The Investigation of Gravity Waves in the Mesosphere / Lower Thermosphere and Their Effect on Sporadic Sodium Layer

Xuguang Cai

Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Physics Commons

Recommended Citation
THE INVESTIGATION OF GRAVITY WAVES IN THE MESOSPHERE /
LOWER THERMOSPHERE AND THEIR EFFECT ON
SPORADIC SODIUM LAYER

by

Xuguang Cai

A dissertation submitted in partial fulfillment of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
Physics

Approved:

Tao Yuan, Ph.D.
Major Professor

Mike Taylor, Ph.D.
Committee Member

Vince Wickwar, Ph.D.
Committee Member

Ludger Scherliess, Ph.D.
Committee Member

Ryan Davidson, Ph.D.
Committee Member

Mark R. McLellan, Ph.D.
Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2017
ABSTRACT

The Investigation of Gravity Waves in the Mesosphere / Lower Thermosphere and Their Effect on Sporadic Sodium Layer

by

Xuguang Cai, Doctor of Philosophy

Utah State University, 2017

Major Professor: Dr. Tao Yuan
Department: Physics

Utilizing the observation data of Na lidar in Utah State University, Logan, Utah, together with the Advanced Mesosphere Temperature Mapper and a numeric model, we investigate the gravity waves in the Mesosphere/ Lower Thermosphere and their effect on sporadic sodium layer.

The first project is the observation of the breaking evidence of gravity wave by Na Lidar and Advanced Mesosphere Temperature Mapper. We found that the breaking of the small-scale high-frequency gravity wave was caused by dynamic instability, which was generated by the combination of background and a two-hour, large-scale gravity wave. In this project, we made full use of the Na lidar full-diurnal cycle observation to
accurately separate the background and perturbations. Furthermore, the extracted
two-hour gravity wave amplitudes in meridional wind and temperature agreed well with
the gravity wave dispersion relationship.

The second project is the calculation of the gravity wave potential energy density
by Na lidar data and the Whole Atmosphere Community Climate Model at the Logan,
Utah location. The result showed that the least-squares fitting of the full diurnal cycle
observation data was effective in extracting gravity wave perturbations in a single
location. By comparing the result with the one obtained from the nocturnal observation
data, we found many differences, which also emphasized the importance of full diurnal
cycle observation.

The third project is the investigation of sporadic sodium layer above 100 km by
numeric modeling. The results suggest that large vertical wind induce the sporadic
sodium layer above 100 km. In addition, the strong eddy coefficient can change the
altitude and time distribution of Na density dramatically. The Na ion chemistry will
decrease the Na density above 100 km.

(156 pages)
PUBLIC ABSTRACT

The Investigation of Gravity Waves in the Mesosphere / Lower Thermosphere and Their Effect on Sporadic Sodium Layer

Xuguang Cai

Gravity waves in the atmosphere are the waves with gravity and buoyancy force as the restoring forces. Gravity waves will significantly impact the Mesosphere Lower / Thermosphere (MLT), and the breaking of gravity waves is the key factor to cause the cool summer and warm winter in the Mesopause region. Therefore, it is important for us to investigate gravity waves. In this dissertation, we mainly use USU Na lidar data to explore gravity waves in the MLT. The exploration is made up of two projects. One is the investigation of gravity wave breaking and the associated dynamic instability by USU Na Lidar and Advanced Mesosphere Temperature Mapper (AMTM). Another is the calculation of gravity wave temperature perturbations and potential energy density by least-squares fitting based on the data from the full-diurnal cycle observation of Na lidar. The sporadic sodium layer is the sharp increase of Na density in a small vertical range (several kilometers) above the Na main layer in the MLT. The formation of the sporadic sodium layer above 100 km remains unknown until now. Here we will investigate the mechanism of the generation of sporadic sodium layer using numeric modeling, including the effect of tide and gravity wave on the variation of Na density.
ACKNOWLEDGMENTS

I would like to thank Dr. Tao Yuan for guiding me and showing me the way in atmospheric science. I have learned a lot from him. I also want to thank Dr. Vince Eccles, who showed me the details of numeric modeling.

I give special thanks to my wife, Xingchen Wang, and my parents. Their support is very important for me to finish my Ph.D. study.

Xuguang Cai
CONTENTS

Page

ABSTRACT ....................................................................................................................... iii
PUBLIC ABSTRACT ......................................................................................................... v
ACKNOWLEDGMENTS ................................................................................................. vi
LIST OF TABLES ............................................................................................................. ix
LIST OF FIGURES ........................................................................................................... x
CHAPTER

I. INTRODUCTION ............................................................................................................. 1
   Mesosphere/Lower Thermosphere Dynamics .............................................................. 2
   USU Na Lidar ............................................................................................................. 5

II. GRAVITY WAVE THEORY ............................................................................................. 16
   Air Parcel Model ........................................................................................................ 17
   Linear Theory ........................................................................................................... 21
   Category of Gravity Waves in the MLT ................................................................. 25
   Gravity Wave Propagation ....................................................................................... 32
   Gravity Wave Saturation and Breaking ................................................................... 33
   Momentum Flux and Potential Energy Density .................................................... 36
   Summary .................................................................................................................... 39

III. JOINT STUDIES OF GRAVITY WAVE BREAKING BY USU NA LIDAR
   AND AIR GLOW INSTRUMENT .............................................................................. 40
   GW Breaking Evidence Observed by AMTM ......................................................... 41
   Na Lidar Observations ............................................................................................. 44
   The GW Breaking Altitude and Background Stability ........................................... 48
   Dynamic Instabilities .............................................................................................. 52
   The Two-Hour Gravity Wave ................................................................................... 58

IV. CALCULATION OF GRAVITY WAVE TEMPERATURE
   PERTURBATION AND POTENTIAL ENERGY DENSITY ........................................... 62
   Methodology .......................................................................................................... 64
   WACCM Data ........................................................................................................ 68
LIST OF TABLES

Table                                                                 Page

5.1. The Main Chemical Reactions related to $NO^+$ and $O_2^-$ from 95 km to 129 km ..... 94
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Mean zonal wind in a) winter, and b) summer at Logan, UT from HAMMONIA</td>
<td>4</td>
</tr>
<tr>
<td>1.2. (a) GW force balanced by the Coriolis force, and (b) scheme of Pole-to-Pole meridional circulation caused by the forcing related to GW breaking from Holton and Alexander [2000]</td>
<td>5</td>
</tr>
<tr>
<td>1.3. Normalized spectra of the Sodium D$_2$ line fluorescence</td>
<td>9</td>
</tr>
<tr>
<td>1.4. Calibration curve to get temperature and line-of-sight wind from the return signal (from Figure 2b in Krueger et al. [2015])</td>
<td>11</td>
</tr>
<tr>
<td>1.5. Main components of the USU Na lidar</td>
<td>12</td>
</tr>
<tr>
<td>1.6. The components and working principle of the Faraday filter</td>
<td>14</td>
</tr>
<tr>
<td>2.1. A schematic representation of air parcel model</td>
<td>19</td>
</tr>
<tr>
<td>2.2. Observed period and horizontal wavelength change of intrinsic period with -120 m/s horizontal wind</td>
<td>27</td>
</tr>
<tr>
<td>2.3. Example of a freely propagating gravity wave in the x-z plane, showing the directions of the phase and group velocities, which are perpendicular to one another</td>
<td>32</td>
</tr>
<tr>
<td>2.4. Example of gravity wave propagation, saturation and breaking</td>
<td>35</td>
</tr>
<tr>
<td>2.5. Scheme of the coplanar method</td>
<td>37</td>
</tr>
<tr>
<td>3.1 Time sequence of OH airglow images from 10:45 UT to 11:35 UT on the night of September 9, 2012, at a 10-min interval</td>
<td>43</td>
</tr>
<tr>
<td>3.2. Contours of the lidar-observed data (a-c) and reconstructed background (d-f) on the night of September 9, 2012</td>
<td>46</td>
</tr>
<tr>
<td>3.3. Time-altitude contours of (a) horizontal wind shear, (b) Brunt-Väisälä</td>
<td>46</td>
</tr>
</tbody>
</table>
frequency squared ($N^2$), and (c) Richardson number ($R_i$) .........................47

3.4. The OH layer peak altitude on the night of September 9, 2012 .................49

3.5. Perturbations (a-c) and corresponding spectra (d-f) from 84 km to 98 km ....53

3.6. Amplitude of Morlet wavelet spectra at 87 km for the lidar measurements
of (a) temperature perturbation, (b) meridional wind, and (c) zonal wind
perturbations ...........................................................................................................54

3.7. The profiles at 11:25 UT of (a) Richardson number ($R_i$), (b) Brunt-Väisälä
frequency squared ($N^2$), (c) meridional wind, and (d) zonal wind .................56

3.8. The mean horizontal wind (solid arrow), the two-hour wave propagation
direction (dashed arrow), and the tidal wind direction (dotted arrow) at
breaking time 11:15 UT, 87 km ...........................................................................59

4.1. (a) The lidar temperature measurements (black) and the reconstructed
background temperatures (red), and (b) the corresponding temperature
perturbations ............................................................................................................66

4.2. The five-year USU Na lidar September monthly climatology of (a) temperature
variance, and (b) potential energy density (PED) obtained from the LSF
method (black) and the Nightly Linear Fit (NLF) (red) ......................................67

4.3. Results of WACCM data (a) mean temperature, (b) temperature variance, (c)
potential energy density (PED), and (d) zonal mean wind .................................70

4.4. Lomb-Scargle power spectra density for the September temperature
perturbations from the lidar data .............................................................................76

5.1. The time and altitude change of nocturnal Na density (a) on UT day 160
and (b) on UT day 155 in 2013 observed by USU Na lidar .................................80

5.2. Nocturnal Na density variations at 102 km with full Na ion-molecular
chemistry (blue solid) and only the three-major ion-molecular reactions (red
solid), and those at 105 km (cross) .....................................................................85

5.3. 24-hour variations of temperature (top left), zonal wind (top right), meridional
wind (bottom left), and vertical wind (bottom right) ..............................................89
5.4. Time and altitude variations of the simulated (a) NO$^+$ and (b) O$_2^+$ in June ........96

5.5. The time and altitude change of Na and Na$^+$ under 1x tidal amplitude (a, b) and 5x tidal amplitude (c, d) .................................................................................................................. 98

5.6. Na density during GW appearance from 4:00 UT to 7:00 UT under 1x tidal amplitude ........................................................................................................................................... 102

5.7. Seasonal variations of (a) vertical wind semidiurnal tidal amplitudes, and (b) their vertical gradients at 105 km, 110 km, 115 km and 120 km. based on CTMT ....................................................................................................................................... 105

5.8. Results of the temporal and spatial variations of the simulated Na$_k$ without Na ion chemistry and using the strong tidal wave scenario with (a) 1x k$_{zz}$, (b) enhanced 50x k$_{zz}$ and (c) 1x k$_{zz}$ but no vertical wind................................. 106

7.1. The temperature perturbations of WACCM from September 8 to September 10 at Logan, UT after zonal wavenumber 6 removal ................................................................. 117

7.2. Temperature from UT day 313 to UT day 314 of 2011, measured by USU Na lidar........................................................................................................................................ 118
CHAPTER 1

INTRODUCTION

The Mesosphere-Lower Thermosphere (MLT, 50 km to 150 km) is the least understood region of the Earth’s atmosphere due to its complex dynamics and the scarcity of experimental observations. This dissertation investigates the behavior of gravity waves in the MLT, and their effects on the high-altitude sporadic Na layer, which is a dynamic feature that hints at complex neutral and ion dynamic processes in the thermosphere.

Chapter 1 introduces the general MLT dynamics and outlines the importance of understanding gravity wave activities in this region. This is followed by a description of the Na lidar at Utah State University (USU), which is the main instrument used to collect the observational data shown in this dissertation. The description of the lidar system includes its basic working principles and main components. Chapter 2 provides a summary of linear gravity wave (GW) theory. Chapters 3 and 4 discuss two GW-related projects that were carried out by the full diurnal cycle observations of USU Na lidar. Chapter 3 presents a case study on GW breaking, using simultaneous observations from the USU Na lidar and an Advanced Mesospheric Temperature Mapper (AMTM). In Chapter 4, we calculate the temperature perturbation and total potential energy density of GW perturbations. Chapter 5 describes a numerical study of the sporadic Na layer at altitudes above 100 km, which involves running a chemical model that accounts for the influence of tides and GWs.
Chapters 6 and 7 present the conclusions and future work of the three projects, respectively.

1. Mesosphere/Lower Thermosphere Dynamics

The Earth’s atmosphere consists of the troposphere (extending from the Earth’s surface to altitudes of about 10-16 km), the stratosphere (10-16 km to 50 km), the mesosphere (50 km to 80 km), the thermosphere (80 km to 500 km), and the ionosphere (60 km to 1000 km). The region comprising the mesosphere and lower thermosphere (MLT), between 50 km and 150 km in altitude, is the least-known region of the atmosphere, mostly due to a lack of experimental observations. The mesopause is the boundary between the mesosphere and the thermosphere and is the location where the coldest temperatures in the atmosphere occur, typically at altitudes about 85 km in summer and up to ~101 km in winter [Yuan et al., 2008]. The temperature in the MLT region also exhibits intriguing seasonal variations, with low temperatures in summer and high temperatures in winter. The mechanism of this counter-intuitive climatology relates to breaking gravity waves (GWs) and their influence on the mean flow in the MLT [Holton, 1983; Holton and Alexander, 2000].

GW is a type of atmospheric wave that has gravity as its restoring force. The basic theory and character of GWs will be presented in section 2. Most GWs are generated in the troposphere/stratosphere through severe atmospheric disturbances, such as thunderstorms
or jet streams. They propagate both horizontally and vertically upward. Some can reach the MLT region, or even the upper thermosphere. Their oscillation periods can vary between a few minutes to approximately the Earth’s intrinsic period \cite{Li et al., 2007; Cai et al., 2014; Yuan et al., 2016; Cai et al., 2017}. Their horizontal scales also vary dramatically from less than 10 km to several thousand kilometers, with vertical scales ranging from less than 10 to 40-50 km (based on five-year observations of USU Na lidar, \cite{Cai et al., [2017]}). These GWs can become saturated and break, thereby releasing their energy and momentum into the background atmosphere. The GW saturation and breaking processes depend on the background wind and temperature, as described in Chapter 2. In general, when the horizontal phase speed of a GW equals the horizontal background wind projected onto its direction of propagation, the GW encounters its critical level and breaks. However, recent studies have shown that, due to the time dependence of the background wind, the critical level becomes transient at a given location and will change location with the phase of the wind \cite[Fritts et al., 2015]{Fritts et al., 2015}. Therefore, there is less time for a wave and a critical level to interact at a given altitude. In addition, the changing background wind tends to change the ground-relative frequency such that a hard critical level does not occur, making GW activities near its critical level much more complex than those described by classical linear theory \cite[Holton, 1982]{Holton, 1982}.

For winter, the Hamburg Model of the Neutral Ionized Atmosphere (HAMMONIA) predicts the dominant zonal wind direction in the stratosphere and
mesosphere above Logan, UT, to be positive, i.e., eastward (see Figure 1.1). Zonal winds, therefore, predominantly filter out eastward propagating GWs, while those propagating westwards can reach the mesosphere, where they generate a westward drag force once breaking. In summer, the situation, including the direction of the drag force, is reversed, due to a reversal of the mean zonal wind direction. In addition, the zonal wave-drag force is balanced by the Coriolis force [Holton and Alexander, 2000] (see Figure 1.2a). The resulting meridional circulation flows from the summer hemisphere to the winter hemisphere (see Figure 1.2b). Based on the law of mass conservation [see Chapter 7 of Andrews et al. (1987) for details], the vertical mass flux gradient \( \frac{\partial \rho w}{\partial z} \) must be positive in winter hemisphere and negative in summer hemisphere. Since the air density,

![Figure 1.1](image-url)  

**Figure 1.1** Mean zonal wind in a) winter, and b) summer at Logan, UT from HAMMONIA. The pink arrows indicate the direction of the GW drag force. The red line marks the zero lines.
Figure 1.2. (a) GW force balanced by the Coriolis force, and (b) scheme of Pole-to-Pole meridional circulation caused by the forcing related to GW breaking from Holton and Alexander [2000].

\( \rho \), decreases with increasing altitude, \( z \), the vertical wind, \( w \), must be negative in winter hemisphere, and positive in summer hemisphere, creating a downward/upward movement of air, respectively, which in turn results in adiabatic cooling/warming in the mesopause regions of summer/winter hemispheres.

2. USU Na Lidar

When meteoroids enter the Earth’s atmosphere, they ablate and deposit metallic elements in their atomic forms (including Na, Fe, Mg, Ca, and K) in Earth’s atmosphere between 80 and 120 km [Vondrak et al., 2008]. Below 80 km, single metallic atoms easily react with other species to form molecules [Plane, 1991] due to the high air density, while above 120 km, most of them become ionized (due to the high densities of \( \mathrm{NO}^+ \), \( \mathrm{O}_2^+ \), and
O\(^+\)). Thus, metal layers are formed around the globe within the MLT, which can serve as tracers that can be observed using resonance fluorescence lidars. Na atoms can be detected using Na lidars, and are particularly well suited to serve as passive tracers for monitoring a range of atmospheric processes. This mesospheric Na layer undergoes strong seasonal variations [Gibson and Sandford, 1971; She et al., 2000; Yuan et al., 2012]. In winter, the layer can reach a thickness of up to 35 km (extending from about 75 km to 110 km) while the layer typically thins during the summer months (extending from 80 km to 105 km). Occasionally, a so-called sporadic Na layer (sharp increase in density in small vertical range) appears a few kilometers above the main Na layer [von Zahn and Hansen 1988; Clemesha, 1995]. The Na lidar is designed to emit laser pulses at the Na resonant frequency to induce laser-induced fluorescence (LIF) of Na atoms in the mesopause region. In general, when the laser frequency is tuned to a resonant line of Na, a Na atom absorbs a photon and is excited to a higher energy level (excited state). Upon returning to the ground state, it re-emits a photon. This is the so-called resonance fluorescence scattering process. For Na atoms that are well mixed with other atmospheric species, their motions and thermal characteristics are good representatives for the wind and neutral atmospheric temperature.

According to the Boltzmann-Maxwell principle, for thermal motion, when the temperature, \(T\), is constant, the number of atoms with a velocity between \(V\) and \(V+dV\) is...
\[ n(V)dV = N \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mV^2}{2k_B T}\right)dV. \]  

(1.1)

where \( m \) is the mass of atom, \( N \) the total number density and \( k_B \) is the Boltzmann constant.

For atoms with a velocity, \( V \), the lidar-detected frequency, \( \nu \), is Doppler shifted from the original frequency, \( \nu_0 \), as

\[ \nu = \nu_0 \left(1 - \frac{V}{c}\right), \]  

(1.2)

where \( c \) is the light speed. Rearranging equation (1.2) yields

\[ V = c(\nu_0 - \nu)/\nu_0 = c\Delta\nu/\nu_0, \]  

(1.3)

so \( dV = -cd\nu/\nu_0 \). Let the left side of equation (1.1) equal \( I(\nu) \), and \( N \sqrt{\frac{m}{2\pi k_B T}}dV = I_0 \), we have

\[ I(\nu) = I_0 \exp\left(-\frac{m c^2 (\Delta\nu)^2}{2k_B T \nu_0^2}\right), \]  

(1.4)

\[ \Delta\nu = -2k_B T \nu_0^2 \ln(I(\nu)/I_0)/mc^2. \]  

(1.5)

Equation (1.5) is known as the Doppler broadening, which increases linearly with temperature. Line-of-sight wind, \( U \), also introduces a Doppler frequency shift: \( \Delta\nu = \pm \frac{U}{\lambda_0} \), where \( \lambda_0 \) is the wavelength.

The sign depends on whether the atoms move toward (+) or away (-) from the detector. Therefore, the atmosphere Na fluorescence spectrum is affected by both Doppler broadening and Doppler shift. The Na D\(_2\) line, which has the largest resonance cross-
section, has ten transitions [Krueger et al., 2015] that must all be included when deriving the temperature and wind information. Thus, the Na fluorescence spectrum can be written as

\[ g(v) = \frac{D}{T \pi} \sum_{i=1}^{10} A_i \exp\left( -\frac{D}{T} (v - v_i - U/\lambda_0)^2 \right), \quad (1.6) \]

where \( D = \frac{m\lambda_0}{2k_B} = 497.62 K(ns)^2 \) for sodium atom with \( \lambda_0 = 589.159\text{nm} \). \( A_i \) is the relative line strength of the hyperfine transitions, \( T \) the temperature and \( v_i \) is the relative frequency. The spectrum thus changes with temperature and wind velocity. As shown in Figure 1.3, when temperature increases, the peak intensity corresponding to the Na D$_2$ line decreases and the linewidth becomes broader, while different wind velocities and directions simply shift the peak to higher or lower frequencies. From equation (1.6), we see that the Na spectrum contains both temperature and wind information, and it is impossible for us to solve for two unknown variables with only one known value. Thus, a three-frequency technique is developed to determine the temperature and wind simultaneously [Bills et al., 1991; She et al., 1992]. One frequency is chosen as the central peak within the spectrum (\( v_0 = 0.656\text{Ghz} \) relative to the center of mass of the Na D$_2$ spectrum), while the other two frequencies are located at the full width at half maximum (FWHM) of the main peak (\( v_+ = -0.656 + 0.63\text{Ghz}, v_- = -0.656 - 0.63\text{Ghz} \)), respectively.

Based on the established lidar equation [Gardner, 1989], the received photon
Figure 1.3. Normalized spectra of the Sodium D\textsubscript{2} line fluorescence. a) Radial velocity = 0 with temperature 150 K (blue), 200 K (red), and 250 K (green). b) Temperature = 200 K with radial velocity 0 (blue), -100 m/s (red), and 100 m/s (green).

Counts, N, can be determined from

\[ N(z) = \eta T_A^2 \left( \frac{E_p}{E_s} \right) \left( \rho \Delta z \sigma Na \right) \left( \frac{A_N}{z^2} \right) T_d + N_0 \quad (1.7) \]

N(z) is the photon counts detected in the altitude range of \( \left( z - \Delta z / 2, z + \Delta z / 2 \right) \).

\( \eta \) is the efficiency of lidar.

\( T_A^2 \) is the two-way transmittance between the bottom of the Na layer and the ground.

\( E_p \) is the energy of a laser pulse.

\( E_s = h \mu c / \lambda \) is the energy of a single photon.
\( \rho_{Na} \) is the number density of Na.

\( \Delta z \) is the altitude range bin.

\( \sigma_{Na} \) is the backscattered cross section of Na.

\( A_R \) is the aperture of the telescope that receives photons.

\[ T_d = \exp \left( -2 \int_{z_1}^{z_2} \alpha(z) dz \right) \] is the two-way transmittance through the Na layer.

\( N_b \) is the background noise.

From Equation (1.7), the only difference between photon counts at two frequencies is the effective Na backscattered cross section, which is related to the fluorescence spectrum.

\[ \sigma_{Na} = \frac{g_2}{g_1} \frac{\lambda^2}{8\pi} A_{21} g(v). \] (1.8)

\( \frac{g_2}{g_1} \) is the degeneracy ratio, which equals two for the Na D2 transition.

\( A_{21} = 6.289 \times 10^2 \text{s}^{-1} \) is the Einstein coefficient. From Yuan [2004], the ratios between the three frequencies are formed as follows.

\[ R_f(T, \nu) = \frac{N_+ + N_-}{2N_0} = \frac{g(v_+)}{2g(v_0)}. \] (1.9)

\[ R_v(T, \nu) = \frac{N_+ - N_-}{N_0} = \frac{g(v_+)}{g(v_-)}. \] (1.10)

\( N_0, N_+, \) and \( N_- \) are the number of photons at frequencies \( \nu_0, \) \( \nu_+ = \nu_0 + 0.63\text{Ghz}, \) and \( \nu_- = \nu_0 + 0.63\text{Ghz}, \) respectively.

A calibration curve can then be constructed by calculating the theoretical ratio
based on line-of-sight wind and temperature (see Li [2005] for details), as shown in Figure 1.4. If we use the real-measured photon counts to calculate the ratios, the corresponding temperature and wind can be found from this curve.

The USU Na lidar system is a narrowband resonance fluorescence Doppler lidar system operating at the Na D2 line (589.159 nm) with a 120 MHz FWHM laser pulse bandwidth. It is based on the Na lidar system developed at Colorado State University (CSU), which was relocated to USU in the summer of 2010, after two decades of operation in Colorado. The main components of the USU Na lidar are shown in Figure 1.5. The pulse of seed laser, with a wavelength of 589.159 nm, is generated by a Coherent Ring dye laser 899-21, which is pumped by a Nd³⁺:YVO₄ laser. Due to the high sensitivity of the

![Calibration Curves](image)

**Figure 1.4.** Calibration curve to get temperature and line-of-sight wind from the returned signal (from Figure 2b in Krueger et al. [2015]).
fluorescence spectrum to the changes in wind velocity (see Figure 1.3b), a sodium Doppler-free spectroscopy cell is set up to achieve precise laser frequency control. For example, 1MHz of frequency drift would result in error of 0.6 m/s in the wind measurement [Yue, 2009]. The pulse from the seed laser is then sent to the Acoustic-Optic-Modulator (AOM) system, which shifts the $v_0$ into $v^+$ and $v^-$ sequentially. The timing order of these three frequencies is controlled by a series of Transistor-transistor Logic (TTL) signals that are synchronized with the Nd$^+$:YAG laser firing. The key components of the AOM are two acoustic crystals attached to piezo-electric transducers, which generate a vibration at a frequency of 0.315 GHz, thereby periodically modulating the index of refraction of each crystal. Accordingly, the

**Figure 1.5.** Main components of the USU Na lidar.
incident light is diffracted with the frequency shifted by 0.315 GHz in a single pass and 0.63 GHz for a round-trip through one crystal.

Each of the three different frequencies from seeding laser are sent into a laser pulsed amplifier to be amplified and converted into laser pulses, which are then sent out to excite the Na atoms in the mesopause region. Through the aforementioned laser-induced fluorescence process, an excited Na atom emits a photon, which carries the temperature and wind information. These returning photons are collected by an astronomy telescope coupled to a Photo Multiplier Tube (PMT) via optical fiber. In order to remove the high background noise during daytime observations, we developed a Faraday filter [Chen et al., 1996; Yuan, 2004], as shown in Figure 1.6. Once the randomly polarized Na lidar return signal passes through the first polarizer, it becomes linearly polarized. If properly designed, the polarization of the signal at the Na D2 line is rotated by 90° upon reaching the end of the Na vapor cell, which is housed in a strong magnetic field. The rotation is due to the combination of the Zeeman Effect and Faraday rotation. The second polarizer is oriented perpendicularly to the first one, thus, only the Na D2 signals are able to pass through the filter with minimum losses, while photons at other frequencies are filtered out. Since the first polarizer induces a 50% loss of the Na signal, the next-generation Faraday filter will be designed to decrease the loss of signal.

Currently, the USU Na lidar is operating in a three-beam set up at night, with east and west pointing beams 20° off zenith and a north pointing beam 30° off zenith. During
Figure 1.6. The components and working principle of the Faraday filter.

The day, even with the removal of background photon noise by the Faraday filter, the Na signal-to-noise (S/N) ratio is still much lower than that of the nighttime measurements. The lidar is therefore operated in a two-beam (east and north) set up during daytime observation to distribute more power to each beam to obtain S/N ratio as high as possible. The lidar return signals are saved every minute during lidar operation with a line-of-sight bin size of ~150 m. In order to achieve high S/N ratios, data are binned into five-min and ten-min intervals when processing the nighttime observation results. As for daytime data processing, larger binning intervals such as 30-min and one-hour are used.

The capability to operate during hours of daylight is one of the most significant
features of the USU Na lidar. It allows the lidar to remain operational for a full diurnal cycle, which makes it possible to extract solar thermal tidal waves and mean field values accurately [Yuan et al., 2006, 2008]. By subtracting the derived mean and tidal waves from observation data, the residuals can be treated as perturbations induced by gravity waves at a particular location. Since the deployment of the Faraday filter in 2002, the Na lidar full-diurnal cycle observation data have been utilized to study not only tide [She et al., 2004; Yuan et al., 2006, 2008, 2012, 2013, 2014a], but also gravity waves including single GW case studies [Yuan et al., 2014b, 2016; Cai et al., 2014; Lu et al., 2015], as well as the potential energy density (Cai et al., 2017).
CHAPTER 2

GRAVITY WAVE THEORY

Gravity waves (GWs) and their associated spectra within the Mesosphere/Lower Thermosphere (MLT) are the key parameters for understanding the energy and momentum transfers between the upper and lower atmosphere. They have been known to control the MLT mean flow seasonal circulation and to generate the counter intuitive cold summer and warm winter in the mesopause region [Holton, 1983; Garcia and Solomon, 1985], along with some irregularities in the ionosphere [Vadas and Nicolls, 2009; Liu and Vadas, 2013]. Yet, after decades of experimental and theoretical investigations, their full characteristics and effects on the upper atmosphere are still not fully understood, mainly due to the random spatial and temporal features with horizontal wavelength changing from less than 20 km to several thousand km and periods varying from a few minutes to several hours. Most of the GWs are generated in the stratosphere or troposphere and propagate upward with growing amplitude to compensate for the decreasing atmospheric density to keep the conservation of wave energy during their upward propagation. A considerable portion reach the critical levels or become unstable and break in the upper atmosphere [Fritts and Alexander, 2003], where secondary waves can then be generated within the breaking region [Vadas et al., 2003; Smith et al., 2013], which further alter the atmosphere above the MLT. The breaking process of a GW deposits momentum and
energy into the mean flow, causing mean flow to accelerate in the direction of GW propagation, thereby changing the thermal structure and generating turbulence around the breaking region. In this section, we present the underlying theoretical concepts that govern GWs, including the definition of GW, and some of the mathematics behind linear theory (including GW dispersion relationship), GW propagation, saturation, and the momentum flux.

1. Air Parcel Model

   For an air parcel in static equilibrium, an external perturbation will cause the parcel to accelerate in the direction of the perturbation. When the perturbation involves vertical displacement, the air parcel is further influenced by both gravity and buoyancy forces. When being placed to a higher altitude, the buoyancy force around the air parcel decreases, due to the exponential decrease of air density/pressure, causing the air parcel to decelerate until it eventually reverses and starts to move downward again. During the course of its downward movement, the buoyancy force steadily increases again, and the parcel will overshoot its initial position due to inertia and sink to regions where the buoyancy force exceeds gravity. At some point, the positive buoyancy becomes sufficiently strong to reverse the air parcel’s downward movement causing it to move upward again. Thus, a vertical oscillation of the air parcel is generated. According to the classical wave theory, such vertical oscillations can propagate in the medium (air),
generating a wave motion. We call this type of wave GW because gravity is part of the restorin

To derive the equations of motion for this oscillation, we assume that the background atmosphere is stratified. An air parcel in its initial rest position has the same density, pressure, and temperature as those of the background. Another assumption is that the vertical motion of the air parcel is adiabatic (no heat exchange with the background) and hydrostatic (pressure within the air parcel and background are the same). The governing equation is thus following Liu and Liu [2011]:

\[ \rho = \rho_0, \quad p = p_0, \quad \frac{\partial p}{\partial z} = \frac{\partial p_0}{\partial z} = -g\rho_0, \quad \frac{d \ln \rho}{d \ln p} = \frac{1}{\gamma}. \]

The subscript zero denotes the background values, while the ones without subscript correspond to the air parcel. \( \rho \) and \( \rho_0 \) are the densities of the background atmosphere and the air parcel. \( p \) and \( p_0 \) are the pressure. \( z \) is the altitude and \( g \) is the gravity acceleration. Equation (2.2) is the adiabatic equation of the atmosphere. \( \gamma = c_p / c_v \) is called the adiabatic index with \( c_p \) and \( c_v \) the specific heat capacities at a certain pressure and volume, respectively. Because the motion of the air parcel is assumed adiabatic, its density will differ from that of the background once it starts to move away from its initial equilibrium altitude. Here, we introduce the important concept of the so-called potential temperature, which stands for the temperature that an air parcel would
have if it displaced adiabatically to a stand pressure level, $P_0$, from the current pressure level, $P$.

\[ \theta = T \left( \frac{P_0}{P} \right)^{R/c_p}. \]  

(2.3)

$R$ is the gas constant per unit mass, $P_0$ the reference pressure and $T$ the current temperature of the air parcel.

Figure 2.1 shows the movement of the air parcel within the reference frame moving with the horizontal background wind. Notice that the direction of the displacement, $ds$, is not vertical, but at angle $\alpha$ relative to the horizontal plane. Two forces act on the air parcel: gravity and the buoyancy force. Therefore, the equation of

\[ F_b = \text{buoyancy force} \]

\[ mg = \text{gravity} \]

**Figure 2.1.** A schematic representation of air parcel model. The blue square is the air parcel. The black double arrow represents the displacement, $ds$. The angle relative to horizontal direction is $\alpha$. The air parcel is influenced by two forces: gravity ($mg$) and buoyancy force ($F_b$).
motion for the air parcel can be described as [Liu and Liu, 2011]:

\[ \frac{d^2 s}{dt^2} = -g \frac{\rho - \rho_0}{\rho} \sin \alpha. \]  

(2.4)

Based on equation (2.1) and the ideal gas law, \( \rho = \frac{p}{RT} = \frac{p_0}{RT} \), the density of the air parcel at a new position becomes

\[ \rho = \rho_0(0) + \frac{dp}{dz} \cdot z, \]  

(2.5)

and the density of the corresponding background is

\[ \rho_0 = \rho_0(0) + \frac{dp_0}{dz} \cdot z. \]  

(2.6)

Substituting equations (2.5) and (2.6) into equation (2.3), we get

\[ \frac{d^2 s}{dt^2} = -g \left( \frac{dp}{dz} \frac{\partial p_0}{\partial z} - \frac{\partial \rho_0}{\partial z} \right) z. \]  

(2.7)

Together with equations (2.1) and (2.2), we obtain

\[ \frac{d^2 s}{dt^2} = -g \left( \rho \frac{\partial p_0}{\partial z} - \frac{\partial \rho_0}{\partial z} \right) z = g \left( \frac{\partial \ln \rho_0}{\partial z} - \frac{1}{\gamma} \frac{\partial \ln p_0}{\partial z} \right) z. \]  

(2.8)

For the ideal gas law \( p = \rho RT \), we calculate the natural logarithm on both sides and then take the partial derivative with respect to altitude, \( z \), yielding

\[ \frac{\partial \ln \rho}{\partial z} = \frac{\partial \ln p}{\partial z} - \frac{\partial \ln T}{\partial z}. \]  

(2.9)

Following the same approach, the equation (2.3) becomes

\[ \frac{\partial \ln \theta}{\partial z} = \frac{\partial \ln T}{\partial z} - \frac{R}{c_p} \frac{\partial \ln p}{\partial z} = \frac{1}{T} \left( \frac{\partial T}{\partial z} + \frac{g}{c_n} \right). \]  

(2.10)

Based on equations (2.7) and (2.8), we have
\[
\frac{\partial \ln \theta}{\partial z} = \left(1 - \frac{R}{c_p}\right) \frac{\partial \ln p}{\partial z} - \frac{\partial \ln \rho}{\partial z}.
\] (2.11)

Therefore, equation (2.7) can be rewritten as

\[
\frac{d^2 s}{dt^2} = -g \frac{\partial \ln \theta_0}{\partial z} z \sin a = -g \frac{\partial \ln \theta_0}{\partial z} ds \sin^2 a
\] (2.12)

By introducing the Brunt-Väisälä frequency, \(N\), with \(N^2 = g \frac{\partial \ln \theta_0}{\partial z}\), thus, equation (2.10) becomes

\[
N^2 = \frac{g}{T} \left(\frac{\partial T}{\partial z} + \frac{g}{c_p}\right) = \frac{g}{T} \left(\frac{\partial T}{\partial z} + \Gamma_a\right),
\] where \(\Gamma_a = 9.5 K/m\) is the adiabatic lapse rate in the MLT region and \(T\) is the temperature. Finally, the equation that represents the motion of the air parcel in the atmosphere can be rewritten as

\[
\frac{d^2 s}{dt^2} = -(N \sin a)^2 s.
\] (2.13)

Equation (2.13) is similar to the equation of motion of an oscillator in classical mechanics, which suggests that the frequency of a GW is always less than the Brunt-Väisälä frequency.

2. Linear Theory

In order to understand the dispersion relation of the GW for each of atmosphere parameter (temperature, wind, and pressure), we will introduce some basic concepts of the atmospheric linear theory. We consider stationary atmospheric conditions, where the motions are isobaric and without friction. Therefore, the equations of motion can be written as (following Fritts and Alexander [2003])
\[
\frac{du}{dt} - f v = - \frac{1}{\rho} \frac{\partial \rho}{\partial x}, \quad (2.14)
\]
\[
\frac{dv}{dt} + f u = - \frac{1}{\rho} \frac{\partial \rho}{\partial y}, \quad (2.15)
\]
\[
\frac{dw}{dt} = -g - \frac{1}{\rho} \frac{\partial \rho}{\partial z}. \quad (2.16)
\]
\[
\frac{d \ln \rho}{dt} + \frac{\partial \tilde{u}}{\partial x} + \frac{\partial \tilde{v}}{\partial y} + \frac{\partial \tilde{w}}{\partial z} = 0, \quad (2.17)
\]
\[
\frac{d \ln \theta}{dt} = 0. \quad (2.18)
\]

\( \frac{d}{dt} \) represents the total derivative and \( P \) is the atmospheric pressure.

\( \rho = \rho_0 \exp[-(z - z_0)/H] \) is the air density as a function of altitude, \( z \), where \( \rho_0 \) is the air density at a reference level and \( H = \frac{RT}{mg} \) is the scale height. \( m \) is the atomic mass of air. \( f = 2\Omega \sin \phi \) is the Coriolis frequency with \( \Omega \) being the Earth’s rotation rate, and \( \phi \) is latitude. \( \theta \) is the potential temperature, which has already been explained in section 2.1. These equations can be linearized by applying the method of small perturbations, according to which any one of these atmospheric parameters can be decomposed into a mean and a small perturbation. \( q = Q + q' \), where \( Q \) stands for the mean and \( q' \) is the small perturbation. The GWs are thus treated as small perturbations of the mean state.

In addition, considering that the vertical wind is much smaller than the horizontal wind, we define the background wind field as \( \bar{u}, \bar{v}, \bar{\theta} \), where the mean vertical wind term
is set to be zero [Yuan et al., 2014a]. Conversely, the vertical gradients of density, potential temperature, and pressure are much larger than their horizontal counterparts are, and we therefore only consider variations in the vertical direction. After making these assumptions, the equations can be linearized as follows:

\[
\frac{Du'}{Dt} + w' \frac{\partial u'}{\partial z} - f v' + \frac{\partial}{\partial x} \left( \frac{p'}{p} \right) = 0, \tag{2.19}
\]

\[
\frac{Dv'}{Dt} + w' \frac{\partial v'}{\partial z} + f u' + \frac{\partial}{\partial y} \left( \frac{p'}{p} \right) = 0, \tag{2.20}
\]

\[
\frac{Dw'}{Dt} + \frac{\partial}{\partial z} \left( \frac{p'}{p} \right) - \frac{1}{H} \left( \frac{p'}{p} \right) + g \frac{\rho'}{\rho} = 0, \tag{2.21}
\]

\[
\frac{D}{Dt} \left( \frac{\theta'}{\theta} \right) + w' \frac{N^2}{g} = 0, \tag{2.22}
\]

\[
\frac{D}{Dt} \left( \frac{\rho'}{\rho} \right) + \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} - \frac{w'}{H} = 0, \tag{2.23}
\]

\[
\frac{\theta'}{\theta} = \frac{1}{c_s^2} \left( \frac{p'}{\rho} \right) - \frac{\rho'}{\rho}, \tag{2.24}
\]

\[
\frac{D}{Dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}
\]

is the linearized form of the total time derivative.

We further assume that the background velocity and Brunt-Väisälä frequency, \(N\), vary very slowly over a wave cycle in the vertical direction, and we can therefore safely neglect any background wind gradient terms in equations (2.19) and (2.20). As for the
GW itself, we assume the solutions to have the following form [Holton, 1992]:

\[
\left( u', v', w', \frac{\theta'}{\rho}, \frac{\rho'}{\rho} \right) = \left( \tilde{u}, \tilde{v}, \tilde{w}, \tilde{\theta}, \tilde{\rho}, \tilde{p} \right) \exp \left[ i(kx + ly + mz -\omega t) + \frac{z}{2H} \right].
\tag{2.25}
\]

Equation (2.25) describes a monochromatic wave perturbation with wavenumber \((k,l,m)\) and observed angular frequency, \(\omega\). By combining equation (2.25) with equations (2.19) to (2.24), we obtain

\[
- \omega \tilde{u} + f \tilde{v} + ik \tilde{p} = 0,
\tag{2.26}
\]

\[
- \omega \tilde{v} + f \tilde{u} + il \tilde{p} = 0,
\tag{2.27}
\]

\[
- \omega \tilde{w} + \left( im - \frac{1}{2H} \right) \tilde{p} = -g \tilde{p},
\tag{2.28}
\]

\[
- \omega \tilde{\theta} + \frac{N^2}{g} \tilde{w} = 0,
\tag{2.29}
\]

\[
- \omega \tilde{\rho} + ik \tilde{u} + il \tilde{v} + \left( im - \frac{1}{2H} \right) \tilde{w} = 0,
\tag{2.30}
\]

\[
\tilde{\theta} = \frac{\tilde{p}}{c_s^2} - \tilde{\rho}.
\tag{2.31}
\]

Here, \(\hat{\omega} = \omega - ku - lv\) is the intrinsic frequency, namely the angular frequency that would be observed in a reference frame moving with the background wind \(\left( \tilde{u}, \tilde{v}, 0 \right)\), and

\[c_s = \frac{c_e}{\rho} \tilde{p} \] is the speed of sound. The equations (2.26) to (2.31) can be combined and converted into one single equation for \(\tilde{w}\), the vertical wind perturbation of GW. By letting the real and imaginary parts of this equation go to zeros, we obtain the following
two equations (see Appendix A for a more detailed derivation of equations (2.32) and (2.33)).

\[
\frac{g}{c_s^2} = \frac{1}{H} \frac{N^2}{g},
\]  

(2.32)

\[
\omega^2 \left( k^2 + l^2 + m^2 + \frac{1}{4H^2} - \frac{\omega^2 - f^2}{c_s^2} \right) = N^2 (k^2 + l^2) + f^2 \left( m^2 + \frac{1}{4H^2} \right). 
\]  

(2.33)

Equation (2.33) supports both acoustic and gravity waves. Notice that the acoustic waves have compressibility in their propagation direction, while GWs we discuss here are free propagating without compression. We can therefore remove the acoustic waves by letting \( c_s \to \infty \), which yields the important dispersion relationships of GWs

\[
\omega^2 = \frac{N^2 (k^2 + l^2) + f^2 (m^2 + \frac{1}{4H^2})}{k^2 + l^2 + m^2 + \frac{1}{4H^2}}. 
\]  

(2.34)

3. Category of Gravity Waves in the MLT

As discussed earlier, the whole GW frequency spectra are between the Brunt-Väisälä frequency and Coriolis frequency. For example, the longest GW period allowed would be the inertial period (inverse of the Coriolis frequency). One way to characterize GWs is to categorize them based on their intrinsic frequencies relative to the Brunt-Väisälä and Coriolis frequencies.

As stated earlier, the intrinsic frequency can be treated as the GW frequency within the reference frame moving with the background wind. For convenience, we will
assume that the GW is propagating in the meridional direction for the remainder of this discussion. The intrinsic frequency then becomes

\[ \hat{\omega} = \omega - lv. \]  \hspace{1cm} (2.35)

If the background meridional wind, \( v \), was southward (\( v<0 \)), then the intrinsic frequency would always be larger than zero. However, if \( v \) was positive (northward), and if there was no limitation of the horizontal (meridional) wavenumber, the intrinsic frequency could theoretically become negative, which is of course impossible in the real atmosphere. We must therefore set \( \omega - lv > 0 \) and have \( \lambda_y > \frac{2\pi v}{\omega} \). This inequality shows that if the background wind is negative, there exists a minimum value for the horizontal wavelength \( (\lambda_y) \) of GW. It also suggests that GWs with smaller periods have shorter horizontal wavelengths than GWs with longer periods.

Based on equation (2.35), a GW with a certain intrinsic frequency, \( \hat{\omega} \), can be seen oscillating at different observed frequency, \( \omega \), depending on the horizontal wind and the horizontal wavelength of GW. Assuming that a meridional wind velocity of -120 m/s, which is extremely large for background wind in the MLT region [Larsen, 2002], observed periods between 5 to 1080 minutes, and horizontal wavelengths that range from 20 km to 1500 km with a step size of 5 km, we can determine the expected range of the associated intrinsic period. According to Swenson et al. [2000], GWs with horizontal wavelengths of less than 20 km can easily be reflected, and therefore cannot propagate too far upward from their initial altitudes. This is the primary reason why we set the
lower limit to 20 km in this simulation. We also limited the horizontal phase speed of GWs to be less than 250 m/s, since GWs with horizontal phase speeds larger than 250 m/s cannot propagate in the MLT (Vadas and Liu, 2009). With the setting of wavelength and the observed period above, the range of the associated intrinsic period is calculated and showed in Figure 2.2. The white area in Figure 2.2 corresponds to the region of “forbidden” horizontal wavelengths. It clearly suggests that, for GWs with short observed periods (high frequencies), the possible horizontal wavelengths are smaller than those GWs with longer periods. In addition, the difference between the observed and intrinsic periods could be significant, depending on the background wind velocity. For the example of our high background wind velocity of -120 m/s, a GW with an intrinsic period of 1080 minutes and a horizontal wavelength of 1500 km is observed with a

![Figure 2.2](image)

**Figure 2.2.** Observed period and horizontal wavelength change of intrinsic period with -120 m/s horizontal wind.
period of only 175 minutes. It should be noted that the results in Figure 2.2 are for our specific scenario with extremely strong background winds, and the differences between the observed and intrinsic periods under more realistic conditions are expected to be much less than in this simulation.

3.1. High-Frequency GWs

We call the GWs with intrinsic frequencies $\omega >> f$ high-frequency GW (HFGW). Here, we want to clarify that the >> or << signs imply inequalities of at least one order of magnitude. The dispersion relationships for HFGW are as follows:

$$\frac{\nabla^2}{\omega} = \frac{(k^2 + l^2)N^2}{k^2 + l^2 + m^2 + \frac{1}{4H^2}},$$  \hspace{1cm} (2.36)

$$w = \left(\frac{-m - \frac{i}{2H}}{\nabla^2} - \frac{1}{N^2 - \omega}\right)\omega p,$$  \hspace{1cm} (2.37)

$$p = \left(\frac{\omega}{ifl + \omega k}\right) u = \left(\frac{\omega}{\omega l - ifk}\right) v,$$  \hspace{1cm} (2.38)

$$\bar{T} = -i*w*N^2*T/g*\omega.$$  \hspace{1cm} (2.39)

As mentioned earlier, HFGWs have short horizontal wavelengths. Thus, the airglow imager, which has a fixed side of view, can easily detect them. The studies of HFGWs have typically been focusing on three aspects. One is the investigation of the spectra of certain single or multiple GWs [Alexander and Teitelbaum, 2007; Yue et al.,
2009; Vadas at al., 2012]. The other is the climatology of HFGWs in a certain location
[Medeiros et al., 2003; Pautet et al., 2005; Hecht et al., 2007; Yue et al., 2010; Li et al.,
2011]. The third is the study of momentum flux (MF) in the MLT region [Acott et al.,
2011; Fritts et al., 2014] because HFGWs are associated with the majority of MF [Ern et
al., 2004], which will be discussed in section 6.

3.2. Medium Frequency Gravity Wave

GWs with \( N >> \omega >> f \) is called medium-frequency GW (MFGW). After we
substitute \( N >> \omega >> f \) into equations (2.26) to (2.31), we can further simplify them as

\[
\hat{\omega} = N \left| \frac{k_h}{m} \right|, \tag{2.40}
\]

\[
\omega = \frac{k_h}{m} u_h, \tag{2.41}
\]

\[
\bar{T} = -i * u_h * N * \bar{T}/g, \tag{2.42}
\]

\[
\bar{p} = -\frac{m}{k_h} w_c_h, \tag{2.43}
\]

\[
\bar{p} = \left( \frac{\omega}{\omega_k} \right) u = \left( \frac{\omega}{\omega_l} \right) v. \tag{2.44}
\]

Due to the simpler dispersion relationships, MFGW has been widely used in the
various theoretical studies on GW dynamics, such as the effects of GW breaking on the
background atmosphere [Fritts, 1984; Holton and Alexander, 2000] and the calculations
of MF [Ern et al., 2004], which will be discussed in sections 4 and 6. Experimentally,
Cai et al. [2014] used collaborative observations of the USU Na lidar and Advanced Mesosphere Temperature Mapper (AMTM) to study a MFGW with two-hour period and obtained horizontal phase speed of 180 m/s, and a vertical wavelength of 17.3 km. They found that the amplitudes of this MFGW in zonal and meridional wind fields agreed with the dispersion relationship described in equation (2.44). Lu et al. [2015] studied a MFGW with period of one hour by the simultaneous Na lidar measurements in Boulder, CO (40°N, 106.5°W) and Logan, UT (41.7°N, 112°W) in combination with the AMTM in Logan, UT. They found that this MFGW, captured by both Na lidars, had a phase speed of 180 m/s, which was identical to that of the two-hour wave reported in Cai et al. [2014].

Yuan et al. [2016] studied the dissipation of a unique spectrally broadened MFGW packet, whose period was changing from ~1.5-hour to one-hour as it was traveling upward through the MLT region.

### 3.3. Inertial Gravity Wave

GWs with intrinsic frequency close to the Coriolis frequency are called inertial GW (IGW) and their polarization relationship can be modified as

\[ \hat{\omega}^2 = N^2 \frac{k^2 + l^2}{m^2} + f^2, \]  
\[ \hat{T} = -i \hat{w} \ast N^2 \ast \hat{T} / \hat{g} \hat{\omega}, \]  
\[ \hat{u} = \left( \frac{i \hat{\omega} k - f \hat{l}}{i \hat{\omega} l + f \hat{k}} \right) \hat{v}, \] 

(2.45) \hspace{1cm} (2.46) \hspace{1cm} (2.47)
\[ w = \left( -m - \frac{i}{2H} N^2 \right) \omega p. \]  

(2.48)

The generation of IGW is related to the geostrophic adjustment, which is the process that restores the geostrophic balance, the equilibrium between the horizontal pressure gradient and the Coriolis force while ignoring friction, acceleration, and vertical velocity (see Holton [1992] for details). IGWs can easily be observed in the troposphere and the stratosphere by balloon and rocket soundings [Sato, 1994]. The vertical wavelength is usually quite small, about two to ten km in the lower stratosphere and 10-20 km in the mesosphere. However, its horizontal wavelength can be several thousand of kilometers [Fritts et al., 1988; Riggin et al., 1995; Nastrom and Eaton, 2006; Yamamori and Sato, 2006], and therefore IGWs are always observed far from their sources in horizontal direction. The ratio of the IGW vertical to horizontal group velocity is equal to \( \left( \frac{\omega^2 - f^2}{N} \right)^{1/2} \), which means that the IGW vertical group velocity is very small compared to the horizontal component and therefore difficult for IGWs to propagate from the stratosphere/troposphere to the MLT region. There have been only a handful of reported observations of IGWs in the MLT [Li et al., 2007; Lu et al., 2009; Nicolls et al, 2010; Chen et al., 2013].
4. Gravity Wave Propagation

As described earlier, GWs propagate vertically and horizontally in the atmosphere, and these propagations are influenced by atmospheric conditions, such as wind and temperature. Figure 2.3 shows a simplified scheme of a monochromatic GW propagating in the x-z plane. It should be noted that the phase velocity of this GW is perpendicular to the group velocity because a GW is a transverse wave. From equation (2.34), we can derive the vertical wavenumber of GWs as follows:

\[
m^2 = k_h^2 \left( N^2 - \frac{\omega^2}{\omega - f^2} \right) - \frac{1}{4H^2},
\]  

(2.49)

Because the IGW has difficulty reaching the MLT region, as mentioned in the last

**Figure 2.3.** Example of a freely propagating gravity wave in the x-z plane, showing the directions of the phase and group velocities, which are perpendicular to one another.
section, we only consider the GWs with \( \hat{\omega} > f \). In addition, \( \hat{\omega} = k \omega (c - U_h) \), where \( c \) is the horizontal phase speed, and \( U_h \) is the background horizontal wind that is projected onto the direction of GW propagation. Thus,

\[
m^2 = \frac{N^2}{(c - U_h)^2} - k_h - \frac{1}{4H^2}.
\]  

(2.50)

Notice that most of the GWs propagate upward with \( m < 0 \). Based on the GW linear theory presented in section 2, we can calculate the group and phase velocities of GWs as

\[
(c_{gr}, c_{ph}, c_{ge}) = \left( \frac{\partial \omega}{\partial k}, \frac{\partial \omega}{\partial m}, \frac{\partial \omega}{\partial m} \right) = \left( \hat{u}, \hat{v}, 0 \right) + \left( \frac{k \left( N^2 - \omega^2 \right), l \left( N^2 - \omega^2 \right), -m \left( \omega - f^2 \right)}{\hat{\omega} \left( k^2 + l^2 + m^2 + \frac{1}{4H^2} \right)} \right)
\]

(2.51)

5. Gravity Wave Saturation and Breaking

To satisfy the principle of conservation of energy, amplitude of a GW must increase exponentially with altitude, in order to compensate for the exponential decrease of air density. However, this growth is limited by the atmospheric instabilities in the MLT. When GW amplitude becomes sufficiently large to cause \( N^2 < 0 \), the GW becomes unstable. This process is called reaching saturation level, which prevents the GW amplitude from increasing any further, and causes them to decrease as their altitude increases beyond the saturation level, as shown by the left part of Figure 2.4. The scenario where \( N^2 < 0 \) is called convective instability because an oscillation can no longer be sustained. The critical level (CL) is defined as the altitude where GW
horizontal phase speed equals to the background horizontal wind velocity projected onto the propagation direction of GW. When a GW encounters its CL, it breaks and all the wave features disappear and evolve into a chaotic turbulence feature. This can be seen in equation (2.50), where $m^2$ becomes infinite (as vertical wavelength becomes zero) when approaching the CL. The GW energy and momentum deposition start to occur during the decaying phase of the GW saturation, as shown in the right part of Figure 2.4. Based on Fritts [1984], the vertical wind perturbation above saturation level is given as

$$w' = w_s \exp(-mi(z - z_o)),$$

(2.52)

$$mi = \frac{1}{2H} - \frac{3}{2} \frac{\partial \bar{u}}{\partial z} \left(\frac{u - c}{u - c}\right).$$

(2.53)

Equation (2.52) shows that the vertical wind decays above the saturation altitude, and the decay factor, $mi$, can be written as equation (2.53). $w_s$ is the vertical wind amplitude at saturation altitude, $z_s$.

It is important to note that the aforementioned convective instability is based on the assumption that the background wind shear can be ignored. However, in the real atmosphere, when large amplitude atmospheric waves exist (IGWs and tidal waves), the wind shear can be greatly enhanced [Yuan et al., 2014b] and may cause the dynamic instability. Therefore, the Richardson number is introduced and defined as

$$Ri = \frac{N^2}{(\partial U_h / \partial z)^2},$$

(2.54)

where $U_h$ is the horizontal wind velocity. When $0 < Ri < 0.25$, dynamic instability
Figure 2.4. Example of gravity wave propagation, saturation and breaking. (Left) Schematic illustrating of the increase in gravity wave amplitude and saturation due to convective instability, where $z_s$ and $z_c$ indicate the saturation and critical levels, respectively. (Right) Change of the corresponding momentum flux with altitude (Figure 9 from Fritts [1984]).

occurs. Since the convective instability requires the $N^2 > 0$, while the dynamic instability can occur for $N^2 > 0$, the dynamic instability should take place earlier than their convective counterpart, which has been demonstrated by Zhao et al. [2003].

As another aspect of the GW saturation process, vigorous eddy diffusion process is generated near and above the saturation level [Fritts, 1984]. The eddy diffusion coefficient during this process is given by Fritts [1984],

$$D = \frac{k(\overline{u-c})^4}{N^3} \left[ \frac{1}{2H} \frac{\partial \overline{u}}{\partial z} \right].$$  \hspace{1cm} (2.55)
6. Momentum Flux and Potential Energy Density

When a GW saturates and breaks, it releases the stored energy as the momentum flux (MF), causing the background flow to accelerate in the direction of the propagation of GW. This parameter is highly important for the development of the climatological numeric models, since it can be utilized to parameterize global effects of GW [Holton, 1983]. It is therefore critical to calculate the MF of GW to a high level of accuracy.

There are two methods of calculating MF from the ground-based measurements. The first is the so-called coplanar method proposed by Vincent and Reid [1983]. It requires two coplanar beams pointing vertically upward at angles $\theta$ and $-\theta$ from the zenith, as shown in Figure 2.5. The velocity perturbations measured by each beam can be divided into two components:

$$v(\theta, R, t) = w'(\theta, R, t) \cos \theta + u'(\theta, R, t) \sin \theta,$$  \hspace{1cm} (2.56)

$$v(-\theta, R, t) = w'(-\theta, R, t) \cos \theta - u'(-\theta, R, t) \sin \theta.$$  \hspace{1cm} (2.57)

Squaring both sides of the above two equations gives

$$v^2(\theta, R) = w'^2(\theta, R) \cos^2 \theta + u'^2(\theta, R) \sin^2 \theta + 2u'w'(\theta, R) \cos \theta \sin \theta,$$ \hspace{1cm} (2.58)

$$v^2(-\theta, R) = w'^2(-\theta, R) \cos^2 \theta + u'^2(-\theta, R) \sin^2 \theta - 2u'w'(-\theta, R) \cos \theta \sin \theta.$$ \hspace{1cm} (2.59)

By assuming the statistics of motions to be independent of their horizontal locations, we obtain

$$u'^2(\theta, R) = u'^2(-\theta, R) \sin^2 \theta = \overline{u'^2(z)},$$  \hspace{1cm} (2.60)

$$w'^2(\theta, R) = w'^2(-\theta, R) \sin^2 \theta = \overline{w'^2(z)},$$  \hspace{1cm} (2.61)
Subtracting equation (2.59) from equation (2.58) yields
\[ v^2(\theta, R) - v^2(-\theta, R) = 4u'w'\cos\theta \sin\theta. \] (2.63)

And the equation to calculate the zonal component of MF becomes:
\[ \frac{u'w'(z)}{2 \sin 2\theta} = \frac{v^2(\theta, R) - v^2(-\theta, R)}{2 \sin 2\theta}. \] (2.64)

The same procedure can be applied for calculating the meridional component of MF.

Coplanar method has been utilized to calculate the MF at a single location using data from radar [Murphy and Vincent, 1993, 1998; Fritts and Janches, 2008; Fritts et al., 2006, 2012] and Na temperature/wind lidar observations [Gardner and Liu, 2007; Acott, 2009; Acott et al., 2011]. Their results vary from several m²/s² to several hundred m²/s², depending on the geographic locations and resolutions in time and space.

**Figure 2.5.** Scheme of the coplanar method. For two beams (y-z or x-z plane) directing at angles \( \theta \) and \( -\theta \) off the zenith, the perturbations on each beam can be decomposed into two directions (line-of-sight wind (parallel to beam) and vertical to beam).
Ern et al. [2004] proposed another method based on GW-induced temperature perturbations. They used the polarization relationship from Fritts and Alexander [2003] to derive the MF equation for a single GW packet:

\[
F_{ph} = \left[1 - \frac{\omega^2}{N^2}\right] \times \left[1 + \frac{1}{2Hm^2}\right]^{-1} \times \left[1 + \left(\frac{f_{M}}{m\omega}\right)^2 \frac{1}{4H^2}\right]^{1/2} \times \frac{1}{2} \rho \frac{k_h}{m}\left(\frac{g}{N}\right)^2 \left(\frac{T}{\bar{T}}\right)^2.
\]  

(2.65)

\(\omega^2\) is the variance of temperature perturbations and \(\bar{T}\) is the background mean temperature. \(H\) is the scale height. \(\omega\) is the intrinsic frequency. \(m\) and \(k_h\) are the vertical and horizontal wavenumbers, respectively. The key parameters are the variance of temperature perturbations and the ratio of horizontal wavenumber to vertical wavenumber. Equation (2.65) suggests that it is the HFGWs and MFGWs that contribute the most to MF because of their much larger \(k_h/m\) ratios compared to that of IGWs.

This method has been successfully applied to satellite data to estimate the climatology of the MF below 80 km [Ern et al., 2004; Alexander et al., 2008; Preusse et al., 2008]. It has also been applied to local observations and investigations of the MF by single GW packet in the MLT region [Fritts et al., 2014; Yuan et al., 2016].

The underlying assumption of the coplanar method is that the perturbations detected by two lidar or radar beams, belong to the same GW packet. In reality, however, the horizontal distance between the two beams above 90 km altitude varies from 45 km to 60 km, depending on the pointing angles. Based on the discussion in section 2.3, this distance is larger than the horizontal wavelength of most, small-scale GWs with high
frequency [Fritts et al., 2014]. This method may therefore not be suitable for calculating the MF of the HFGWs. Nevertheless, one of the main advantages of this method is that it can provide an altitude profile of MF, which allows the body force, namely the vertical gradient of MF, to be calculated [Yuan et al., 2014a]. The temperature perturbation method, proposed by Ern et al. [2004], can calculate the temperature variance and mean background temperature precisely for a single GW, especially for small-scale GWs. However, the horizontal wave number provided by the airglow measurements is only available at one particular altitude [Yuan et al., 2016]. MF can therefore only be determined at a particular altitude, such as the OH level ~ 87 km, while the induced body force cannot be calculated.

7. Summary

This section provides some basic descriptions of GW theory in the MLT region. It includes the definition of GW, GW dispersion relationship, the categorization of GWs and the propagation of GW. The chapter also presents the saturation of GW and methods to measure the MF of the GW, which is an important atmospheric parameter to quantify the GW effects on the background atmosphere. With these basic theories on GW dynamics, we are able to explain various observed GW-related phenomenon observed by the USU Na lidar and other instruments.
CHAPTER 3

JOINT STUDIES OF GRAVITY WAVE BREAKING BY USU NA LIDAR AND AIRGLOW INSTRUMENT

The observations by Na Temperature/Wind lidar and airglow instruments, especially the Advanced Mesosphere Temperature Mapper (AMTM) [Pautet et al., 2014], are ideal to study breaking of gravity wave (GW) and the associated atmospheric instabilities [Swenson and Mende, 1994; Fritts et al., 1997; Hecht et al., 1997; Yamada et al., 2001; Franke and Collins, 2003; Li et al., 2005; Eijiri et al., 2009]. However, one of the major challenges is to extract the short-period GW modulations from the data of these experimental observations. Without full diurnal cycle observation, it is highly difficult to separate these GW-induced short-period perturbations from the large-scale background variations precisely, which mainly includes the mean fields and tidal waves. This difficulty brings considerable uncertainties in the exploration of atmospheric instability and GW dynamics, such as GW breaking. Compared with the scenarios in the stratosphere, these uncertainties become much more significant in the MLT. In this Chapter, we utilize the coordinated measurements from the Utah State University (USU) Na lidar and a collocated AMTM to give a comprehensive investigation on a GW breaking event, and the associated atmospheric instability, which occurred on the night of September 9, 2012. The lidar had been running for 38 hours continuously from
September 8, 2012 to September 9, 2012 (UT day 252 to UT day 253), which allows us to separate the background and the GW perturbations, along with the characterization of some of the strong gravity waves (GWs). With the assistance of AMTM, we can not only determine the GW breaking level and uncover detailed mechanism of the wave breaking but also provide a clear description of the roles of different atmospheric wave components in initiating instability layers. The content of this Chapter has already been published in Journal of Geophysical Research (JGR) (Cai et al., [2014]).

1. GW Breaking Evidence Observed by AMTM

The USU AMTM [Pautet et al., 2014] has been designed to measure the mesospheric OH (3, 1) rotational temperatures over a large area, centered on the zenith. This instrument uses a fast (f:1) 120° field-of-view telecentric lens system designed and built at the Space Dynamics Laboratory (SDL), Logan, Utah, and three 4" narrow band (2.5-3 nm) filters centered at 1523 (P12), 1542 (P14), and 1520 (BG) nm, mounted in a temperature-stabilized filter wheel. The detector is an infrared camera fitted with a 320x256 pixels InGaAs sensor, thermoelectrically cooled down to -50°C to limit electronic noise, and controlled through a USB port by a Windows computer. The exposure time for each filter is typically ten seconds, giving a temperature measurement for each of the 81,920 pixels every ~30 seconds. This imager can operate in the presence of aurora and acquires data in full moon condition. Two of these instruments have been
built so far. The first one has been in operation at the South Pole Station (90°S) since 2010 and the second one at the Arctic Lidar Observation for Middle Atmosphere Research in Northern Norway (69°N) since the winter 2010 and 2011. During the summer months, the later instrument has been brought back to Utah and deployed on the USU campus (41.7°N, 112°W), alongside the Na lidar (Pautet et al., 2014).

The AMTM was operating from 2:46 UT to 12:17 UT on September 9, 2012. The data were processed with a temporal resolution of 30 seconds, generating temperature measurements within the OH layer with uncertainty of 1 K. A wave, which appeared between 10:45 UT and 11:05 UT in the AMTM field of view, was propagating toward the azimuth angle of 332° (+/-5°) relative to north with an observed horizontal phase speed of 74 ± 5 m/s and a horizontal wavelength of 20 ± 3 km. It had an observed period of 4.5 minutes and was seen breaking between 11:05 UT and 11:35 UT, as shown in Figure 3.1.

The images of OH airglow emissions are presented at a 10-minute interval from 10:45 UT to 11:35 UT in the figure in order to match the temporal resolution applied to Na lidar data analysis. The red line marks the wave front before the breaking and the blue arrow, indicating the propagation direction of the wave. This wave propagated into the AMTM’s field of view from the bottom right corner at about 10:46 UT, and started to lose its waveform between 10:55 UT and 11:05 UT. It is worth noting that a ripple-like feature (marked by red circle) appeared at 10:55 UT ahead of the wave front, and was
Figure 3.1. Time sequence of OH airglow images from 10:45 UT to 11:35 UT on the night of September 9, 2012, at a 10-min interval. The red line stands for the wave before breaking, and the blue arrow is the propagation direction of the wave. The movement of ripple at 10:55, 11:05, and 11:15 are marked with a red circle. The yellow arrow is the direction of horizontal wind. The breaking, which induced the turbulence feature, intensified after 11:15 UT, which is marked by the pink arrow. The three lidar beams are marked in a green point in the first figure, the north beam is up, the west beam is left, and the east beam is right.

traveling northward until it disappeared around 11:15 UT. After 11:15 UT, the breaking became more vigorous and produced some turbulence, which is marked by the pink arrow. The turbulence covered almost the entire field of view of the AMTM (180 km × 144 km). The existence of an accompanying ripple structure hints at a possible dynamic instability region nearby, since the ripple’s direction of propagation was almost identical to the direction of horizontal winds (marked by the yellow arrow), which was determined by the lidar-measured zonal and meridional wind. The green dots in the first plot
represent the positions of the USU Na lidar beams at 90 km altitude. The top dot is the north beam, the left dot the west beam and the right dot the east beam. The lidar data provide information on the atmospheric background condition, which will be discussed in the next section. The presence of ripples during dynamic instabilities in the mesopause region is to be expected and their existence has been reported in previous studies [Li et al., 2005a, 2005b]. Apart from the direct evidence of GW breaking, the AMTM also observed several ripple events during the night, implying the possible existence of other dynamic instability layers during that night within the mesopause region.

2. Na Lidar Observations

The contour plots of the Na lidar-observed temperature and wind measurements are presented in the left column of Figure 3.2. Here, the lidar data were processed with 10-minute temporal and 2-km spatial resolutions. The figure suggests that the lidar-measured temperature and meridional wind are clearly dominated by a large-scale, coherent oscillation with a downward phase progression of 3 km/hour, likely a semidiurnal tide in nature, while the zonal wind structure shows no clear tidal structure at all. The semidiurnal tidal phase, which marks the tidal maximum modulation, is plotted for each lidar altitude in each contour. Based on the tidal wave dispersion relationship [Yuan et al., 2008], the zonal wind semidiurnal maxima occur in the morning and late afternoon in the mesopause region, thus bearing little significance during the nighttime.
The tidal results are based upon the outputs of the least-squares fitting that is applied on the lidar diurnal measurements. During the wave breaking period that was observed with the AMTM, a strong temperature increase was observed by the Na lidar within the altitude range of OH layer after 10:00 UT. This increase of temperature reached about 215 K around 11:00 UT, which appeared to be part of a two-hour wave modulation that was well separated from the tidal feature between 84 km and 90 km. This is in excellent agreement with the OH rotational temperature modulation that was observed by the AMTM during the same period, showing a similar two-hour wave modulation. Also from Figure 3.2b, the increase of temperature was accompanied by a large wind shear (40 m/s/km) in the meridional wind field at around 11:00 UT at altitudes of 85–88 km, which was superimposed on the meridional semidiurnal maxima. Over the same period, however, the zonal wind did not show any significant variation.

From the high-resolution Na lidar data, we calculated the shear of the zonal and meridional wind, as well as important atmospheric dynamic parameters, such as Brunt-Väisälä frequency squared ($N^2$) and Richardson number ($Ri$), by using the following equations:

\[
N^2 = \frac{g}{T} \left( \frac{dT}{dz} + \frac{g}{C^p} \right), \quad (3.1)
\]

\[
Ri = \frac{N^2}{\left( \frac{dU}{dz} \right)^2 + \left( \frac{dV}{dz} \right)^2}, \quad (3.2)
\]
Figure 3.2. Contours of the lidar-observed data (a-c) and reconstructed background (d-f) on the night of September 9, 2012. The black point and line in (a) temperature and (b) meridional wind is the phase profile of semidiurnal tide. The phase profile of semidiurnal tide of (c) zonal wind field is out of the night’s temporal range. The reconstructed background (d) temperature, (e) meridional wind, and (f) zonal wind are deduced by utilizing the mean fields and tidal parameters deduced from the linear square harmonic fitting algorithm.

where \( g = 9.5 \times 10^{-3} \, \text{m/s}^2 \) is the gravitational acceleration at the altitude of the mesopause region. \( C_p \) is the specific heat per unit mass at constant pressure, and \( T \) is the temperature. In the MLT region, the lapse rate is \( \frac{g}{C_p} = 9.5 \, \text{K/m} \). \( \frac{dT}{dz} \), \( \frac{dU}{dz} \), and \( \frac{dV}{dz} \) are the vertical gradients of temperature, and the shears of the zonal wind and the
Figure 3.3. Time-altitude contours of (a) horizontal wind shear, (b) Brunt-Väisälä frequency squared ($N^2$), and (c) Richardson number ($R_i$). Altitude range is from 84 km to 90 km, and time range is from 9:25 UT to 12:45 UT.

The variations of horizontal wind gradient during the latter part of night 253 are shown in the contour plots of Figure 3.3a, along with those of $N^2$ and $R_i$. A GW breaks when propagating into either an unstable region ($N^2 < 0$ for convective instability or $0 < R_i < 0.25$ for dynamic instability), or when encountering its critical level (CL) where the horizontal wind speed equals the wave’s horizontal phase speed in its direction of propagation [Fritts and Alexander, 2003]. Transient instabilities can also be induced by
the superposition of a small-scale GW on the tidal-modulated background, thus triggering wave breaking.

During the second half of night 253, there were two increases in $N^2$ at altitudes of 85–89 km that were roughly two hours apart, as shown in Figure 3.3b. The horizontal wind shear was also modulated by a strong two-hour wave after 9:30 UT with its maximum lagging that of $N^2$. However, the wave’s downward phase progression feature was disturbed after 11:30 UT below 87 km. In Figure 3.3c, the dynamic instability between 11:00 UT and 11:30 UT occurred near the maximum of horizontal wind shear (~40 m/s/km) near 86 km. Therefore, the two-hour GW played an important role in the atmospheric variations that occurred after 9:30 UT at altitudes of 84–90 km. This two-hour wave remained active throughout the rest of the night below 92 km, and can be clearly seen in the lidar’s subsequent daytime observations.

3. The GW Breaking Altitude and Background Stability

In order to locate the GW breaking altitude range precisely, the OH layer peak altitude during the breaking event has to be calculated to account for the layer’s possible vertical fluctuations, mainly due to the modulations of tidal waves and large-scale GWs. This is achieved by employing the approach presented by Zhao et al. (2005), as shown in Figure 3.4. For each time point, the lidar temperature profile is Gaussian height-weighted with a Hamming window of 9 km full width at half maximum (FWHM), which is a good
proxy for the OH layer thickness and shape (*Baker and Stair*, 1988). The center of the Gaussian profile was then moved upward from 82 km to 92 km. For altitudes below 90 km, it is necessary to truncate the lidar Gaussian window because of the low quality of lidar data below 85 km. Next, we compared the height-weighted lidar temperature results with the simultaneously measured AMTM temperature values, and chose the altitude with the smallest difference as the OH peak altitude. Figure 3.4 shows the results based on this algorithm during the night. Based on the 30-minute smoothing results, the peak altitude of the hydroxyl layer is seen to vary a lot during the night from around 83 km at the beginning of the observation to above 90 km by 12:00 UT. Between 3:00 UT and 6:00 UT, the peak of OH layer climbed up 1.5 km per hour. From 7:00 UT to 12:30 UT,

![Figure 3.4](image)

**Figure 3.4.** The OH layer peak altitude on the night of September 9, 2012. The black dots in the figure stand for the peak altitude deduced from the lidar observed temperatures with a 5-min temporal resolution. The red line is in a 30-minute smoothing.
the OH peak altitude was oscillating with a two-hour period wave with two peaks at 10:00 UT and 12:00 UT. The amplitude of the oscillation is estimated to be slightly less than 2 km. Such overall upward OH layer fluctuations throughout the night are very different from those of the previous similar studies [Reisin and Scheer, 1996; Hecht et al., 1998; Taylor et al., 1999; Pendleton et al., 2000; Zhao et al., 2005], which presented a clear downward trend due to either a diurnal tide or a tide superimposed with a short-period GW. From the fluctuation of the OH peak altitude, we can see that this two-hour GW activity was quite strong during that night from 10:00 UT to 12:00 UT and dominated variations near the OH layer altitude. During the time of the wave breaking (11:05 UT to 11:35 UT), the OH peak was pushed upward from 87 km to 89 km and this two-hour wave feature is consistent with the simultaneous Na lidar temperature results presented in Figure 3.2a and Figure 3.3c. In order to reveal the wave breaking mechanism, we therefore focus on the lidar temperature and wind data during this breaking event within this altitude range (83-90 km) to have a better understanding of the mechanism involved.

We first calculated the $N^2$ values around 11:00 UT from 85 km to 90 km using equation (3.1). From Figure 3.3b, $N^2$ is mostly positive from 9:25 UT to 12:45 UT, indicating a convectively stable environment. Thus, the observed GW breaking was unlikely triggered by convective instability. To have a good estimation of the GW propagation condition and the critical levels of the GWs on that night, the background
temperature and horizontal wind need to be precisely determined. We define this background as the large-scale variations due to the combination of the mean field and tidal waves. Here, a Least-Squares Fitting (LSF) [Yuan et al., 2006, 2010] is applied to the lidar measurements at each altitude throughout the entire 38-hour continuous data set:

\[
S = A_0 + \sum_{i=1}^{4} A_i \cos \left( \frac{2\pi t}{24} - \phi_i \right),
\]

(3.3)

where \( S \) can be either the wind or the temperature, \( A_0 \) stands for the mean field, \( i = 1 \) for the diurnal tide, \( i = 2 \) for the semi-diurnal tide, \( i = 3 \) for the terdiurnal tide, \( i = 4 \) for the quarter-diurnal tide, and \( \phi_i \) represents the tidal phase for the \( i \)th tidal component.

We calculated the reconstructed background temperature, meridional wind, and zonal wind on that night using equation (3.3), which are shown in Figures 3.2d, 3.2e and 3.2f, respectively, with a ten-minute time resolution. These results are then employed to calculate the horizontal velocity projected onto the GW propagation direction (334° relative to the north). Due to the high horizontal phase speed (74 ± 5 m/s) of the GW before breaking, which is much faster than the background horizontal wind projected onto the direction of the GW, its CL was not formed within the hydroxyl layer. On the other hand, as shown in Figure 3.3c, there was a dynamically unstable region from 11:00 UT to 11:30 UT between 84 km and 87 km, which matches the time and altitude ranges of the GW breaking event observed by AMTM in Figure 3.1. This, along with the
AMTM-observed ripple feature during the same time, indicates the essential role played by dynamic instability in this wave-breaking event.

4. Dynamic Instabilities

In this section, we investigate more closely the formation of the above-mentioned dynamic instability that triggered the GW breaking. To discover the dominant GW components on the later part of night on September 9, 2012, the GW perturbations (Figures 3.5a, 3.5b, and 3.5c) are obtained by subtracting the reconstructed background from the original measurements (ten-min temporal resolution). The two-hour perturbations are clearly visible in the temperature and meridional wind field, but much weaker in the zonal wind field, most likely due to its propagating direction (345° relative to the north). It is worth noticing that the meridional wind perturbation appears to be lagging behind the temperature perturbation by roughly 30 minutes. This is in good agreement with the polarization relationship of medium-frequency GWs [Fritts and Alexander, 2003], which shows that the temperature perturbation should lead the horizontal wind perturbation (in this case the horizontal wind is mostly in the meridional direction) by π/2. We will discuss this in more detail in section 5. The spectra of the GW perturbations in the temperature and horizontal wind fields show the existence of four large-scale GWs during that night, as shown by Figures 3.5d to 3.5f. A 1.5-hour GW and a two-hour GW are clearly visible from 84 km to 92 km in the temperature and
Figure 3.5. Perturbations (a-c) and corresponding spectra (d-f) from 84 km to 98 km.

meridional wind field. A three-hour GW is dominant from 89 km to 91 km, and 96 km to 98 km in the meridional wind field and 87 km to 90 km in the zonal wind field. A four-hour period GW appears from 88 km to 91 km, and 95 km to 98 km in the meridional wind field and from 84 km to 89 km, and 93 km to 97 km in the zonal wind field. The 1.5-hour and two-hour GWs are very strong in the meridional wind field, while the three-hour and four-hour GWs are more pronounced in the zonal wind field. To determine the variations of the dominant GWs during different parts of the night within the hydroxyl layer, we also calculated the Morlet wavelet spectra of T, V, and U [Torrence and Compo, 1998], as shown in Figure 3.6.
Looking at the Morlet wavelet spectra at 87 km in Figure 3.6, it is obvious that different GWs components dominated different parts of the night. For example, the temperature and meridional wind fields are dominated by GWs with periods between 1.5 and three hours during the second half of the night, especially the dominance of the two-hour modulation from 10:00 UT to 12:00 UT. The zonal wind field, however, experienced strong GW modulations with periods around three to four hours during the first half of the night. Through spectral analysis, we can further reconstruct the temperature and wind variations to include these major GW modulations and characterize
their roles of in the formation of the instabilities that likely were the cause of the observed wave breaking. From the Morlet wavelet spectrum, the 1.5-hour and two-hour GWs are the main perturbations during the second half of the night. Therefore, we fitted the lidar-measured temperature and wind perturbations with 1.5-hour and two-hour GWs and applied a two-hour sliding window, moving with a time step of 10-minute over the data between 8:45 UT and 12:45 UT. At each step, the amplitude and phase of each GW can be calculated, allowing for the reconstruction of the modulations from each GW. Then we combined the reconstructed background with these reconstructed GW perturbations to simulate the temperature and wind variations, as well as the Ri and N² during this event.

Figure 3.7a shows the Ri profiles at 11:25 UT that are associated with the instability layer below 90 km. These profiles include the Ri directly calculated through lidar measurements and the one to be calculated through the reconstructed background alone, and the one deduced from the combination of the reconstructed background and the major GW modulations. From this plot, we can clearly see that the Ri values calculated from the combination of the reconstructed background and the GWs are very similar to those deduced directly from the lidar profiles, indicating the success of the reconstruction. To investigate the GWs’ impacts on the atmospheric dynamics during this wave-breaking event further, we applied the same algorithm and calculated the profiles of N², the meridional wind (V) and the zonal wind (U) at 11:25 UT as well, and listed them.
Figure 3.7. The profiles at 11:25 UT of (a) Richardson number (Ri), (b) Brunt-Väisälä frequency squared ($N^2$), (c) meridional wind, and (d) zonal wind. The black lines correspond to the original data and the associated error. The red lines are for the reconstructed background data, and the blue lines are for the reconstructed background with reconstructed gravity waves. The purple line in (a) stands for $Ri=0.25$.

In Figure 3.7 (b-d). From Figure 3.7a, both Ri values are near or less than 0.25 at 85.5 km and 86 km. On the other hand, the Ri calculated from the reconstructed background alone, termed as background Ri, are much larger than 0.25, implying that the background is fairly stable. Also from the figure, the background $N^2$ does not change that much from 85 km to 88 km, and is very different from the observed $N^2$, while the reconstructed $N^2$ profile is similar to the observed $N^2$, indicating the GW effect on the convective
instability in this region. As the figure shows, after the major GWs were included, the \( N^2 \) values near the dynamic unstable region dropped significantly from \( 6.367 \times 10^{-4} / s^2 \) and \( 6.468 \times 10^{-4} / s^2 \) to \( 1.88 \times 10^{-4} / s^2 \) and \( 0.97 \times 10^{-4} / s^2 \), respectively. However, given the range of uncertainty, the atmosphere is still convectively stable near the wave breaking level. As for the wind shear, when only the background is taken into account, the reconstructed meridional wind shear was calculated as 14.57 m/s/km and 13.9 m/s/km at 85.5 km and 86 km, respectively. These values almost doubled to 27.05 m/s/km and 22.78 m/s/km once the major GWs were included. The zonal wind shear was less than 1 m/s/km, and, is therefore not expected to contribute significantly to the total horizontal wind shear. It is also worth noticing that the differences between both wind shears become much larger above 87 km most likely due to the growing of the amplitudes of GWs. Based upon the above detailed discussion on the GW impacts on \( N^2 \) and wind shear value, it is clear that large amplitude GWs played a significant role in the evolution of \( N^2 \) and Ri, and the associated atmospheric instability locally over a short time scale.

Looking back to the Figure 3.6, the 1.5-hour and two-hour GWs were very consistent and were the dominant GW components after 10:00 UT in both the temperature and meridional wind perturbations, while they were very weak in the zonal wind field. Therefore, these two GWs played crucial roles in the formation of the dynamic instability between 10:00 UT and 11:30 UT, which induced the breaking of the subsequent small-scale high frequency GW. There was another dynamic instability
region observed between 7:30 UT and 9:00 UT during the same night between 91 km and 93 km, due to the combination of background and a large-amplitude GW, likely a GW with a short vertical wavelength. In that case, the dominant GW component was a four-hour GW, which affected mostly the zonal direction. Unfortunately, this intriguing region is above the hydroxyl layer; thus, the behaviors of the GWs within this dynamic instability layer are unable to be measured by the AMTM.

5. The Two-Hour Gravity Wave

The harmonic fit showed that the 1.5-hour modulation was relatively weaker compared to the two-hour wave. Thus, the two-hour wave becomes particularly important in this study. In addition, this GW was observed by both the Na lidar and the AMTM and played a significant role in the observed wave breaking process. The comprehensive coverage in both the horizontal and vertical directions by the two instruments provided an opportunity for a detailed study of the full characteristics of this wave. As shown by Figures 3.3a and 3.3c, it generated very large wind shear in the meridional direction when superimposed on the semidiurnal tide, and decreased Ri to near or less than 0.25, initiating the breaking of a small-scale GW around 11:05 UT.

Figure 3.8 presents a direct view of the directions and magnitudes of the mean wind, the total tidal wind, and the horizontal wind component of two-hour wave wind at 87 km when the instability occurred. Since we can extract the tidal wind by LSF as
Figure 3.8. The mean horizontal wind (solid arrow), the two-hour wave propagation direction (dashed arrow), and the tidal wind direction (dotted arrow) at breaking time 11:15 UT, 87 km.

mentioned earlier, the sum of the extracted tidal zonal and meridional wind yields the total tidal wind at 11:15 UT. It shows that the tidal wind aligned with the mean horizontal wind fairly well, while the horizontal wind component of the two-hour wave was also in the similar direction to the tidal wind. Combined with the lidar horizontal wind measurement, we calculated that the intrinsic period of the two-hour GW was 2.45 hours with intrinsic phase speed of 148.72 m/s during the wave breaking time. Utilizing the deduced fitting parameters, we estimate its vertical phase speed of $2.4 \pm 0.3$ m/s, associated with a vertical wavelength of $17.3 \pm 2.1$ km, which is in agreement with the magnitude of the downward progression feature in the lidar’s GW perturbation
observations in Figures 3.5a and 3.5b. The AMTM data revealed a horizontal phase speed of $180 \pm 9$ m/s and a horizontal wavelength of $1230 \pm 61.5$ km. This wavelength is also in the same order of magnitude with the one associated with the IGW that had been reported at the same latitude in Fort Collins, CO by Li et al. [2007], which had a horizontal wavelength of $1800 \pm 150$ km. This IGW dissipated quickly and disappeared around 102 km, due to the filtering by its CL. However, in this case, this two-hour GW was still propagating beyond 98 km, most likely because of its rapid horizontal phase speed. It was also found that during the wave breaking process, its amplitude was 15 K in the temperature field and $>20$ m/s in the meridional wind field around 87 km. Its zonal wind amplitude was found to be $\sim 5.7$ m/s at the same altitude. According to the GW polarization relationship with medium-frequency approximation [Fritts and Alexander, 2003],

$$\tilde{w} = \frac{k_h}{m} \tilde{u}_h,$$

where $\tilde{w}$ and $\tilde{u}_h$ are vertical wind perturbation and horizontal wind perturbation, respectively. From the deduced horizontal wind amplitude (mainly in the meridional direction) and the above-calculated wavelength, we estimate the vertical wind perturbation to be roughly 0.35 m/s and lead the horizontal wind perturbation by 180°.

These results are in accordance with the observed oscillations of the hydroxyl layer’s peak altitude presented in Figure 3.4 and described in section 3.3. Based upon another polarization relationship of the medium-frequency wave [equation (26) in Fritts and Alexander, 2003], we have the relationship between the meridional wind perturbations
and zonal wind perturbations without dissipation as

$$\tilde{u} = \left( \frac{i \omega k - f l}{i \omega l + f k} \right) \tilde{v}. \quad (3.4)$$

A further simplification ($N >> \omega >> f$, and $l > k$, due to the wave propagation direction of 345°) gives $\tilde{u} \approx \frac{k}{l} \tilde{v}$, where $k$ and $l$ are zonal and meridional wave numbers, respectively. From the AMTM observation, this ratio is ~0.27, which gives the zonal wind amplitude of ~5.4 m/s. This is in excellent agreement with the lidar-observed zonal wind amplitude of this two-hour wave mentioned above, and explains why the meridional component is much larger than the zonal component. This simplified polarization relationship also implies that the zonal wind component and meridional wind component are mostly in phase with each other, which is also highly consistent with the lidar-observed phase relation (not shown) of the two-hour wave. Such good agreement with the medium-frequency GW for the case of no dissipation may imply that this two-hour wave experienced little dissipation during this transient dynamic instability.
CHAPTER 4

CALCULATION OF GRAVITY WAVE TEMPERATURE PERTURBATION AND POTENTIAL ENERGY DENSITY

Based on Ern et al. [2004], the gravity wave (GW) potential energy density (PED) is directly related to the GW momentum flux (MF) and the associated body forcing on the mean flow. The calculation of PED requires precise measurements of both slow varying background temperature modulated by large-scale tidal waves and the GW induced temperature perturbations through carefully detrending of the data. So far, experimental investigations on GW perturbations and dynamics mostly focus on the nocturnal small-scale individual events with short periods [Cai et al., 2014; Fritts et al., 2014; Bossert et al., 2015; Yuan et al., 2016]. Except a few studies in the polar region [Chen et al., 2013], the long period (period more than two hours) large-scale GWs that play critical roles in the Mesosphere and Lower Thermosphere (MLT) dynamics, are not well studied, leaving contributions from most of these waves unchecked. The ground-based experimental studies on climatological GW perturbations [Gardner and Liu, 2007; Rauthe et al., 2006, 2008] have been limited to nocturnal observations. In this Chapter, we utilize a total of 433 hours of continuous Na lidar full diurnal cycle temperature measurements acquired during boreal equinoxes (September) from 2011 to
2015 to derive the monthly profiles of GW-induced temperature variance, \( T'^2 \), and the associated PED. Operating at Utah State University (USU) (41.7°N, 112°W), these lidar measurements reveal severe GW dissipation near 90 km, where both \( T'^2 \) and PED drop to their minima (~ 20 K^2 and ~ 50 m^2s^2, respectively). The study also shows that GWs with periods three to five hours dominate the midlatitude mesopause region during the summer-winter transition. The fitting algorithm that we have developed in the previous chapter was modified to detrend the lidar temperature measurements and calculate GW-induced perturbations. It completely removes the tidal perturbations from the lidar temperature measurement and provides the most accurate GW perturbations for ground-based observations. To evaluate this new algorithm, we introduce the latest Whole Atmosphere Community Climate Model (WACCM) mesoscale simulation results, which generates global distribution and variations of GWs. This new and unique study covers the GW spectrum up to inertia period, providing the most comprehensive results of large-scale GWs from ground based observations. As the only investigation of GW based on full diurnal cycle observations, these results can be utilized to evaluate the current effort of GW simulations from general circulation models that can simulate impacts and evolution of GWs from the Earth’s surface up to the lower thermosphere. The content of this Chapter has already been published in Annales Geophysicae (ANGEO) (Cai et al., 2017).
1. Methodology

From Ern et al. [2004], the GW-induced PED can be calculated as follows:

\[
PED = \frac{1}{2} \frac{g^2}{N^2} \frac{T''}{\bar{T}^2},
\]

(4.1)

where \( g = 9.5 m/s^2 \) is the gravitational constant and \( \bar{T} \) is the mean temperature with tidal modulations removed, \( T'' \) is the variance of the temperature perturbation, \( T' \), which, in fact, is the fitting residual that will be discussed later. \( N^2 = \frac{g}{T} \left( \frac{d\bar{T}}{dz} + \frac{g}{C_p} \right) \) is the background Brunt-Väisälä frequency squared with the lapse rate \( \frac{g}{C_p} = 9.5 K/km \) in the mesopause region.

Similar to the methodology applied in the previous chapter, to calculate the GW-induced temperature perturbations at each lidar-measured altitude, the slow varying background temperature, \( T_0 \), has to be calculated and subtracted from the lidar data. This could be achieved through a two-step algorithm. First, we applied a least-squares fitting (LSF) algorithm (shown in equation (3.3)) that covers all the tidal periods to the multi-day of continuous lidar measurements to derive the mean temperature and tidal period perturbations, provided that tidal period inertia GWs did not exist or were insignificant. Second, by utilizing the derived mean temperature and tidal period perturbations from step 1, we reconstructed the background temperature, \( T_0 \), in the mesopause region using equation (3.3).
Considering the large day-to-day tidal variability, possibly due to tidal-planetary wave interaction [Palo et al., 1999], a 24-hour fitting window sliding across the multi-day lidar measurements at one-hour step is applied. The abovementioned LSF is conducted within each of the fitting windows to derive the tidal period perturbations that are regarded as the good representatives of solar thermal tidal waves. The tides, along with the simultaneously deduced mean temperature, $\bar{T}$, were utilized to reconstruct the slow varying background temperature, $T_0$. After detrending the temperature at each lidar altitude, the differences between the lidar temperature measurements and this reconstructed background temperature were treated as GW-induced temperature perturbations, $T'$. The temperature variance, $T'^2$, which is an important for the calculation of PED, can be calculated at the same time. The window then moved one hour forward to repeat the above process until it reaches the end of the data set. The mean monthly profiles of $T'^2$ and PED, representing the monthly mean GW activity, were obtained by averaging all September data for each of the five years from 2011 to 2015. Figure 4.1 shows an example of this new method of determining the GW-induced temperature perturbations. Figure 4.1a is the comparison between the hourly lidar temperature measurements during a lidar campaign from UT Day of Year (DOY) 252 to 255 (September 9 to 12) in 2015 and the reconstructed slow varying background temperature, $T_0$, at 89 km, 92 km, and 95 km. Figure 4.1b is the corresponding $T'$, which is the difference between the lidar measured temperatures and the reconstructed $T_0$. As
Figure 4.1. (a) The lidar temperature measurements (black) and the reconstructed background temperatures (red), and (b) the corresponding temperature perturbations. The data are at 89 km, 92 km, and 95 km from UT day 252 to 256 in 2015. The temperatures at 92 and 95 km are increased by 50 K and 100 K, respectively, and perturbations at 92 and 95 km are increased by 20 K and 40 K, respectively. The magenta lines represent the zero-point lines for the temperature perturbation.

shown in Figure 4.1a, the reconstructed $T_0$ from this new method fits the lidar temperature measurements extremely well, reproducing the large-scale, day-to-day variability of MLT temperature, highlighting the capability of this new algorithm to generate realistic background temperature and GW-induced temperature, which is the difference between the lidar measured temperatures and the reconstructed $T_0$. Finally, as discussed earlier, the climatological profiles of $T^{aw}$ and PED based upon the five-year lidar temperature observations are established for the autumnal equinox in the midlatitude mesopause region, as shown in Figure 4.2a and Figure 4.2b.
There have been several reports of the total GW-induced temperature perturbation based on nighttime lidar observation [Rauthe et al., 2006; Gardner and Liu 2007; Rauthe et al., 2008]. To compare the new results from the full diurnal cycle to those based solely upon nocturnal lidar measurements, we processed the USU Na lidar nighttime data, between UT 2:00 and UT 12:00, by following the algorithm from Gardner and Liu [2007]. The nocturnal results are presented alongside the full-diurnal cycle results in Figure 4.2. Here, the lidar measured nighttime temperature variations at each lidar altitude during the whole night were linearly fitted to calculate and later remove the
background temperatures, $T_0$. This is justified by the fact that the large-scale waves, such as tides and planetary waves, are all slow varying components in the time domain, and can therefore be treated as constant during the night. This was followed by another linear fit in the vertical direction of the resulting temperature perturbation profiles to calculate and remove the residual tidal biases from the first fitting attempt. Because tidal waves, especially the semidiurnal tide, usually have very long vertical wavelength compared to the altitude range of lidar observation, their effects could be further removed by this second fitting process. These resulting profiles from this approach, termed as Nightly Linear Fit (NLF), are presented in Figure 4.2, alongside those derived by the LSF on the lidar’s full diurnal cycle data. From the figure, the differences are most apparent below 92 km, where nighttime results are significantly larger than those derived from the diurnal cycle lidar data are. We will discuss the differences in detail in section 4.4.

2. WACCM Data

To evaluate this LSF algorithm from previous section and see whether the deduced GW perturbations can precisely represent that of the entire temporal spectrum, we introduce the latest high resolution WACCM data in September that are able to simulate global GW distribution and variations. Here the WACCM data are processed in two completely different approaches to generate two sets of GW perturbations. For the first approach, we processed the high-resolution outputs of WACCM at the USU location
with the same LSF tidal fitting algorithm discussed in the previous section, generating background \( T_0 \) and GW-induced temperature perturbations, \( T' \), and variances, \( T'^2 \), along with the associated PED profiles. We again use the two-step fitting algorithm (NLF) for nighttime WACCM data at Logan, UT to study the differences. In the second approach, we applied the fit to the hourly WACCM temperature data from longitudes 0° to 360° at every altitude and latitude of 41°N with zonal wavenumbers 0 to 10 and then removed all these small zonal number components and the zonal mean at 41°N from the initial WACCM data to obtain the fitting residuals. These removed small zonal wavenumber components are most likely induced by the major large-scale waves, such as tidal waves and planetary waves, including the large-scale waves due to tidal-planetary wave interactions. The background temperature, \( T_0 \), can then be reconstructed with these low wavenumber components and the zonal mean. The residuals obtained from this Zonal Wavenumber 10 Removal approach, henceforth termed ZW10R, were treated as GW-induced temperature perturbations, \( T' \). Finally, \( T_0 \), \( \overline{T} \) (zonal mean), and \( T' \) at Logan, UT were chosen to calculate \( T'^2 \) and PED profiles, as shown in Figures 4.3a, 4.3b, and 4.3c. This algorithm distinguishes GWs from tidal/planetary waves based on their spatial scale. Thus, due to having the complete 24-hour coverage in this WACCM simulation, the obtained results are the most reliable representatives of the GW-induced temperature perturbations and the related body forcing generated within this model simulation.
Figure 4.3. Results of WACCM data (a) mean temperature, (b) temperature variance, (c) potential energy density (PED), and (d) zonal mean wind. In (a-c), black, red and blue lines represent results calculated by the LSF, ZW10R and the two-step NLF, respectively. Red line in (d) is the zonal mean zonal wind on September 5 in WACCM and green line is the lidar-measured mean zonal wind.

3. Difference Between Lidar and WACCM Results

As shown in Figures 4.2a and 4.2b, the derived lidar-measured temperature variance and PED profiles for September share very similar vertical structure within the mesopause region between 84 km and 99 km. The errors represent the averaged goodness of fit (chi-squared divided by the difference between the number of sampling and fitting
parameters) in the temperature variance profiles and the PED calculation. From 84 km to about 89 km to 90 km, both slowly decrease to their minima, only to quickly increase again above 90 km. For example, the GW-induced temperature variance drops from ~ 50 K² at 84 km to ~ 20 K² near 90 km, but starts to increase fairly quickly going at higher altitudes, with values reaching almost 90 K² at 99 km. This feature was called node by Rauthe et al. [2006], who used a Potassium lidar and Rayleigh lidar to study GW activity in midlatitude from 30 km to 105 km. In a later study, Lu et al., [2009] reported the similar node feature in Hawaii. And several other single-location observation studies mention this feature. For example, Mzé et al. [2014] studied the seasonal variation of the GW PED in Haute-Provence Observatory (43.93°N, 5.71°E) using Rayleigh lidar, and argued that GWs were dissipating above 70 km during all seasons. Tsuda [2014] reported the occurrence of considerable GW attenuation between 65 km and 90 km based on the measurements by Middle and Upper Atmosphere (MU) Radar at Shigaraki, Japan (34.85°N, 136.10°E). All of these observations indicated that there might exist one or even several GW saturation regions between 50 km and 84 km. However, the node feature is not obvious for satellite observations [Ern et al., 2011], which only showed a monochromatic increase of PED with increasing altitude. We speculate that it may be caused by the intensive spatial smoothing of many GW structures in the process of the zonal averaging, leading to the exclusion of the part of the GW spectrum. The node feature could be interpreted through the simplified relationship between the GW
horizontal wind perturbation, $u'$, and its vertical wavenumber, $m$, proposed by Booker and Bretherton [1967], which indicates a dramatic increase of $u'$ and $m$ when the GW is approaching its critical level. However, to our knowledge, no theoretical simulation has been able to duplicate these observations, and therefore, no solid dynamic scheme exists.

Although neither $T'^2$ nor PED from WACCM show minima at 90 km that are present in observations, they do show local minima in the upper mesosphere (82 km in ZWR10 and 78 km in LSF, respectively). Based on the GW linear saturation theory, when approaching its critical level, a GW becomes unstable and experiences saturation, leading to a damping of the GW amplitude due to the dissipation process [Fritts and Alexander, 2003]. The node, or local minimum feature, is therefore likely associated with the location of the GW critical level, which in turn is determined by the local horizontal wind profile. Figure 4.3d shows the comparison between the lidar-measured eastward zonal wind in September, and the zonal mean zonal wind profile on September 5 from WACCM data. As shown in the figure, although two wind profiles are both eastward in the mesopause region, they exhibit quite distinct behaviors in the vertical direction. The lidar-measured zonal wind speed tends to decrease with increasing altitude, while the mean zonal wind from WACCM reverses direction from westward to eastward near 79 km and then further accelerates up to an altitude of 90 km. It is worth noting that, in the mesopause region below 88 km, the mean zonal wind measured by lidar is much larger than the zonal wind of the model. The differences in background wind amplitude and the
altitude at which the wind reverses, which is most likely due to deficient GW forcing in this simulation, probably results in the altitude difference of GW filtering between the simulation and observations. This difference prevents direct comparisons between the model results and those based on lidar measurements. However, more importantly, the comparison between the two sets of WACCM data can certainly help determine whether the LSF approach we are taking can provide the true total GW perturbations.

4. Difference Between LSF and NLF Results and Validation of LSF

As mentioned earlier, we compared the LSF results with those based on nocturnal data and the NLF algorithm, and found considerable differences below about 92 km in the lidar results, as shown in Figure 4.2. For example, near 86 km, the results deduced from the NLF were approximately 50% larger than those derived from LSF were. This is most likely due to the bias introduced by the diurnal tide, which peaks during equinox at midlatitude [Yuan et al., 2010] and could not be completely removed by the NLF algorithm. Presumably, the magnitude of the difference depends on the diurnal tidal amplitude at these altitudes. Figure 4.2 also showed that, although the NLF results were still mostly larger than those obtained from the LSF, these differences were considerably reduced above 90 km. When we applied both algorithms to the WACCM local outputs, both sets of WACCM profiles (Figures 4.3b and 4.3c) showed even larger differences: near 75 km and between 80 km and 92 km, the results from the NLF overestimated GW
perturbations by greater than or equal to 100%. Again, significant differences between 80 km and 92 km in the WACCM results were most likely related to the bias caused by the diurnal tide, while those near 75 km could have been induced by planetary waves, as it has indeed been reported that transient planetary waves could be quite active during the autumnal equinox [Liu et al., 2007]. Although the planetary wave effects had not been completely removed by the LSF, their impact was considerably decreased in the 24-hour fit window.

To validate the above-mentioned new LSF algorithm and the associated GW results, we presented both kinds of profiles that were derived from the WACCM data using the two different algorithms. As shown in Figure 4.3a, both WACCM $T^2$ profiles are very similar. The differences between the two were mostly smaller than the fitting uncertainties. Within the mesopause region, both of the temperature variances were less 20 K$^2$ near 80 km, but grow quickly and continuously above 80 km to values as large as ~ 50 K$^2$ near 95 km. However, above ~ 97 km, both profiles showed significant differences: values deduced from the ZW10R are considerably higher, close to 70 K$^2$ at 100 km, compared to those calculated from the LSF at the same altitude (~ 40 K$^2$). One possible explanation is the presence of GWs with periods near six hours and eight hours in the model, whose amplitudes grew significantly above 97 km. These tidal period GWs were removed by the LSF method, but remained in the perturbations deduced by the ZW10R. The associated PED profiles in Figure 4.3c behaved similarly to temperature variance
profiles except that above ~ 97 km, the differences were smaller than the fitting uncertainties. Intriguingly, there is no node structure in neither the LSF nor the ZW10R profiles. The mean temperature profiles (Figure 4.3a) deduced from these three algorithms were almost the identical, given the fitting uncertainty.

Overall, based on the comparisons between the two WACCM profiles in Figure 4.3, both the LSF and ZW10R algorithms generate very similar GW temperature perturbations and PED profiles. Therefore, both can be utilized with a high degree of confidence to investigate GWs. For the studies of ground-based observation without horizontal coverage, the LSF tidal removal algorithm is able to provide the most reliable GW results.

Figure 4.4 illustrated the Lomb-Scargle power spectra density for the temperature perturbations during each of the lidar campaigns in September between 2011 and 2015. The lidar temperature perturbations were derived from the LSF tidal removal algorithm. Here, only modulations with significant level larger than 50% are shown in the figure. The figure showed the dominance of long-period large-scale GWs’ modulations in the mesopause region with periods between three hours and five hours in almost all of the six campaigns, except the campaign from DOY 251 to 253 in 2012. This indicates that during the fall equinox at midlatitudes, GW perturbations in the mesopause region are mostly generated by this part of the GW spectrum. To confirm this conclusion, we reanalyzed the lidar data using a 15-minute time resolution and applied the same LSF
Figure 4.4. Lomb-Scargle power spectra density for the September temperature perturbations from the lidar data. The observation periods are a) 2011 (UT day 262-266), b) 2012 (UT day 249-250), c) 2012 (UT day 251-253), d) 2013 (UT day 262-264), e) 2014 (UT day 253-256), and f) 2015 (UT day 252-255). Only components with significance level larger than 50% are shown.

algorithm. This change extends the lowest detectable GW period from two hours to 30 minutes, thus covering more high-frequency components. The Lomb-Scargle results from these high-resolution data, again, exhibited the dominance of this part of GW spectrum in the midlatitude mesopause region. Unfortunately, GWs with periods between three to five hours are difficult to be detected by the airglow equipment due to their large scales and long periods, and therefore horizontal information is not available. On the other hand,
the lidar can only measure these GWs’ vertical wavelengths and periods, which is not sufficient to reveal the full picture. Nonetheless, the GWs with periods three to five hours may play an important role in the MLT region and even in the upper thermosphere because some of these GWs with large vertical wavelengths (as high as 40-50 km) have been observed by lidar. These GWs thus have high vertical phase speeds and can reach upper thermosphere (above 250 km) with little dissipation. In addition, some of the GWs with periods between three to five hours can cause dramatical variations of Na density. Future plans to investigate this type of GW will be discussed in Chapter 7, future work.
CHAPTER 5
INVESTIGATION OF HIGH-ALTITUDE SPORADIC NA LAYER

The sporadic sodium layer (Naₘ), especially the high-altitude Naₘ above 100 km in the E region (90-140 km), can be described as a sharp increase in Na density in the mesopause region with the layer’s vertical distribution of merely a few kilometers [von Zahn and Hansen, 1988; Clemesha, 1995]. Previous Naₘ studies have reached two main conclusions. First, Naₘ has strong seasonal variations with its rate of occurrence highest in summer and lowest in winter [Fan et al., 2007; Dou et al., 2013; Yuan et al., 2014c]. Second, both the Naₘ and sporadic E layers (Eₛ) show downward progression most of the time, implicating that their close relationship to the tidal and/or gravity waves. Compared to the well-studied Eₛ, the mechanism of the formation of Naₘ above 100 km is not understood at all. The most widely accepted hypothesis is the neutralization of the elevated Na⁺ within Eₛ [Daire et al., 2002; Williams et al., 2006, 2007; Yuan et al., 2014c]. This mechanism is quite feasible below 100 km, where the atmospheric density and pressure increase dramatically as the Eₛ descend towards lower altitudes. Indeed, in one theoretical Na chemistry model simulation, Plane et al. [1999] simulated an intense ion layer downward to altitudes below 100 km and observed the formation of a sporadic Na layer out of the Eₛ. However, the observed formation of Naₘ above 100 km in the thermosphere around the globe [Clemesha et al., 2004; Sarkhel et al., 2012; Gao et al.,
remains puzzling and unable to be reproduced in model studies. The existing Na ion-molecular chemistry theory seems to prohibit Na\(^+\) neutralization above this altitude, thus, other mechanisms must play significant roles to generate high-altitude Na\(_s\). The investigation on this eminent feature is not only important to the understanding of the dynamic and chemical processes it implicates but also enable the resonant fluorescence lidar to measure the needed neutral winds and temperature profiles in the E region directly [Krueger et al., 2015; Liu et al., 2016]. Intense Na\(_s\) has also been frequently observed by the USU Na lidar during summer months above 100 km over the years of its operating at Logan, Utah (41.7°N, 112°W) [Yuan et al., 2014c]. Figure 5.1 shows a couple of such examples captured by the lidar during Es/Na\(_s\) campaign in summer 2013. The one observed on UT day 160 (June 9) with peak density ~ 200/cm\(^3\) shows continuous downward propagation of the Na\(_s\) during the first half of the night. The other event on the night of UT day 155 (June 4), which had peak density over 300/cm\(^3\), was modulated by a gravity wave with period of about one hour.

So far, most of the experimental studies have been focusing on simultaneous observations of Na\(_s\) and E\(_s\) evolution [Kane et al., 2001; Williams et al., 2007; Yuan et al., 2013; Raizada et al., 2015], which can only provide qualitative evidence for a possible correlation between Na\(_s\) and E\(_s\), but cannot identify the full mechanism that links E\(_s\) with Na\(_s\) formation. By studying the seasonal variation of the vertical ion drift at midlatitude caused by tidal waves, Yuan et al. [2014c] provided a more quantitative analysis on the
Figure 5.1. The time and altitude change of nocturnal Na density (a) on UT day 160 and (b) on UT day 155 in 2013 observed by USU Na lidar.

variations of ion density in the lower E region in midlatitudes during summer and winter. This investigation concluded that the frequent occurrence rate of Es and Na, during summer months could be due to the dominance of the ion transport convergence at this time of year. However, they still did not address the complex dynamic and chemical details connecting these two sporadic features in the upper atmosphere. To investigate Na, formation in a more comprehensive manner, it is necessary to build a numeric model that includes both the chemistry and the dynamics of the lower E region.

In this chapter we utilized the classic chemical model of Na by Plane [2004],
combined with the dynamic transport effects induced by atmospheric tidal and gravity waves, to investigate the formation of these high altitude Na₈ in midlatitude. After we established the modified Na ion-molecular chemistry model, we first introduced tidal waves into the model to set up the standard background neutral winds, densities, and temperatures at midlatitude. By changing the tidal amplitudes, we studied the contribution of tidal waves to the formation of the Na₈. Later, a large amplitude gravity wave (GW) modulation was added to investigate the associated variations of Na₈. The role of the eddy diffusion coefficient in this process was also studied in this simulation.

1. Numerical Framework

*Carter and Forbes* [1999] modeled the spatial and temporal evolution of Fe and Fe⁺ in the lower E region above Arecibo (18°N, 67°W geographic; 30°N, 2°E geomagnetic) by using the following system of continuity equations for neutrals and ions:

\[
\frac{\partial N_M}{\partial t} = S_M + P_M - L_M - \nabla \cdot \left( N_M \vec{v} \right),
\]

where \( N_M \) represents the number density of either neutrals or ions, and \( t \) is time. \( S_M \) is the source function. \( P_M \) and \( L_M \) are the production and loss terms, respectively, due to chemical reactions and ionization processes. \( \vec{v} \) is the transport velocity, which can be either \( v_i \) for ion transport or \( v_n \) for neutral transport. Following this approach, this Na simulation is conducted in the altitude range between 95 km and 129 km. The source function, which is the metal ablation of the meteoroids entering the Earth, is adopted.
from the WACCM-Na study [Marsh et al., 2013]. The lower boundary is set to 95 km because Na ion-molecular chemistry dominates above this altitude [Cox and Plane, 1998].

The winds are assumed uniform in the horizontal direction within the grid. To validate the Na model, we ran the simulation on June 15 and at Logan, UT (41.7°N, 112°W), where the USU Na lidar is located. The continuity equation was solved for each major constituent using the method of finite difference.

There are some important aspects, which need to be considered in solving the equation. First, the flux ($N_M \vec{v}$) must be conserved and be one single variable. It cannot be divided into $v \frac{\partial N_M}{\partial z} + N_M \frac{\partial \vec{v}}{\partial z}$. The gradient of mass flux could be written as follows (we use n in place of N for convenience):

$$\frac{\partial n v}{\partial z} = n_{i+1/2}v_{i+1/2} - n_{i-1/2}v_{i-1/2}. \quad (5.2)$$

Then we use donor cell flux as follows:

$$n_{i+1/2}v_{i+1/2} = \min(n_{i+1}v_{i+1}, 0) + \max(n_i v_i, 0), \quad (5.3)$$

$$n_{i-1/2}v_{i-1/2} = \max(n_{i-1}v_{i-1}, 0) + \min(n_i v_i, 0). \quad (5.4)$$

Regarding the boundary conditions:

in the upper boundary, $n_{i-1/2}v_{i-1/2} = \min(n_i v_i, 0)$, \quad (5.5)

in the lower boundary, $n_{i+1/2}v_{i+1/2} = \max(n_i v_i, 0). \quad (5.6)$

For the second derivative of diffusion, we set it to zero in the upper and lower boundaries, and the derivative is written as:
\[
\frac{\partial^2 n}{\partial^2 z} = \frac{n_{i+1} - 2n_i + n_{i-1}}{(dz)^2}.
\]  
(5.7)

The second is the setting of the time step in the simulation. Since there are chemical reactions in partial differential equations (PDE), we should set the step based on the rules of chemical kinetics and stability of PDE. For each of the three species chemical reactions, the maximum time step was set to

\[
\Delta t = \frac{1}{k_i BC},
\]

(5.8)

where \( k_i \) is the reaction coefficient, and B and C are background chemicals. For example, for the reaction \( Na^+ + N_2 (M) \rightarrow NaN_2^+ \), \( Na^+ \) is the species need to be solved, and \( N_2 \) and \( M = N_2 + O_2 \) are the background chemical species. Therefore, the time step for this reaction must satisfy \( \Delta t < \frac{1}{k_i [N_2][M]} \). For two element reactions, such as \( Na + O_2^+ \rightarrow Na^+ + O_2 \), there is only one background chemical \( O_2^+ \) in this case, solving for Na). Furthermore, the time and spatial step must follow the rule of the stability of PDE

\[
\frac{v_i \Delta t}{\Delta h} < 0.5,
\]

(5.9)

where \( v_i \) is the transport velocity of the species, and \( \Delta h \) the spatial step. Based on all the requirements of time and spatial steps discussed above, we set time step as 1s, and the spatial step in vertical direction as 1,000 m.
2. Chemistry and Transport

2.1 Na Ion Chemistry

Based on the chemical reactions and associated reaction rates articulated in Plane [2004], we could reproduce the mesospheric main Na layer. However, since we focus on the region above 95 km where the ion chemistry predominates [Cox and Plane, 1998], only the reactions of ion chemistry and photochemistry were included in our model simulations. Furthermore, the reaction rate of the temperature-dependent Na\(^+\) neutralization process, which involves several molecular Na\(^+\) species, is considerably smaller than those of the Na ionization, even during the night when the photo ionization process is at the minimum. This indicates that the ionization processes dominate Na density variations in the E region, and there are very few Na\(^+\) ions converting to Na atoms during nighttime above 100 km. Without sacrificing any model accuracy, we can therefore simplify our simulation even further by including only the following two major Na ion-molecular chemical reactions, and one photochemistry reaction as follows [Plane et al., 2015]:

\[
\begin{align*}
Na + NO^+ & \rightarrow Na^+ + NO, \\
Na + O_2^+ & \rightarrow Na^+ + O_2, \\
Na + hv & \rightarrow Na^+ + e.
\end{align*}
\]

The reactions rates of R1, R2 and R3 are: \(2.7 \times 10^{-9}\), \(8 \times 10^{-10}\) and \(2 \times 10^{-5}\) [Plane, 2004; Badnell, 2006], respectively, with units of \(cm^3\text{molecule}^{-1}\text{s}^{-1}\). For the NO\(^+\) and O\(_2^+\)
densities, we use the main chemical reactions and associated reaction rates from Table 5.1 in Schunk and Nagy [2009] to solve equation (5.1). Figure 5.2 shows the variations of Na density at 102 km and 105 km when either the scheme that includes the full Na ion-molecular chemistry and photochemical reactions (Table 1 of Plane et al. [2015]) or the simplified one with only the three-major reactions is applied in the simulation. As expected, the density variations derived from the full Na ion-molecular chemistry scheme are slightly higher than those based on the simplified chemical scheme (less than 2/cm³ at 102 km). There is almost no difference at and above 105 km, where the pressure and atmospheric density become even less than those at 102 km. The over-night increasing of Na density is because of the decreasing of Na ionization due to the decreasing densities.

Figure 5.2. Nocturnal Na density variations at 102 km with full Na ion-molecular chemistry (blue solid) and only the three-major ion-molecular reactions (red solid), and those at 105 km (crosses).
of NO$^+$ and O$_2^+$, and weak photo ionization rates, along with the Na source function.

However, as mentioned earlier, Na ion-molecular chemistry alone is not adequate to describe the formation of these high-altitude Na$_s$, which clearly shows dynamical features, such as downward progression and periodic modulations by atmospheric waves, etc. The lidar-observed C-shape Na structure above 100 km associated with Na$_s$ hints at vigorous vertical transport of Na from the main layer [Kane et al., 2001; Clemesha et al., 2004; Sarkhel et al., 2012]. Thus, we must introduce the dynamical effects including vertical transport velocity due to neutral winds into this new Na chemistry simulation to understand the high-altitude Na$_s$ phenomenon. We have to point out that, due to the exclusion of Na neutral chemistry and uncertainties in Na source function and vertical transport processes (such as eddy diffusion coefficient), the simulation results are not self-consistent.

2.2. Transport

In the lower E region of midlatitude, the ion vertical transport is mainly determined by the ion vertical drift velocity that heavily depends on horizontal winds [Mathews, 1998]. In addition, due to the relatively higher atmospheric density in this region compared to the F region, eddy diffusion needs to be included when studying the transport of ions. Thus, based on the description by Plane [2004] and Yuan et al. [2013] the vertical drift velocity of ions in the lower E region of midlatitude is given by
\[ v_i = -\frac{V \cos I \sin I + U \gamma \cos I}{1 + \gamma^2} + k \left( \frac{1}{H} + \frac{1}{T} \frac{dT}{dz} + \frac{1}{N_M} \frac{dN_M}{dz} \right). \]  

(5.10)

The second item on the right is the vertical ion drift velocity. \( I \) is the magnetic dip. \( V \) and \( U \) are the background meridional and zonal wind, respectively. \( \gamma = \frac{mv_{in}}{eB} \) is the ratio of ion-neutral collision frequency to ion-gyrofrequency, \( m \) the ion mass, \( e \) the electron charge, \( B \) the magnetic field, and \( v_{in} \) is the ion-neutral collision frequency. The collision frequency is expressed through the following formula [Macleod, 1966; Banks and Kockarts, 1973]:

\[ v_{in} = \left[ \frac{1.8}{\sqrt{A(A+28)}} \cdot n_{N_2} + \frac{1.83}{\sqrt{A(A+32)}} \cdot n_{O_2} + \frac{1.80091}{\sqrt{A(A+16)}} \cdot n_O \right] \times 10^{-14}, \]  

(5.11)

\( A \) represents the atomic mass of metallic ion with \( A = 23 \) for \( \text{Na}^+ \). \( n_{N_2}, n_{O_2} \) and \( n_O \) are the densities of \( \text{N}_2, \text{O}_2, \) and \( \text{O}, \) respectively.

The second term in equation (5.10) is the transport caused by eddy diffusion. \( H \) is the scale height and \( T \) is the background temperature. \( k_{zz} \) is the eddy diffusion coefficient. For this simulation, \( k_{zz} \) was adopted from the Whole Atmosphere Community Climate Model (WACCM) [Garcia et al., 2007]. The values in the lower E region during summer are around 10 m\(^2\)/s\(^2\) to 20 m\(^2\)/s\(^2\). However, \( k_{zz} \) is a poorly known, but critical parameter for vertical transport, and the \( k_{zz} \) in WACCM may contain considerable uncertainty. Based on Carter and Forbes [1999], the transport velocity of neutrals, \( v_n \), in the midlatitude lower E region, is controlled by horizontal wind (here we
use meridional wind), tidal vertical wind, \( w \), and eddy diffusion.

\[
V_n = -w - k_{zz} \left( \frac{1}{H} + \frac{1}{T} \frac{dT}{dz} + \frac{1}{N_M} \frac{dN_M}{dz} \right).
\]  

(5.12)

In this simulation, we assumed the mean vertical wind to be zero and, thus, only wave-induced, vertical wind modulations are considered. Since we focus on the vertical direction, contributions from the horizontal winds were not included in our simulation.

Equation (5.12) shows that the vertical transport velocity of neutral species is a function of vertical wind, temperature and the density gradient. However, by comparing the magnitude of each term, the vertical wind component plays a pivotal role in vertical transport of neutral species in the static atmosphere.

2.3. The Description of Tidal and Gravity Waves

The effects of tides on Na density above 95 km include dynamic transport due to the tidal vertical and horizontal winds, as well as the tidal wave induced perturbations of atmospheric density, temperature, and pressure. Based on the Climatological Tide Model of Thermosphere (CTMT) (see model details in Oberheide and Forbes [2008] and Oberheide et al. [2011]) and monthly climatological means from the Hamburg Model of the Neutral and Ionized Atmosphere (HAMMONIA) [Schmidt, et al., 2006], we are able to produce the climatological time and altitude distributions of vertical and horizontal winds, atmospheric density and temperature between 95 km and 129 km by superimposing the diurnal and semidiurnal tidal wave components, in the shape of two
cosine functions, on the monthly mean. Figure 5.3 shows the 24-hour variations of temperature and winds generated by this hybrid approach. The semidiurnal tide modulation is the prominent feature in the lower E region above Logan, UT. The tidal outputs of CTMT have been compared with the Na lidar observations, and the comparison shows good agreement in the seasonal variations \[Yuan et al., 2014a\]. However, the model considerably underestimates the tidal amplitude due to the inherent temporal and spatial smoothing \[Yuan et al., 2014a\]. Thus, the scenario with the standard tidal output (1x standard tidal amplitude) is considered as the weak tide condition. To

Figure 5.3. 24-hour variations of temperature (top left), zonal wind (top right), meridional wind (bottom left), and vertical wind (bottom right). They are constructed by adding standard CTMT tidal components to the HAMMONIA monthly mean climatological values as cosine functions.
explore the effects of the increasing tidal amplitude on Na, we use 1x and 5x the standard CTMT tidal amplitude in our numerical simulation to represent the weak and strong tidal modulation scenarios, respectively. For example, the standard CTMT zonal wind semidiurnal amplitude near 110 km is ~ 20 m/s in June, while the 5x CTMT amplitude would increase to close to 100 m/s. Following the same approach, in this simulation, the atmospheric density variations are derived from the combination of CTMT relative density outputs, and the daily mean atmospheric density profile on June 15 that are obtained from the NRLMSISE00 model. The ion densities, especially the long-lived Na+, will be influenced significantly by tidal horizontal wind because the vertical ion drift is altered considerably by the horizontal wind through equation (5.10).

In addition to the tidal wave modulations, the GWs behavior in the lower thermosphere becomes quite complex, as most of them can become unstable, saturate, and even break in this region, causing vigorous changes in temperature and wind fields. As discussed in section 2.2, Fritts and Alexander [2003] provided dispersion relations for GWs propagating between the troposphere and thermosphere. Later, Vadas and Fritts [2005] presented the detailed description of the GWs (period less or equal to one hour) that penetrate high into the thermosphere. These GWs are damped and dissipated by the molecular viscosity and thermal diffusivity, rather than through turbulence diffusion. Based on Vadas and Fritts [2005], if these high-frequency GWs are able to reach the upper thermosphere (120-250 km), their amplitudes in the lower E region must be very
small (vertical wind velocity of about $10^{-4} - 10^{-3}$ m/s) to avoid wave saturation. Their contribution can therefore be safely neglected. Long-period inertial gravity waves (IGWs) have difficulty reaching above 100 km [Tsuda et al., 2000] due to their relatively slow vertical phase velocities that make them quite sensitive to atmospheric dissipation. On the other hand, the signatures of medium-frequency GWs (MFGWs) are frequently observed in the lower E region by the USU Na lidar [Cai et al., 2014; Yuan et al., 2016]. Occasionally, they even appeared in the ionosonde data [Yuan et al., 2013], indicating their activities in the middle and upper E regions. Therefore, we add a one-hour GW modulation in this simulation to study the contributions from the GW perturbation and its saturation process to the formation of high-altitude Na.

Based on Fritts [1984], when the background Brunt-Väisälä frequency, $N$, and the GW horizontal phase speed, $c$, are determined, the vertical wave number of a MFGW can be described as

$$m = \frac{N}{u - c},$$

(5.13)

where $\bar{u}$ is the background horizontal wind projected on the GW propagation direction. We set the vertical wind perturbation as

$$w' = A \exp\left(\frac{(z - z_0)}{2H}\right) \cdot \exp\left(k_H x + m(z - z_0) - k_H ct\right),$$

(5.14)

where $z_0$ is the initial altitude where the GW is generated. $H$ is the scale height, $k_H$ is the horizontal wavenumber, $A$ the constant related to the initial GW forcing, $c$ the
horizontal phase speed of GW, and $x$ is the horizontal distance that GW propagates.

Based on Fritts and Alexander [2003], we can calculate the horizontal wind perturbations for MFGWs as

$$u_H' = -mw' / k_H,$$  \hspace{1cm} (5.15)

and the associated pressure perturbation can be related to the vertical wind component as

$$p' = c_{ge} w' = -\frac{\omega}{m} w', \hspace{1cm} (5.16)$$

where $c_{ge}$ is the vertical group velocity and $\omega$ is the intrinsic frequency. The associated temperature perturbation can be derived from equation (18) in Fritts and Alexander [2003],

$$T' = -i * w' * N^2 * T / g \omega,$$  \hspace{1cm} (5.17)

where $T$ is the mean temperature, and $g$ is the gravitational acceleration ($g = 9.8 m/s^2$). Equation (5.17) suggests that the temperature perturbation should lead the horizontal wind perturbation, $u_H'$, by $\pi / 2$, which has been demonstrated in Chapter 3 in this dissertation. Notice that equation (5.17) is not only suitable for MFGW, but also for other GWs because the temperature perturbation is related to the vertical wind perturbations through the adiabatic process of atmosphere.

As was described in the previous sections, when a GW is generated at a lower altitude, its initial amplitude is small. During the upward propagation, its amplitude increases exponentially to compensate for the decrease of atmospheric density due to
energy conservation, until the wave begins to become saturated. The GW saturation and
decay processes are extremely complex, and critical to the dynamics of the ionosphere D
and E region. Thus, both processes and the associated effects on vertical transport have to
be carefully addressed in this model simulation.

As described in Fritts [1984], when the GW temperature amplitude increases and
results in \( N^2 < 0 \), the GW will become saturated. Its amplitude stops growing and
begins to decay until the GW approaches a critical level where the horizontal phase speed
equals the component of the background horizontal wind projected onto the GW
propagation direction. To address the decay process of the GW amplitude, we use
equation (24) and equation (26) from Fritts [1984].

\[
\begin{align*}
\dot{w'} &= w'_s \exp(-mi(z - z_s)), \\
mi &= \frac{1}{2H} - \frac{3}{2} \frac{\partial u / \partial z}{u - c}.
\end{align*}
\]

Equation (5.18) shows how the vertical wind decays above the saturation altitude with the
decay factor, \( mi \), given by equation (5.19). \( w'_s \) is the vertical wind amplitude at the
saturation altitude, \( z_s \).

More importantly, Fritts [1984] indicates that vigorous eddy diffusion process
will be generated when the GWs begin saturation. Like the vertical wind shear, also the
vertical shear of eddy diffusion can cause the vertical upward displacement of Na from
the main mesospheric Na layer. Thus, the eddy diffusion coefficient above the GW
saturation level will no longer be the climatological one adopted from WACCM, but has to be calculated from equation (33) in Fritts [1984] as follows:

\[ k_{zz} = \frac{k(u - c)^4}{N^3} \left( \frac{1}{2H} - \frac{3}{2} \frac{\partial^2 \tilde{u}}{\partial z^2} \right). \] (5.20)

3. Numerical Simulations

As shown in reactions R1-R3 on page 84, the densities of two major ionospheric species, \( \text{NO}^+ \) and \( \text{O}_2^+ \), are closely related to the Na ion chemistry and their densities have to be carefully calculated. Both are determined by the photochemical ionization process, which depends on the solar radiation angle. The variations of \( \text{NO}^+ \) and \( \text{O}_2^+ \) density mainly follow the changing of photo ionization rate of \( \text{O}_2 \), as shown in Table 5.1. We therefore need to calculate the photochemical ionization rate of \( \text{O}_2 \) (Ra1 in Table 5.1) first.

When the solar zenith angle, \( \chi \), is less than 75°, the photochemical reaction rate can be described by Chapman production rate

\[ P = I(z, \chi) \eta \sigma^a n(z) = I_\infty \exp[-H n(z) \sigma^a \sec \chi] \eta \sigma^a n(z), \] (5.21)

where \( I_\infty \) is the unattenuated solar flux at the top of the atmosphere. \( \sigma^a \) is the absorption cross section and \( n(z) \) is the atmospheric neutral density. \( H \) is the scale height and \( \eta \) is the probability of photon absorption. From Appendix J in Schunk and Nagy [2009], we obtain the cross sections and the solar fluxes. For the lower E region, we only use the second and third spectra intervals in Appendix J since other intervals only ionize at higher altitudes due to the larger absorption cross sections. Both spectra
Table 5.1. The Main Chemical Reactions related to $NO^+$ and $O_2^+$ from 95 km to 129 km

<table>
<thead>
<tr>
<th>Number</th>
<th>Reaction</th>
<th>Rate coefficient$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ra1</td>
<td>$O_2 + hv \rightarrow O_2^+ + e$</td>
<td>Photochemical</td>
</tr>
<tr>
<td>Ra2</td>
<td>$O_2 + NO \rightarrow O_2^+ + NO$</td>
<td>$4.6 \times 10^{-10}$</td>
</tr>
<tr>
<td>Ra3</td>
<td>$O_2^+ + e \rightarrow O + O$</td>
<td>$(2.4 \times 10^{-7})(300/T)^{0.5}$</td>
</tr>
<tr>
<td>Ra4</td>
<td>$NO^+ + e \rightarrow N + O$</td>
<td>$(4 \times 10^{-7})(300/T)^{0.5}$</td>
</tr>
<tr>
<td>Ra5</td>
<td>$O_2^+ + N_2 \rightarrow NO^+ + NO$</td>
<td>$5 \times 10^{-16}$</td>
</tr>
</tbody>
</table>

$^a$Unit: Unimolecular, s$^{-1}$, bimolecular, cm$^3$s$^{-1}$.

Intervals have an ionization probability of 100%. Based on equations (9.19) and (9.20) from Schunk and Nagy [2009], we can calculate the daytime solar flux. $n(z)$ from equation (5.21) is the $O_2$ vertical density distribution. The mean value of $O_2$ density is taken from the NRLMSISE-00 model [Picone et al., 2002] and then we add tidal perturbations to form the diurnal variation of $O_2$ (The same method was used for $N_2$ and $O$, which are needed in equation (5.11). The temperature used for the scale height, $H$, is the hybrid output from the CTMT and HAMMONIA. Finally, we can get the photochemical production rates for zenith angles less than 75°. For the angles larger than 75° and less than 95°, we use equations (9) to (11) from Smith et al. [1972] to calculate the photochemical reaction rates. The resulting variations of $NO^+$ and $O_2^+$ are presented in Figure 5.4. As the figure shows, in the daytime (0:00 UT-3:00 UT, 12:00 UT-24:00
UT), the densities of NO$^+$ and O$_2^+$ are high (~10$^5$ cm$^{-3}$), while they become very low at night (~10$^3$ cm$^{-3}$). Their peak densities appear around noon, when the photo ionization reaches its maximum of the day. Furthermore, both of the species do not exhibit the tidal feature due to their much shorter lifetime compared to those of the metal ions. Thus, the densities of NO$^+$ and O$_2^+$ in the lower E region are mainly determined by photochemical reactions, rather than dynamic transport. For the main neutral atmospheric constituencies, the N$_2$, O$_2$, and O densities are again obtained from the NRLMSISE-00 [Picone et al.,

Figure 5.4. Time and altitude variations of the simulated (a) NO$^+$ and (b) O$_2^+$ in June.
2002] (mean values) and the CTMT (tidal perturbations). Also in this simulation, the density of NO was obtained from the mixing ratios in Anderson et al. [1986], and the electron density was calculated as the total charges of all ions under the assumption of quasi-neutrality. The initial value of Na density was obtained from the climatology of Na densities for the month of June, as observed by USU Na lidar [Yuan et al., 2012]. With the E region background established as mentioned above, we apply Na/Na\(^+\) chemistry and the transport schemes into the model to simulate the Na and Na\(^+\) density in the lower E region of midlatitudes.

4. Results and Discussion

4.1. Results Due to Tidal Waves

As mentioned earlier, we ran the simulation with 1x and 5x tidal amplitudes provided by CTMT to represent the weak and strong tidal wave scenarios, and the results are shown in Figure 5.5. With the negative ion vertical drift velocity and shear induced by tidal winds; Na\(^+\) was converged to layer structure. For the case of standard CTMT tidal output (weak tidal waves), the sporadic Na\(^+\) layer (Figure 5.5a) in the first half of the night is near and above 115 km (from 0:00 UT to 5:00 UT), and the second Na\(^+\) layer is near the morning hours between 12:00 UT and 16:00 UT, with a slightly larger density at higher altitude compared to the first one. The two Na\(^+\) layers are not identical because of the contribution from diurnal tide and strong photo ionization during the day. However,
under the weak tidal wave condition, there is no Na\textsubscript{s} formed above 95 km (Figure 5.5b) during the first half of the night. The variations of Na above 95 km appear to be merely some extension of the main mesospheric Na layer. Above 100 km, the Na density started to increase during the second half of the night, forming a high-altitude Na cloud in the morning hours between ~ 11:00 UT and ~ 16:00 UT. The Na cloud has some downward propagating features but the peak density is quite low (about 60/cm\textsuperscript{3} and ~ 5% of peak density).
density of the main mesospheric Na layer), which is significantly less than those observed by the Na lidar (see Figure 5.1).

Turning to the scenario with 5x tidal amplitude in Figures 5.5c and 5.5d, two very strong Na\(^+\) layers are generated from 2:00 UT to 10:00 UT, and from 12:00 UT to 20:00 UT, respectively. Based on equation (5.10), the vertical ion drift velocity and the associated shear in this case are several times larger than in this case compared to the weak tide scenario. The negative velocity and shear are progressing downward over time due to strong tidal winds and, thus, a much denser Na\(^+\) layer is generated. The ranges of the two sharp Na\(^+\) layers overlap with the region of the negative shear of ion drift velocity. This again demonstrates that the E\(_s\) in the lower E region of midlatitude is mainly controlled by the vertical ion drift. Looking at the effects on the Na layer, the 5x tidal amplitude produced the large vertical wind perturbations (~ 1 m/s) near 119 km and lead to large vertical transport velocity of Na. A strong Na\(_s\) event occurred in this simulation from 12:00 UT to 18:00 UT in Figure 5.5d with a peak density higher than 120/cm\(^3\), culminating between 13:00 UT and 16:00 UT (early morning) near the altitude range of 98-104 km. The simulation shows that this second layer, which clearly has a downward phase progression feature that is common to tidal waves, can extend well beyond 110 km right before dawn. Figure 5.5d also indicates that the majority of the Na\(_s\) occurs in the region where the vertical transport velocity (equation (5.12)) has negative vertical gradient, which causes the Na atoms to converge into a layer structure. However, below
100 km, the eddy diffusion effect becomes more prominent and the peak of the Na layer is moving slightly away from the negative gradient region. In addition, in this simulation run, a relatively weak Na, was generated between 97 km and 100 km starting at 0:00 UT and lasting for the first half of the night. Its peak stayed near 99 km until about 4:00 UT and started propagating downward, merging into the main layer below 95 km after 6:00 UT.

Overall, this simulation indicates that the large horizontal wind shear due to strong tidal wave can, indeed, lead to a large vertical ion drift and shear causing the convergence of Na\(^+\) and the formation of Es. The same strong tidal wave, through its vertical wind component, induced a large negative gradient in vertical neutral transport velocity, \(v_n\), which can converge the Na atoms into a layer structure above 100 km. This process is also facilitated by the decreasing number of densities of NO\(^+\) and O\(_2\)\(^+\) (weakening Na ionization in R1 and R2) in the night. The weakening Na photo ionization in the lower E region also promotes this process during the night. It is worth noticing that converged Na\(^+\) densities in both the weak and strong tidal wave scenarios peaked at 2,000/cm\(^3\) and 3,000/cm\(^3\), respectively. The overall Na\(^+\) density should be less than 10% of the metal ion density of Es. From this, we can deduce that the density of entire metal ions is around 10\(^3\)-10\(^4\)/cm\(^3\), which agrees with the ion density observed by Kopp [1997].
4.2. Effects due to GW in weak tidal wave condition

To investigate the role of GW modulations in Es and Naε formation and evolution, we ran the model by adding a GW modulation to the standard CTMT tidal amplitude scenario during the first half of the night. The GW was chosen to have a horizontal phase speed of 80 m/s with a period of one hour (horizontal wavelength is 288 km), and propagating in the northward direction with an initial vertical wavelength of 10 km. The meridional wind we used is always less than the horizontal phase speed of this GW (maximum absolute value is 45 m/s). Thus, this northward-propagating GW did not encounter its critical level in the lower E region throughout the simulation. The source altitude was chosen to be 30 km and the initial vertical wind amplitude 0.006 m/s. Compared with the much slower modulations by tidal waves, GWs contribute to Naε formation mostly through the neutral dynamic transport process induced by wind perturbations, and do not significantly affect the Na ion chemistry. Based on the dispersion relationship described in section 5.2.3, the time and altitude changes of GW temperature, horizontal wind, and vertical wind amplitudes were obtained. The $N^2$ variation increased with the altitude due to the growth of this GW’s temperature modulation as it travels into high altitudes. Around 102 km, this GW eventually saturated when its vertical wind amplitude reached around 1.7 m/s. From 102 km upward, the GW decayed and dissipated due to its saturation process. Below 102 km, the GW contributions to the neutral vertical transport velocity, $v_n$, in equation (5.12), were
therefore mainly due to the GW vertical wind modulation, \( w' \). However, the eddy diffusion process induced by GW saturation started to alter the Na transport process considerably due to the significantly enhanced \( k_{zz} \) above the saturation level.

The simulated GW effects on Na, are shown in Figure 5.6, which shows a few relatively weak sporadic Na layers from 100 km to 105 km with a peak density less than 40/cm\(^3\). The changing of the vertical wavelength in the Na modulation is due to the changing of meridional wind direction with the GW propagating to higher altitudes, following equation (5.13). From the figure, the simulated Na exhibits some features that are, in general, similar to some of the Na events observed by USU Na lidar, such as Na observed on UT day 155 (June 4) in 2013 shown in Figure 5.1b. On the other hand, the

![Na density with GW under 1x tidal amplitude](image)

**Figure 5.6.** Na density during GW appearance from 4:00 UT to 7:00 UT under 1x tidal amplitude.
simulated density is considerably lower than what the lidar measured, only about 10% to 20% of the observed peak Na\textsubscript{a} density. The lidar-observed GW perturbations during that night also appeared to be much more complex in temporal and spatial structures than our over-simplified GW simulation, indicating there are some major GWs and small-scale dynamic processes missing in the current model. Our simulated background atmosphere, with climatological monthly mean fields and tidal waves, also underestimates the tidal wind magnitudes considerably in comparison to the real conditions on UT day 155 (June 4) in 2013. Since tidal waves tend to have strong day-to-day variability, likely through nonlinear interactions between tidal and planetary waves [She \textit{et al.}, 2004; Yuan \textit{et al.}, 2013], this monthly mean tide may not be suitable for the investigation on specific case of Na\textsubscript{a}. Furthermore, the Na source function is also a monthly mean result adopted from WACCM, while the meteor input of Na atoms on that day above the Western United States might have been much larger than this monthly mean value. Furthermore, the most significant defect of this GW simulation is that it was unable to precisely reproduce the interactions between the GW and background flow, including the momentum and energy transfer, that are critical for the neutral transport of all the atmospheric constituents. For example, as Figure 5.6 shows, when the GW was propagating upward, its vertical wavelength would decrease quickly when the meridional wind switched to northward. This would certainly induce large horizontal wind shear near 110 km, likely causing dynamic instability and overturning of the atmospheric constituents. However, these
interaction processes cannot be reproduced by the current simplified numerical scheme. Thus, a much more sophisticated GW and atmospheric model would be needed to fully address GW contributions to the formation of these high-altitude Naₜ.

### 4.3. The Roles of Vertical Wind and Eddy Diffusion

Equation (5.12) highlighted the importance of atmospheric wave-induced, vertical wind component, \(w'\), and its associated vertical gradient in the vertical transport velocity of Na, \(v_n\), and the formation of the high-altitude Naₜ. Since the semidiurnal tide dominates the modulation in the lower E region, its seasonal variation could be related to that of these high-altitude Naₜ. Figure 5.7a shows the seasonal variation of the CTMT vertical wind semidiurnal tidal amplitude near 105 km, 110 km, 115 km, and 120 km. The tidal amplitudes at these altitudes reach their annual maximum in June, except that around 105 km. More importantly, the seasonal variation of the vertical gradient of the semidiurnal tidal amplitude (Figure 5.7b) peaks in May or June at Logan, UT, depending on the altitudes. During the winter months, however, both the amplitude and gradient are near their annual minimums. Intriguingly, these high-altitude Naₜ are indeed mostly observed during the northern hemisphere summer months, while very few cases have been reported during winter.

To further investigate the dynamic effects on Naₜ, we set up a simulation without the Na ion-molecular chemistry using the strong tidal wave scenario (with 5x the CTMT
Figure 5.7. Seasonal variations of (a) vertical wind semidiurnal tidal amplitudes and (b) their vertical gradients at 105 km, 110 km, 115 km, and 120 km, based on CTMT.

tidal amplitude). Here, the standard WACCM $k_{zz}$ values are utilized. This set up aimed to evaluate what effects of the vertical wind component and eddy diffusion have on the formation of these NaNs individually. The results show the formation of two strong NaNs by the large-amplitude tidal vertical wind component alone: one occurring from 0:00 UT to 8:00 UT and another from 12:00 UT to 20:00 UT. The densities in both peak regions are considerably larger than those presented in Figure 5.5d, because there is no Na ion-molecular chemistry to ionize the Na atoms. To further demonstrate the critical
contributions of the vertical wind component, we completely removed the tidal vertical wind component, \( w' \), in another simulation, without turning on the Na chemistry. The results are shown in Figure 5.8c. As expected, there is no Na\(_s\) formed at all without vertical wind component. The weak Na layer near 105 km between 8:00 UT and 16:00 UT is generated by the Na source function adopted from the WACCM-Na model.

**Figure 5.8.** Results of the temporal and spatial variations of the simulated Na\(_s\) without Na ion chemistry and using the strong tidal wave scenario with (a) 1x \( k_{zz} \), (b) enhanced 50x \( k_{zz} \) and (c) 1x \( k_{zz} \) but no vertical wind. The contour lines in (a) and (b) are the vertical gradient of neutral transport velocity in equation (5.12). Purple lines represent the negative vertical gradients.
Another big uncertainty for the generation of Na$_s$ is the eddy diffusion coefficient, $k_{zz}$, in the lower E region. As mentioned earlier, the absolute value and the seasonal variation of $k_{zz}$ are still under debate and, this value probably represents the biggest uncertainty in this simulation. Here, the $k_{zz}$ that we have been utilizing so far is adopted from WACCM [Garcia et al., 2007], with values from ~10 m$^2$/s$^2$ to 20 m$^2$/s$^2$. During the GW saturation process, the $k_{zz}$ calculated from equation (5.20) was significantly larger above 102 km, around 1,000-2,000 m$^2$/s$^2$, due to the large-amplitude GW prior to its saturation. In a separate simulation, still with five times the standard CTMT tidal output and no Na ion-molecular chemistry, we increased the $k_{zz}$ by 50 times to see how a different $k_{zz}$ affects the Na$_s$ formation (Figure 5.8b). Comparing with Figure 5.8a, when the $k_{zz}$ increases, the peak density of the Na$_s$ decreases significantly. The 1x $k_{zz}$ in Figure 5.8a generates the peak around 550/cm$^3$, while the 50x $k_{zz}$ in Figure 5.8b results in ~250/cm$^3$ in peak Na$_s$ density. This is in agreement with the $k_{zz}$ effect on the Na main layer revealed by Plane [2004], which also exhibited the smaller peak Na density in the main layer with larger $k_{zz}$ (see Figure 6a in Plane [2004]). The increased $k_{zz}$ also alters the neutral vertical transport velocity, $v_n$, and its associated vertical gradient. The net result is that the slope of the Na$_s$ peak region is also modified slightly by the changing of $k_{zz}$. The larger $k_{zz}$ causes a steeper slope of Na$_s$ (faster downward progression).

Furthermore, the Na$_s$ in this case extends to higher altitudes and occurs at earlier times compared with the run with 1x $k_{zz}$. The two peak density regions fall from 22:00 UT to
4:00 UT (early evening) and between 8:00 UT and 16:00 UT (early morning). Overall, the change of $k_z$ affects the Na$_s$ formation significantly by altering both the temporal and vertical dimensions of Na$_s$. Thus, concrete knowledge of eddy diffusion is essential for Na$_s$ studies.

In Chapters three to five, we have discussed three important topics in the MLT region: gravity wave breaking, gravity wave induced potential energy density and high altitude sporadic Na layers. All three projects have been accepted for the peer-review journal publications. However, more problems and potential topics were discovered during these investigations. In the next two chapters, we will present conclusions and future work.
CHAPTER 6
CONCLUSIONS

In the dissertation, the first project (Chapter 3) focused on the investigation of gravity wave (GW) breaking evidence captured during a collaborative campaign of USU Na lidar and AMTM. Using both the USU Na lidar and Advanced Mesosphere Temperature Mapper (AMTM) observations, we were able to calculate the variations of the OH layer’s peak altitude throughout the night, and reveal that this wave-breaking event was taking place near 87 km. The observations showed that wave breaking occurred when a small-scale GW propagated into a transient dynamical unstable region formed around 11:00 UT, where it started to break immediately. The centroid height of OH layer was found to oscillate between 85 km and 92 km with a large amplitude, indicating considerable modulations by atmospheric waves. Benefiting from the lidar’s full diurnal cycle observation, we were able to separate the GW-induced perturbations and slowly changing large-scale background, which consisted of mean fields and tidal wave modulations. This enabled us to evaluate their roles in the formation of the atmospheric instability. The spectra analysis uncovered that there were several active large-amplitude GWs in the mesopause region during the night. Among them, a two-hour wave was clearly dominant during the wave breaking process. This large-amplitude GW signature was also captured by the collocated AMTM. It had an observed horizontal
phase speed of $\sim 180 \pm 9$ m/s, horizontal wavelength $\sim 1230 \pm 61.5$ km, vertical phase speed of $2.4 \pm 0.3$ m/s, and vertical wavelength of $17.3 \pm 2.2$ km. It maintained its propagating waveform within the mesopause region and beyond 98 km mainly due to its fast horizontal and vertical phase speeds. During the small-scale GW breaking process, this two-hour GW reached the amplitude of $\sim 15$ K in temperature and $\sim 20$ m/s in meridional wind, while its modulation of the zonal wind was relatively weak. The superposition of this large-amplitude GW and the semidiurnal tidal maximum in meridional wind in the early morning hours induced a large horizontal wind shear in the meridional direction near 87 km, causing a transient, dynamically unstable region. In addition, this combination caused a decreasing of $N^2$ value of 80%, which further aided the formation of the dynamic instability. The observed polarization relationship of this two-hour wave agrees extremely well with that of the medium-frequency GW under a no-dissipation scheme, implying that the two-hour wave dissipated very little during the process.

Overall, this study revealed another important mechanism for the atmospheric instability in the Mesosphere and Lower Thermosphere (MLT): the superposition of strong large-scale GWs and tidal waves can cause large vertical shear in the temperature and horizontal wind fields. The Na lidar full diurnal cycle measurements enabled us to identify the roles of large-scale background and GWs in facilitating this wave-breaking event. This study also demonstrated the capabilities of the coordinated observations
between the AMTM and the Na Doppler/temperature lidar for detailed and comprehensive investigations on GW dynamics within the MLT.

In the second project, by applying a Least-Squares Fitting (LSF) tidal removal algorithm to full diurnal cycle Na lidar temperature observations over Logan, Utah in September between 2011 and 2015, we were able to derive the monthly averaged profiles of the mesopause region temperature variance and the associated PED induced by the large-scale GWs during the boreal autumnal equinox at midlatitudes. The study covered the GW temporal spectrum from two hours up to the inertial period, providing the most comprehensive large-scale GWs results in the MLT. It revealed a node structure near the middle of the mesopause region in both profiles, decreasing GW modulations between 84 km and 90 km, but a reversed trend above 90 km where temperature variance increased from its minimum of less than 20 K² up to over 90 K² near 99 km. This node feature indicates the existence of a persistent GW dissipation layer near the middle of the mesopause region that cannot be resolved or well presented in current General Circulation Models (GCMs). The possible explanation may be related to the potential GW critical level near 90 km that prevents a significant portion of the GWs propagating beyond this altitude, while the GWs at other parts of the spectrum could penetrate or leak through the layer due to their faster horizontal phase speed. The dominance of long period GWs in the MLT region, especially of those with periods between three and five hours, is evident in these lidar observations, suggesting that these
GWs likely contribute the most to the GW energy and momentum in the MLT region.

On the other hand, this investigation found that results solely based on nighttime observations could overestimate the GW perturbations by ~50% or more near and below 90 km, possibly due to the bias brought by tidal components that cannot be fully removed. A similar test performed on the local results of the Whole Atmosphere Community Climate Model (WACCM) indicated a possible overestimation of the GW perturbations by more than 100% near 75 km and 90 km. Thus, full diurnal measurements are critical not only for tidal wave studies, but also for the precise calculation of GW perturbations. This new algorithm for detrending lidar measurements is validated using the latest high-resolution global GW simulation results from WACCM.

Besides applying the same LSF tidal removal approach on the WACCM output at the USU location, a spatial zonal wavenumber removal algorithm was also developed and applied to the WACCM data of all longitudes at 41°N to remove all large-scale tidal/planetary modulations with a zonal wavenumber less or equal than 10. The fitting residuals from these two completely different algorithms were treated as the GWs-induced temperature perturbations. The investigation revealed that within the fitting uncertainties, these two sets of WACCM GW temperature variance and PED profiles are almost identical in the MLT. This comparison clearly demonstrated that the new LSF tidal removal approach generates the most accurate GW results for ground-based observations. This study builds a concrete foundation for future
investigation on the climatology of GW perturbations and the associated forcing for other ground-based observations.

The third project dealt with the ion-neutral coupling processes in the lower E region, by combining the modified classic chemical model of Na with ion-neutral transport schemes induced by tidal winds in the lower E region, we have successfully simulated the qualitative Na density variations above 95 km. We were also able to separate the causal influences of tides, GWs, ion chemistry, and eddy diffusion. The simulation included photo ionization effects on major ionospheric species, such as NO$^+$ and O$_2^+$ that are critical to Na ion-molecular chemistry in the E region. The high-altitude sporadic Na (Na$_s$) features could be numerically generated with features similar to those observed by the Na lidar. In addition, using this numerical simulation, we have investigated and discussed the role of medium frequency GWs in the formation and evolution of Na$_s$ in the lower E region. The results of this numerical simulation work can be summarized as follows:

First, the Na ion-molecular chemistry is mostly working to ionize the Na atoms and, thus, reducing the density of the high-altitude sporadic Na layers while increasing the Na$^+$ density. These Na layers above 100 km do not result from the recombination of elevated Na$^+$ within sporadic E layers. The nighttime Na ionization processes are still relatively fast in comparison to the neutralization processes. Thus, the high density of Na$^+$ in the E layer cannot easily convert into neutral Na atoms.
Second, the neutral dynamic processes, especially the vertical wind modulations induced by tidal and gravity waves, are critical to the formation and evolution of the Na\textsubscript{s} above 100 km. Strong tide with sufficiently large vertical wind (value near or above ~1 m/s) can induce Na\textsubscript{s} naturally above 100 km through the combination of the changing in Na ion-molecular chemistry and neutral transport from the main mesospheric Na layer.

Third, the role of GWs on the Na\textsubscript{s} is mostly associated with the vertical transport of Na from the main mesospheric Na layer, rather than changing the Na\textsuperscript{+} density in the Na ion chemistry. When the GW vertical wind and the associated transport are large enough (around 1 m/s), Na\textsubscript{s} can be generated in the lower E region with distinct spatial features to those tidal wave-induced Na\textsubscript{s}. In this case, Na\textsubscript{s} can be formed without a strong E\textsubscript{s} layer.

Fourth, the eddy diffusion coefficient, \( k_{zz} \), is also critical for the formation of Na\textsubscript{s} above 100km. It is very important to have a comprehensive understanding of \( k_{zz} \), since it can change both the altitude and time distribution of Na\textsubscript{s} dramatically. Smaller \( k_{zz} \) can result in Na\textsubscript{s} with larger density at lower altitudes, compared with those generated with larger \( k_{zz} \).
CHAPTER 7

FUTURE WORK

Due to the complex dynamics of the Mesosphere and Lower Thermosphere (MLT) and the high variability of gravity wave (GW) features, the aeronomy community, after decades of numerical and theoretical investigations, still does not fully understand the behavior of GW and its associated impacts on the dynamics and chemistry of the upper atmosphere. For example, in the second collaborative project with the Whole Atmosphere Community Climate Model (WACCM), there are many other intriguing phenomena discovered. The most prominent one is that, after the tidal removal, both the USU Na lidar and WACCM temperature perturbations were dominated by GWs with periods from three to six hours (less than six hours). We call this type of GW quasi-medium-frequency gravity wave (QMFGW). The dominant QMFGW perturbations in USU Na lidar data were discovered by Cai et al. [2014] and had been reported in several observations at low and high latitudes. However, their horizontal information, so far, is derived indirectly from the nightglow measurements in the MLT, which have a limited field of view. The WACCM [Liu et al., 2014] we used in Chapter 4 lacks the GW parameterization and, thus, its results could not be compared directly with the lidar observations, prohibiting further investigation on this subject. However, the most recent version WACCM-DART [Pedatella et al., 2014] could potentially be used to investigate the QMFGW. This
version of WACCM is improved by adding data assimilation capacity using Data Assimilation and Research Testbed (DART), which constrains the model based on observations, so that the model can represent the state of the atmosphere more realistically [Pedatella et al., 2014]. The GWs within the model are self-generated, and the latest GW parameterization scheme is applied within the model. The model resolution is $2^\circ \times 2^\circ$ (220 km × 220 km max), which is sufficiently high to study the QMFGW, given the fact that the horizontal wavelength of QMFGW is usually larger than 700 km based on limited literatures (730 km in Li et al. [2007] and 2800 km in Fritts et al. [2014]). The WACCM altitudes cover the range from the Earth’s surface to 145 km, which includes the range (105-130 km) where most remote sensing instruments cannot conduct full-diurnal cycle observation. Therefore, one of our future works is to study QMFGWs systematically using WACCM-DART, such as their source and behavior in the thermosphere, along with its effects on ion-neutral coupling processes in the ionosphere.

The second interesting phenomenon was the tidal period modulations that are not induced by tidal waves, showing up in both the lidar observations and the model outputs. When we first utilized zonal wave number removal algorithm on the WACCM5 data, we removed all the modulations with zonal wave numbers equal or less than six. However, a strong 12-hour component appeared in the residuals at USU location and distorted the semidiurnal tidal phase line. We initially treated it as a tidal period IGW (see Figure 7.1), since its vertical wavelength was around 18 to 20 km, which is well below
Figure 7.1. The temperature perturbations of WACCM from September 8 to September 10 at Logan, UT after zonal wave number 6 removal.

the semidiurnal tide, but close to the IGW.

However, when we removed wavenumbers up to and including 10, this 12-hour component disappeared from the residual, indicating that this wave belonged to global-scale wave category. Indeed, as Palo et al. [1999] pointed out in his study on tidal-planetary wave interaction, this 12-hour component with short vertical wavelength is not an IGW, but very likely a global scale child wave with zonal wavenumber seven and westward direction, resulting from the nonlinear interaction between a two-day wave and migrating semi-diurnal tide. The USU Na lidar detected a similar event from UT day 313 to UT day 314 in 2011: an eight-hour component with a vertical wavelength around 20 km, as shown in Figure 7.2.

The preliminary analysis revealed that this eight-hour component was the
Figure 7.2. Temperature from UT day 313 to UT day 314 of 2011, measured by USU Na lidar.

combination of terdiurnal tide and some other unknown wave components. However, due to the limitation of measurements from a single location, we could not proceed with further investigation. Therefore, another part of future work of the second project in Chapter 4 is to employ the WACCM-DART to study the child waves systematically.

As discussed in the third project in Chapter 5, current simplified Na model did not include Na neutral chemistry, which could affect our Na density results below 100 km in the simulation. It is also important to run the numerical model under winter conditions, when the main mesospheric Na layer gets thicker and extends to higher altitudes. The comparison between results in summer and winter may be utilized to explain why Na\textsubscript{s} always take place in summer but not in winter. Furthermore, the variations of Na density will become more complicated when adding the GW activities and interactions with
mean flow such as mean flow acceleration. Thus, a better GW model would be required as well. Therefore, the second phase of our future work is to investigate GW saturation and breaking more extensively to obtain much more accurate variation of the eddy diffusion coefficient near the saturation level, which was found to play a significant role in the formation and evolution of high-altitude Na. As a final step, we are planning to expand the current one-dimensional model to three-dimensional, which would allow for the investigations of the influence by large-scale horizontal transport on the sporadic Na layer.
REFERENCES


Bossert, K., et al. (2015), Momentum flux estimates accompanying multiscale gravity
waves over Mount Cook, New Zealand, on 13 July 2014 during the DEEPWAVE

coordinated investigation of the gravity wave breaking and the associated dynamical
instability by a Na lidar and an Advanced Mesosphere Temperature Mapper over Logan,
UT (41.7°N, 111.8°W), *J. Geophys. Res. Space Physics*, 119, 6872–6864,

Cai, X., T. Yuan, and H-L. Liu (2017), Large Scale Gravity Waves perturbations in
mesosphere region above northern hemisphere mid-latitude during Autumn-equinox: A

Carter, L. N. and J. M. Forbes (1999), Global transport and localized layering of metallic

temperature measurements using a sodium-vapor dispersive Faraday filter in lidar

Inertia-gravity waves in Antarctica: A case study using simultaneous lidar and radar
measurements at McMurdo/Scott Base (77.8°S, 166.7°E), *J. Geophys. Res. Atmos.*, 118,

structures in the atmospheric sodium layer, *J. Geophys. Res.*, 109, D11306,

Clemesha, B. R. (1995), Sporadic neutral metal layers in the mesosphere and lower

Cox, R. M., and J. M. C. Plane (1998), An ion-molecule mechanism for the formation of
neutral sporadic Na layers, *J. Geophys. Res.*, 103(D6), 6349–6359,
doi:10.1029/97JD03376.

A theoretical study of the ligand-exchange reactions of Na`X complexes (X = O, O₂, N₂,


Gardner, C. S., and A. Z. Liu (2007), Seasonal variations of the vertical fluxes of heat


Krueger, D. A., C.-Y. She, and T. Yuan (2015), Retrieving mesopause temperature and


Li, T. (2005), Sodium lidar observed variability in mesopause region temperature and horizontal wind: planetary wave influence and tidal-gravity wave interactions, PhD Dissertation, Colo. State Univ., Fort Collins.


Rauthe, M., M. Gerding, and F.-J. Lübken (2008), Seasonal changes in gravity wave


Vadas, S. L., and H. Liu (2009), Generation of large-scale gravity waves and neutral


Yue, J. (2009), Effects of background winds and temperature on bores, strong wind shears and concentric gravity waves in mesopause region, PhD dissertation, Colo. State Univ., Fort Collins.


APPENDICES
Now we derive the dispersion relationship from the linearized equations:

\[-i \omega \hat{u} - f \hat{v} + ik \rho = 0, \quad (2.26)\]
\[-i \omega \hat{v} + f \hat{u} + il \rho = 0, \quad (2.27)\]
\[-i \omega \hat{w} + \left( im - \frac{1}{2H} \right) \rho = -g \rho, \quad (2.28)\]
\[-i \omega \hat{\theta} + \frac{N^2}{g} \hat{w} = 0, \quad (2.29)\]
\[-i \omega \hat{\rho} + ik \hat{u} + il \hat{v} + \left( im - \frac{1}{2H} \right) \hat{w} = 0, \quad (2.30)\]
\[\hat{\theta} = \frac{\hat{p}}{c_s^2} - \rho. \quad (2.31)\]

First we get \(u\) and \(v\) from equations (2.26) and (2.27):

\[\hat{u} = \frac{k \rho \omega + il \rho f}{\omega - f^2}, \quad (A.1)\]
\[\hat{v} = \frac{l \rho \omega - ik \rho f}{\omega - f^2}, \quad (A.2)\]

Then from equation (2.29), we obtain

\[\hat{\theta} = \frac{N^2 \hat{w}}{ig \omega}. \quad (A.3)\]

From equations (2.28) and (2.30), we get
After obtaining these four parameters in $\tilde{w}$, we substitute equations (A.1) to (A.5) into equation (2.30). Let us calculate these terms one by one:

$$-i \omega \rho = \frac{1}{c_i^2} \frac{i \tilde{w}}{\omega} \left( \frac{\omega^2}{\omega - N^2} \right) + \frac{N^2}{g \omega},$$

(A.6)

$$i k \tilde{u} + i l \tilde{v} = \frac{ik^2}{\omega - f^2} + \frac{i p \omega (k^2 + l^2)}{\omega - f^2} = \frac{i p \omega (k^2 + l^2)}{\omega - f^2}.$$  

(A.7)

Then we substitute (A.5) into (A.6) and (A.7), and we get

$$i k \tilde{u} + i l \tilde{v} = \frac{ik^2}{\omega - f^2} + \frac{i p \omega (k^2 + l^2)}{\omega - f^2} = \frac{i p \omega (k^2 + l^2)}{\omega - f^2}. 

(A.8)

Finally, the equation (2.30) can be rewritten as

$$w \left( \frac{1}{c_i^2} \frac{\omega^2}{\omega - N^2} + \frac{N^2}{g \omega} \right) - \frac{N^2}{g \omega^2} \left( \frac{\omega^2}{\omega - f^2} \right) + \frac{1}{2H} \left( \frac{\omega^2}{\omega - f^2} \right) \tilde{w} = 0. 

(A.9)$$
Multiplying \( \left( \omega - f^2 \right) \left( \frac{g}{c_s^2} + im - \frac{1}{2H} \right) \) on each term of equation (A.9) yields

\[
Y1 + Y2 = 0, \quad \text{(A.10)}
\]

\[
Y1 = w \left[ \frac{\omega - f^2}{c_s^2} + \frac{N^2}{g} \left( \omega - f^2 \right) \left( \frac{g}{c_s^2} m - \frac{m}{H} \right) - \left( k^2 + l^2 \left( \omega - N^2 \right) \right) \right], \quad \text{(A.11)}
\]

\[
Y2 = w \left[ \left( \omega - f^2 \right) \left( \frac{g}{c_s^2} + im - \frac{1}{2H} \right) \left( \frac{g}{c_s^2} m - \frac{m}{H} \right) \right]. \quad \text{(A.12)}
\]

Then we let the imaginary part and real part go to zero simultaneously. We focus on imaginary part:

\[
i \frac{N^2}{g} \left( \omega - f^2 \right) m + i \left( \omega - f^2 \right) \left( \frac{g}{c_s^2} \frac{m - m}{H} \right) = 0, \quad \text{(A.13)}
\]

\[
\frac{N^2}{g} m + \left( \frac{g}{c_s^2} \frac{m - m}{H} \right) = 0. \quad \text{(A.14)}
\]

Removal of \( m \) in equation (A.14) results in equation (2.32) in Chapter 2:

\[
- \frac{N^2}{g} + \frac{1}{H} = \frac{g}{c_s^2}. \quad \text{(A.15)}
\]

For the real part of equations (A.10) to (A.12), we utilize the equation (2.32) to obtain:

\[
X1 + X2 = 0, \quad \text{(A.16)}
\]

\[
X1 = \omega \frac{\omega - f^2}{c_s^2} + \frac{N^2}{g} \left( \omega - f^2 \right) \left( \frac{1}{-2H} \right), \quad \text{(A.17)}
\]

\[
X2 = - \left( k^2 + l^2 \left( \omega - N^2 \right) \right) + \left( \omega - f^2 \right) \left( \frac{1}{2H c_s^2} + \frac{1}{4H^2} - m^2 \right) = 0. \quad \text{(A.18)}
\]

Combined with equation (A.15), we simplify equations (A.17) and (A.18) as follows:
\[ \omega^2 \left( k^2 + l^2 + m^2 + \frac{1}{4H^2} - \frac{\omega^2 - f^2}{c_s^2} \right) = N^2 \left( k^2 + l^2 \right) + f^2 \left( m^2 + \frac{1}{4H^2} \right), \tag{A.19} \]

which is identical to equation (2.33) in Chapter 2.
APPENDIX B

COPYRIGHT PERMISSIONS FROM AGU

AGU grants permission for individuals to use figures, tables, and short quotes from AGU journal and books for republication in academic works and to make single copies for personal use in research, study, or teaching provided full attribution is included. There is no need to request this permission from AGU.
APPENDIX C

COPYRIGHT PERMISSIONS FROM ANGEO

As all articles published by ANGEO have been licensed under the Creative Commons Attribution License together with an author copyright, it is not necessary to seek our permission for the reproduction of the articles. It is only important that the work be cited correctly.
CURRICULM VITAE
XUGUANG CAI
(Dec 2017)

EDUCATION
Ph.D., Physics expected by Apr 2018
Utah State University (USU), Logan, Utah
Thesis: The investigation of gravity waves in Mesosphere/ Lower Thermosphere and their Effect on sporadic sodium layer.
Advisor: Dr. Tao Yuan.

B.S., Physics. July 2011
Anhui University, Hefei, Anhui, China
Advisor: Dr. Liang Lv.

RESEARCH INTEREST
Gravity wave dynamics in Mesosphere/ Lower Thermosphere.
Sporadic Na and Sporadic E layer in the lower E region mid latitude.
Coupling of Ionosphere and Thermosphere.

RESEARCH EXPERIENCE
Graduate Student/Research Assistant, Jan 2012—Present.
CASS, Department of Physics, USU

- Investigated gravity wave breaking evidence by the combined effort of Na Lidar and Advanced Mesosphere Temperature Mapper (AMTM) at USU. Collaboration with Dr. Mike Taylor, Department of Physics, USU.

- Proved that the least-squares fitting was effective in extracting gravity wave perturbations for the full-diurnal cycle observation by joint study of USU Na Lidar and Whole Atmosphere Community Climate Model (WACCM) data. Collaboration with Dr. Han-li Liu, National Center for Atmospheric Research (NCAR).

- Developed a numerical model of Na ion chemistry to explore the formation of sporadic Na layer above 100 km. Collaboration with Dr. Vince Eccles, Space Environment Corporation.

Undergraduate Student, Applied Physics Department, College of Physics and Material Science, Anhui University Hefei, China Oct 2009-July 2011.

- Worked on numeric simulations of vibration measurement based on self-mixing
effect of laser

TEACHING EXPERIENCE
Teaching Assistant Aug 2011—May 2013
Utah State University, Logan, UT
Developed curriculum in all areas including instruction, grading, holding office hours and assigning final grades.

Tutor Aug 2011—May 2013
Utah State University, Logan, UT
Tutored instructions to all the undergraduate physics courses.

PUBLICATIONS
Journals


**SKILLS**
Languages: English (fluent), Chinese (fluent).
Computer skills: MATLAB, FORTRAN, C, IDL, Microsoft word/excel, LINUX.
Other: Data analysis, Data modeling, Numeric Simulation, Parallel computing.

**PRESENTATIONS AT PROFESSIONAL MEETINGS**

Presentations


**Cai, X.** T. Yuan “Model investigation of Na neutralization within the sporadic E layer”, CEDAR workshop Seattle, WA, 2015.

Posters


**Cai, X.** T. Yuan, “Investigations on seasonal variations of gravity waves forcing by Na lidar over Logan, Utah (41.7°N, 111.8°W)”, CEDAR workshop Seattle, WA, 2015.

**Cai, X.** T. Yuan “Model investigation of Na neutralization within the sporadic E layer”, CEDAR workshop Seattle, WA, 2015.


PROFESSIONAL AFFILIATIONS
Student membership of American Geophysical Union (AGU).

AWARDS AND HONORS
AGU meeting travel award 2015.
Teaching Instructor Certificate, Utah State University, 2011.