these neutral atmosphere-ionosphere dynamics.
2. Examples of Neutral Atmosphere Influences in the TI

a. Tidal influences

Both non-migrating and migrating tides are expected to penetrate to high altitudes in the TI and to have significant dynamic and electrodynamic effects. The diurnal eastward-propagating wavenumber 3 (DE3) is seen to make the dominant non-migrating component contribution at high altitudes, where it is seen to dominate the neutral wind field near the dip equator (see Figure 1) and to induce significant longitudinal perturbations in the equatorial anomaly (Immel et al. 2006). Importantly, however, the maximum magnitude of this component is only ~10 m/s. Non-migrating tide signatures are also seen in exospheric temperatures inferred from CHAMP accelerometer data (see Figure 2) and suggest significant neutral and associated plasma densities at these altitudes.

Migrating tides were anticipated to be significant in the TI, based on TIME GCM simulations and limited ground-based measurements, but less attention has been paid to defining their component amplitudes to date. These modes may also be estimated from CHAMP wind data averaged over a yaw cycle, however (see Figure 3), and the results of such an assessment show diurnal and semidiurnal amplitudes that are substantially larger that the corresponding non-migrating modes (despite averaging over a yaw cycle), with terdiurnal components having comparable amplitudes.

Thus we might expect that the migrating tides in the E and F layers and above may have comparable or larger effects to those already highlighted for the non-migrating modes in previous studies. New, comprehensive, global observations of these tides in the TI are needed to quantify these tidal contribution to TI dynamics and energetics.

![Image of tidal influences](image)

Figure 1. Amplitudes of the diurnal (DE3, DW5), semidiurnal (SE2, SW6), and stationary (sPW4) components contributing to the wave-4 structure in the zonal wind along the dip equator (Häusler and Lühr 2009).

![Image of exospheric tidal temperatures](image)

Figure 2. Exospheric diurnal tidal temperatures from the CHAMP accelerometer (Forbes et al. 2009).

![Image of tidal wind](image)

Figure 3. Total zonal wind with LT and latitude (top left), mean zonal wind (middle left), and diurnal, semidiurnal, terdiurnal, and 6-hour migrating tides inferred from DELFT CHAMP winds averaged over a yaw cycle. Note that the three dominant migrating modes are all larger than the non-migrating modes (image by R. Lieberman).
b. PW influences

PWs likewise exhibit interesting, but not understood, influences in the ionosphere in various data sets. Figure 4 shows an example of vertical plasma drifts measured by the Jicamarca radar that exhibit 2-day and 16-day wave periodicities that become much more pronounced during a northern hemisphere SSW.

A second example shown in Figure 5 reveals a correlation between electrojet current and North Pole temperature at 32 km also occurring during the 2002-3 SSW event. These results suggest that even slowly-varying, global-scale motions have pronounced responses in the E and F layers that represent coupling from the neutral atmosphere into the ionosphere accompanying waves propagating upward from lower altitudes. The physical processes accounting for these responses, and their implications for ionospheric behavior and variability, are completely unknown at present, however, and new global measurements are required to define these influences quantitatively.

c. GW influences

As seen above for tides, GW perturbations in winds, temperatures, and electron densities are directly observable in many data sets. As a result, it has been possible to compare observed GW spatial scales and frequencies with those predicted based simply on viscous propagation theory (Vadas and Fritts 2005). The results of such comparisons show generally good agreement, despite the linear description of GW structure and propagation inherent in ray tracing theory (e.g., Earle et al. 2008; Vadas and Nicolls 2008, 2009). An example of the variations of GW vertical wavelengths with altitude seen in electron densities and observed with the MU radar in Japan is shown in Figure 6. This reveals an approximately exponential increase in vertical wavelength with altitude that is largely consistent with theory (Vadas 2007). A second example using the Poker Flat Incoherent Scatter Radar (PFISR) in Alaska for a multiple-beam experiment is shown in Figure 7 and

![Figure 4](image-url)

**Figure 4.** Vertical plasma drifts measured by the JRO ISR during the 2002-3 SSW (after 12/28) indicating strong plasma motions modulated by PWs at ~2- & 16-day periods (Fejer et al. 2010).

![Figure 5](image-url)

**Figure 5.** Electrojet current measured by CHAMP (top) and stratospheric temperature over the North Pole at 32 km (bottom, red). Note the strong ~16-day modulation in both data (Fejer et al. 2010).

![Figure 6](image-url)

**Figure 6.** Variations of GW vertical wavelength seen by the MU radar under quiet solar conditions (Oliver et al. 1997).
reveals fractional electron density perturbations exceeding 20%, with GW structures clearly penetrating to altitudes above ~300 km. Also seen here is clear downward phase progression, which is indicative of upward energy propagation for GWs, clearly pointing to a source in the lower atmosphere.

Finally, there is evidence of GW perturbations in electron densities also extending to significantly higher altitudes, despite the dramatic increases in viscosity with altitude. Two examples are shown in Figure 8. The left image is from measurements using the Arecibo ISR during strong auroral energy inputs at high northern latitudes; the right image is digisonde data obtained at Fortaleza, Brazil during the Spread F Experiment (SpreadFEx) (Fritts et al. 2009). The Arecibo measurements indicate penetration of aurorally-generated GWs to altitudes well above 500 km and apparent initiation of plasma instabilities above ~300 km. The electron density perturbations at right exhibit large vertical motions and correlate closely with the seeding of equatorial plasma bubbles, suggesting their role in the seeding process (Abdu et al. 2009). In both cases, GW phase structures exhibit downward propagation, confirming upward propagation from sources at lower altitudes.

3. A Canonical Neutral Atmosphere – Ionosphere Coupling Mission

a. Satellite parameters

Neutral waves obviously influence the ionosphere in many ways, their dynamics are expected to be functions of the wave sources, the environments through which they propagate, and the ionospheric conditions that they encounter (e.g., magnetic field orientation), and some of
the neutral dynamics impacting large- and small-scale wave responses extend to polar latitudes. Thus, a mission to explore these coupling dynamics in a comprehensive manner must perform global measurements of the neutral fields defining the various waves and the environments through which they propagate. Such a mission must also provide sensitivity to the spatial scales (horizontal and vertical) of the dominant wave motions and to their seasonal variations and geographic variability. Spatial scales are simple to define for mean structures, tides, and PWs, as limb measurements of horizontal winds and temperatures are sufficient for this purpose. Limb measurements are not sufficient to define GWs arising from lower atmosphere sources below ~150 km altitude, however, so these motions must be viewed in the nadir or sublimb in order to capture the important horizontal scales. These depend on the altitudes from which it is desired to track GW propagation, as the dominant GW scales increase strongly with altitude into the TI.

The desire for near-global coverage and a reasonable precession rate drives a trade-off between a high inclination (needed for global coverage with limb viewing) and a lower inclination (required for more rapid precession and tidal definition). This choice will also depend on the latitudes over which nadir or sublimb viewing of GWs is desired in order to provide sensitivity to the major source regions. Alternatively, these needs could be separately addressed with two satellites in different inclinations, and likely having different instrumentation.

b. Instrument measurement requirements

Horizontal winds should be defined in the TI, as well as at lower altitudes in order to define the environment through which the various wave motions must propagate en route to the TI.

Limb temperatures are required to define the mean thermal structure and the tides and PWs. However, nadir or sublimb viewing of suitable emissions will be necessary to provide the higher horizontal resolution required to define the GWs arising in the lower atmosphere, depending on the TI dynamics that are intended to be addressed.

References


