On the solar cycle variation of the winter anomaly

A. G. Burns¹, W. Wang¹, L. Qian¹, S. C. Solomon¹, Y. Zhang², L. J. Paxton², and X. Yue³

¹High Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado, USA, ²Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland, USA, ³COSMIC Program Ofice, University Corporation for Atmospheric Research, Boulder, Colorado, USA

Abstract Constellation Observing System for Meteorology, Ionosphere and Climate, Ionosonde, and Global Ultraviolet Imager data have been used to investigate the solar cycle changes in the winter anomaly (the winter anomaly is defined as the enhancement of the F₂ peak electron density in the winter hemisphere over that in the summer hemisphere) in the last solar cycle. There is no winter anomaly in solar minimum, and an enhancement of about 50% in winter over summer ones on the same day of the year at solar maximum. This solar cycle variation in the winter anomaly is primarily due to greater winter to summer differences of [O]/[N₂] in solar maximum than in solar minimum, with a secondary contribution from the effects of temperature on the recombination coefficient between O⁺ and the molecular neutral gas. The greater winter increases in electron density in the Northern Hemisphere than in the Southern Hemisphere appear to be related to the greater annual variation of [O]/[N₂] in the north than in the south.

1. Introduction

The winter anomaly in F₂ region electron densities (henceforth the winter anomaly) is the condition in which greater daytime electron densities are seen at the F₂ peak (NmF₂) in winter than in summer [e.g., Rishbeth and Garriott, 1969]. Berkner et al. [1936] discovered the winter anomaly and Berkner and Wells [1938] determined that there was also an annual anomaly in which NmF₂ is greater in December than in June. If measurements are only made in one hemisphere, the two effects cannot be differentiated; in the south, the annual anomaly will be subtracted from the winter anomaly, and in the Northern Hemisphere, it will be added. Burns et al. [2012] suggested that this problem should be addressed by defining the winter anomaly as being the relationship between the electron density at the F₂ peak in the winter hemisphere and the equivalent electron density at the conjugate latitude in the opposite hemisphere at the same longitude (to remove any longitudinal effects) and on the same day of the year. This definition removes contamination by the annual anomaly (or asymmetry—e.g., Zeng et al., 2008).

Over the years, many studies have described both the winter anomaly and the annual anomaly. Lee et al. [2011] recently described these ably in their introduction, so only a few papers will be mentioned here. A number of studies extended Berkner et al.’s [1936] original work [e.g., Yonezawa, 1971; Torr and Torr, 1973; Rishbeth and Muller-Wodarg, 2006; Pavlov and Pavlova, 2009]. The Torr and Torr [1973] study described the difference in the winter anomaly at solar maximum and minimum. Their global map showed that the winter anomaly was large at solar maximum and small at solar minimum. They also found large differences between the southern and northern hemispheres. Their work, however, used a definition of the winter anomaly that effectively included the annual anomaly as well and did not attempt to describe how the winter anomaly developed over a solar cycle. Pavlov and Pavlova [2009] suggested that the Torr and Torr [1973] study might not make an accurate calculation of the intensity of the winter anomaly because of contamination by magnetically disturbed conditions. However, as geomagnetic storms are relatively rare events (as defined by Kp > 5), their contribution to the averages used by Torr and Torr was probably relatively small. Given the relative rarity of geomagnetic storms (typically only a few days a month), using medians, which we do in this study, largely eliminates this problem. There is still the issue of whether the background level of geomagnetic activity contributes to the winter anomaly either positively or negatively over the solar cycle; it is not clear whether this effect should be considered separately from the solar cycle variations. It is also difficult to remove such effects. The best resource for this is general circulation models, but they can have problems in accurately describing the winter anomaly [Qian et al., 2013].

The generally accepted explanation for the winter anomaly is that increases in O density relative to N₂ density in winter compared to summer are sufficiently large to not only counteract solar zenith angle effects (there is
These ionosonde data were primarily automatically scaled, and the identification of the frequency at the $F_2$ peak is mostly robust, so there should not be an offset between different stations. The difference in the absolute geographic latitudes of the comparison stations potentially introduces an error into our calculations, as the ionospheric response may change as a result of zenith angle and background composition. However, we undertook tests of the effect of solar zenith angle on the $N_mF_2$ at the various ionosondes, which showed that this did not significantly affect the conclusions that we have drawn.

Neutral composition data are also included here. They were measured by the Global Ultraviolet Imager (GUVI) instrument [Paxton et al., 1999] that flew on board the thermosphere-ionosphere-mesosphere electrodynamics (http://www.timed.jhuapl.edu/WWW/index.php) satellite. This paper will use these data to describe the evolution of the winter anomaly from 2001 to 2011, and provide a possible explanation for this solar cycle variation of the winter anomaly.

2. The Theoretical Basis for the Winter Anomaly

Rishbeth and Setty [1962] developed a theoretical understanding for the winter anomaly based on neutral composition. This can best be understood in terms of the ion continuity equation [e.g., Schunk and Nagy, 2000] for O+, as that species is the dominant ion near the $F_2$ peak.

$$\frac{\partial [O^+]}{\partial t} = q[O] - \beta [O^+][M] - \nabla \cdot (\mathbf{v}[O^+])$$

(1)

where $[O^+]$ is the density of O ions, $t$ is the time, $q$ is a production rate as a result of ionization by either solar EUV radiation or precipitating electrons and ions, $[O]$ is the neutral O density, $\beta$ is the loss rate for recombination, $[M]$ is the density of the molecular ions $O_2$ and $N_2$, and the last term includes all of the transport processes.

2.1. The Possible Effects of Transport on Solar Cycle Variations of the Winter Anomaly

Any possible source of the solar cycle variation of the winter anomaly involves equation (1), as the composition of the ionosphere at the $F_2$ peak is dominated by $O^+$ [e.g., Schunk and Nagy, 2000]. The transport terms can largely be eliminated as a possible source of this solar cycle variation for several reasons. The most critical reasons are given by Rishbeth and Edwards [1989, 1990], who showed that the daytime ionosphere is best ordered on constant pressure surfaces. There are two relevant implications of this: first, transport terms move the $F_2$ layer off this constant pressure surface by definition, and second, $N_mF_2$ then represents the height change of a pressure surface, which is primarily caused by the integrated effect of changes of neutral temperature. If $[O]/[N_2]$ remains constant on a pressure surface over a solar cycle, then a further conclusion...
can be drawn that \( N_mF_2 \) changes over a solar cycle will be primarily driven by changes in the amount of ionizing radiation (see equation (2)).

Some other considerations also militate against transport being important in the changes of winter anomaly over a solar cycle. To the best of our knowledge, daytime neutral meridional winds change by less than 50 m/s over a solar cycle during the daytime [Hagan, 1993; Buonsanto and Witasse, 1999] and are relatively more southward (equatorward) in winter at solar minimum than at solar maximum. Poleward winds tend to decrease electron density; so if they are more equatorward in solar minimum, it would be expected that the winter anomaly would be more pronounced in solar minimum than in solar maximum. Emmert et al. [2006] presented wind data from several Fabry-Perot interferometer (FPI) stations, but because FPIS only measure winds at night, this is of limited relevance to daytime \( N_mF_2 \). Wind Imaging Interferometer 630.0 nm data are available to analyze the change in meridional wind [e.g., Emmert et al., 2002] over a solar cycle and have been included in the horizontal wind model [Drob et al., 2008], the solar cycle variation of these winds has not yet been included explicitly in a scientific publication. Solar cycle variations in 557.7 nm emissions have been analyzed [Liu and Shepherd, 2007].

The daytime, low-latitude, eastward electric field should also not change much with solar cycle as it is primarily driven in the \( E \) region [Richmond, 1985], where solar cycle changes are minimal. Fejer [1993] did show that there were some significant changes in the zonal ion drift, but little change in the vertical ion drift: the component that would cause changes in electron density. Thus, it is unlikely that solar cycle changes in the low-latitude electric field drives changes in the winter anomaly.

The last transport term, ambipolar diffusion, is more problematic: difficulties in modeling plasma temperature and ion flux, and consequently the topside scale height and ambipolar diffusion, occur. Solar cycle changes will occur, but their nature is not clear, nor is their impact on the winter anomaly. The solar cycle variations of all of the transport terms need to be revisited at a later date, but for the present, we will continue under the assumption that they are not the prime drivers of the changes that we will describe.

### 2.2. The Reduced Continuity Equation

Without transport, equation (1) can be reduced to

\[
\frac{d[O^+]}{dt} = \frac{q[O]}{\beta[M]}
\]  

(2)

If it is also assumed that the production and loss of \( O^+ \) is not time dependent, which is reasonable for long-term medians.

For simplicity, it is often assumed that \( q \) and \( \beta \) are constant and that \( [O]/[M] \) can be reduced to \( [O]/[N_2] \), as \( [N_2] \) is much greater than \( [O_3] \) at these heights. However, both \( q \) and \( \beta \) vary with latitude.

The variation in \( q \) relates to the change in solar zenith angle with latitude (henceforth, it will be assumed that all calculations are for local noon). The latitudinal variation of \( q \) does not change with solar cycle and can, for simplicity, be regarded as being roughly the cosine of the zenith angle. Thus, the latitudinal variation in \( q \) cannot be the source of the solar cycle variation in the magnitude of the winter anomaly.

Another process needs to be considered: the change of \( \beta \) with latitude. The loss coefficient \( \beta \)'s variations with latitude actually reflect a temperature dependency. The importance of this temperature dependency has been invoked previously to describe changes in the ionosphere due to geomagnetic storms [Seaton, 1956], but it was found to be relatively unimportant in determining the cause of negative storm effects. We reexamine the importance of temperature effects on the winter anomaly in this paper using the formulation described by Roble [1995] for the reaction of \( O^+ \) with \( N_2 \)

\[
O^+ + N_2 \xrightarrow{\beta_1} NO^+ + N(4S)
\]  

(3)

\[
\beta_1 = 1.533 \times 10^{-12} = \frac{5.92 \times 10^{-13}B_1 + 8.6 \times 10^{-14}B_2}{T_1} \leq 1700 K \text{ and } B_1 = T_1/300,
\]

\[
B_2 = B_1^2,
\]

\[
T_1 = 0.636 \times T_f + 0.364 \times T_n
\]
The resulting molecular ions recombine relatively quickly with electrons and are thus not rate limiting. A similar relationship occurs when recombination with O₂ is considered.

We also examine whether the summer-to-winter ratio of [O]/[N₂] changes over a solar cycle. NRL-MSIS (Naval Research Laboratory-Mass Spectrometer and Incoherent Scatter) [Picone et al., 2002] produces a weaker solar cycle variation, which is consistent with the authors’ description of the model; as we will show later GUVI measurements show a considerable variation.

3. Results

Figure 1 shows the change of $F_{10.7\text{ median}}$ (a running median calculated using ~3 solar rotations of data) from 2001 to 2011. $F_{10.7\text{ median}}$ reached values of over 200 in 2002 and then decreased rapidly to a value of below 100 in 2005, before reaching a minimum in 2007. After this $F_{10.7\text{ median}}$ stayed low until 2011, when there was an increase in activity up to a level of about 110 solar flux units. $F_{10.7\text{ median}}$ (or average—they are essentially equivalent where the variable approximates a normal distribution) may not be a good proxy for the EUV radiation that is reaching the thermosphere in 2008 and 2009 [Solomon et al., 2011]. Direct EUV measurements from Solar Extreme ultraviolet Experiment [Woods et al., 1994] indicate that EUV radiation was even lower in these 2 years. Supporting evidence for this can be found in both neutral densities [Solomon et al., 2011] and electron densities [Burns et al., 2012]. In both cases, densities were lower in 2008 and 2009 than in 2007.

Figure 2 gives $N_mF_2$ values from COSMIC for the last solar minimum at ± 45 geomagnetic latitude, using 6 h medians from 0900 to 1500 local solar time, and a 30 day running median. There are several features of this plot, which were described in detail by Burns et al. [2012]. Uncertainties for these measurements can be found in that paper. The dominant climatological feature is that $N_mF_2$ is greatest at the equinoxes—the equinoctial anomalies or peaks [see Zhang et al., 2010]. There is one unusual aspect of these equinoctial anomalies: the peaks are usually closer in time across the winter solstice than they are across the summer solstice. Another feature of these plots is that an annual anomaly consistently occurs from year to year, insofar as the average $N_mF_2$ of the two hemispheres is greater in December than in June.

Neither of these anomalies are the ones that interest us in this paper. Instead, we are interested in the winter anomaly that was defined earlier as the enhancement of the peak electron density in the winter hemisphere over that in the summer hemisphere in both June and December. In all solstices in Figure 2, the summer values of $N_mF_2$ were greater than the winter ones on the same day of the year. In other words, there is no winter anomaly in
these data. This leads to the questions:
does this lack of winter anomaly continue over the whole solar cycle; and what could be driving the lack of a winter anomaly at solar minimum?

A data set is needed that is continuous over a solar cycle to address this question of winter anomaly persistence. COSMIC has not been operating long enough for this purpose. So instead of satellite measurements, data from ionosondes are used to address this issue. Figure 3 presents \( N_mF_2 \) values for stations in Southern Africa and Europe. Three stations are used in this figure: Grahamstown, South Africa; Chilton, England, and Rome, Italy (see Table 1 for locations). The two Northern Hemisphere stations bracketed the magnetic latitude of Grahamstown. Two other Northern Hemisphere stations were also studied in this longitude region: Juliusruh, Germany, and D’Elebre, Spain (see Table 1 for locations). The climatology at these stations was very similar to that at Chilton and Rome, respectively, so they were not plotted in this paper. Figure 3 is complicated: close study shows that winter values of \( N_mF_2 \) are much greater than summer ones at conjugate stations (on the same day of the year) at solar maximum (2002), but summer values are slightly greater than winter ones at solar minimum (2008). The picture is complicated by the differing solar zenith angles at the stations. Grahamstown is located at –33.3 south geographic latitude, a significantly lower latitude than Rome (41.8°N), and a much lower latitude than Chilton (51°N). There is a distinct difference between the climatological values on \( N_mF_2 \) at Chilton and Rome, but this can be attributed to the difference in solar zenith angle. All of the geographically close ionosondes were studied (including a number that are not shown in this paper) have a very similar climatology. Most of the differences between these stations can be attributed to differences in solar zenith angle.

To provide greater clarity about the solar cycle variations, we looked at the data in a solar maximum year (2002—Figure 4a) and a solar minimum year (2008—Figure 4b). The behavior of the equinoctial peaks in these two figures provides an interesting study, but it is not our purpose to discuss them here. Instead, we want to consider what is happening at the beginning of the year and in the middle of the year (left- and right-hand sides near the tick mark in the center of the plots). At solar maximum, there is a clear seasonal pattern at the solstices: the blue (northern) curves were much greater than the red (southern) one at the December solstices; and the red (southern) one were much greater than the blue (northern) ones at the June solstice. Thus, there is clearly a strong winter anomaly in solar maximum with the winter value of \( N_mF_2 \)

### Table 1. The Location of the Ionosondes Used in This Study

<table>
<thead>
<tr>
<th>Station</th>
<th>Geographic Longitude</th>
<th>Geographic Latitude</th>
<th>Geomagnetic Longitude</th>
<th>Geomagnetic Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grahamstown, South Africa</td>
<td>26.5</td>
<td>–33.3</td>
<td>91.0</td>
<td>–41.7</td>
</tr>
<tr>
<td>Chilton, UK</td>
<td>–1.3</td>
<td>51.6</td>
<td>77.6</td>
<td>48.3</td>
</tr>
<tr>
<td>Rome, Italy</td>
<td>12.5</td>
<td>41.8</td>
<td>86.5</td>
<td>36.0</td>
</tr>
<tr>
<td>Juliusruh, Germany</td>
<td>13.4</td>
<td>54.6</td>
<td>90.1</td>
<td>51.0</td>
</tr>
<tr>
<td>D’Elebre, Spain</td>
<td>0.3</td>
<td>–40.8</td>
<td>76.0</td>
<td>36.0</td>
</tr>
<tr>
<td>Port Stanley, UK</td>
<td>–57.8</td>
<td>–51.7</td>
<td>10.4</td>
<td>–38.4</td>
</tr>
<tr>
<td>Wallops Island, USA</td>
<td>–75.5</td>
<td>37.9</td>
<td>1.6</td>
<td>48.2</td>
</tr>
<tr>
<td>Eglin, Air Force Base, USA</td>
<td>–86.7</td>
<td>30.4</td>
<td>–15.0</td>
<td>40.9</td>
</tr>
<tr>
<td>Hobart, Australia</td>
<td>147.3</td>
<td>–42.9</td>
<td>–133.7</td>
<td>–53.8</td>
</tr>
<tr>
<td>Magadan, Russia</td>
<td>151.0</td>
<td>60.0</td>
<td>–140.2</td>
<td>53.8</td>
</tr>
<tr>
<td>Petropavlovsk, Russia</td>
<td>158.7</td>
<td>53.0</td>
<td>–132.4</td>
<td>46.4</td>
</tr>
</tbody>
</table>
being 40–80% greater than the summer one, if the winter-summer \( N_mF_2 \) ratio is calculated for a particular day of the year. The annual anomaly was also clear in this plot, with the January values being much greater than the June values in the Northern Hemisphere (125% for Rome and 200% for Chilton). The December values were also greater: 50% for Rome and 130% for Chilton.

The difference was small in Grahamstown, where the effects of the winter anomaly counterbalance those of the annual anomaly.

The situation was quite different in 2008 (Figure 4b). Now the summer values were greater than the winter values on the same day of the year. In January (southern summer), Grahamstown \( N_mF_2 \) values were about 33% greater than those at Chilton, and slightly greater than those at Rome. In December (also southern summer), they were 66% greater in Grahamstown than they were in both Chilton and Rome. In the northern summer (June), \( N_mF_2 \) values were about 33% greater at Rome than they were at Grahamstown. They were roughly the same at Grahamstown and Chilton. Just as there was no evidence of a winter anomaly on the same day of the year in the COSMIC data, there was no evidence of a winter anomaly during 2008 in these ionosonde data.

Two other longitudes were also studied: one in the American sector and one in the East Asian sector. The first consisted of the ionosondes in Port Stanley, Wallops Island, and Eglin Air Force base, whereas the second consisted of the ionosondes at Hobart, Magadan, and Petropavlovsk (see Table 1 for locations). Christchurch was also considered in this last longitude sector group, but its climatology was very similar to Hobart and its data density was less, so it is not considered further here.

It is necessary to look at one other longitude sector to confirm the results of Figure 4, but not both as all three sectors show a qualitatively similar variation in the nature of the winter anomaly over a solar cycle. The American longitude sector is shown in Figure 5 as it presents some interesting features that differ from those in the other two sectors. The 2 years are slightly different, as there were missing data in 2002 and 2008. Instead, 2002 and 2007 are described here. \( N_mF_2 \) values were greater at Eglin and Wallops Island than at Port Stanley in both January and December (Figure 4a) in 2002. The difference was about 33% for both northern stations in January and 50–60% in December. Port Stanley \( N_mF_2 \) was slightly bigger than Wallops Island in June, but about the same as at Eglin in this time. This apparent lack of winter anomaly in June is almost certainly a result of the vastly different latitudes of the stations. The solar zenith angle at midday at Port Stanley in the June solstice is about 74°, whereas the solar zenith angle at Eglin is about 8° (Wallops Island is about 25°). The ionization rate is dependent on the cosine of the zenith angle (approximately), so the difference is that ionization rate would compensate for differences in composition if they were the source of the winter anomaly. Note that in the December solstice, the solar zenith angle at Port Stanley is considerably less than it is at Eglin at the same time, so the reverse argument in December is not fully applicable.

There was no evidence of any significant winter anomaly for these three stations in 2007 (Figure 5b). \( N_mF_2 \) at Port Stanley was about equal with that of Eglin and greater than that of Wallops Island in January. It was also
greater than $N_{m}F_2$ at Wallops Island in December. In June, $N_{m}F_2$ at the northern sites was greater than that at Port Stanley. One surprising feature of this plot was that the relative June difference between northern and southern stations did not change much between 2002 and 2007, indicating that the balance between the processes that cause the winter anomaly does not change much over a solar cycle in June in this meridian. Also, the values of $N_{m}F_2$ only dropped from about $6 \times 10^5$cm$^{-3}$ to $3 \times 10^5$cm$^{-3}$ in June. It is not so surprising for the summer (northern) stations, which exhibit similar behavior at other longitudes. But it is an indication of the lack of a winter increase of electron densities at Port Stanley in any part of the solar cycle, even in 2002. In this case the compositional effects appear to be balanced by the zenith angle effects.

The purpose of this paper is not only just to contrast solar maximum with solar minimum but also to provide an understanding of the way in which the winter anomaly changed in the declining phase of the previous solar cycle. To do this we introduce a concept called the winter anomaly ratio (WAR), which is the ratio of $N_{m}F_2$ for a median of 30 days for a station in the winter hemisphere divided by the same median for an approximately conjugate station in the summer hemisphere for the same solstice (December or June).

$$\text{WAR} = \frac{N_{m}F_2 \text{ winter 30 day median around solstice}}{N_{m}F_2 \text{ summer 30 day median around solstice}}$$

Figure 6 shows the WAR calculated by comparing $N_{m}F_2$ values at European/African longitudes. In June, it is winter in the south and summer in the north, so the WAR is calculated by dividing $N_{m}F_2$ at Grahamstown by its value at the northern stations. In December, $N_{m}F_2$ at the northern stations is divided its value at Grahamstown. The winter anomaly is seen clearly in this plot in solar maximum. In 2002 and 2003, there were WAR values of 1.5 and above on average. The WAR was lower in 2001 and 2004 and had decreased to close to (see blue triangles) or below 1 (no winter anomaly) by 2006. There were also indications of the WAR being slightly greater than 1 in 2011. A comparison with
Figure 2 suggests that the winter anomaly does not exist when the \( F_{10.7} \) is below a value of about 90–100 solar flux units. This result was consistent with that seen in the other longitude sectors (not shown), indicating that the amount of solar flux is critical in determining if the winter anomaly is seen anywhere in the world.

There were longitudinal differences in the WAR at solar maximum. The value of the WAR 2002 was much greater over East Asia and much less over the Americas. Most of these differences may be attributable to the different geographic latitudes of the stations used for comparisons and hence different zenith angle effects, but other factors may be involved as well. For example, the location of the geomagnetic poles may change the quiet time, neutral composition in some sectors relative to others.

4. Some Simple Theoretical Aspects

Equation (2) gave the expected O\(^+\) density in the absence of any transport. We examined this equation to see if it could give any insight into why the WAR varied with solar cycle. If we treat the \([O]/[N_2]\) as being one term (as measured by GUVI, for example), then the left-hand side of equation (2) has three parts: the ionization rate (\(q\)), the recombination rate (\(\beta\)), and the \([O]/[N_2]\). Although the magnitude of \(q\) changes markedly over a solar cycle, it depends on the output from the Sun, not the latitude of the observation on the Earth. The only part of \(q\) that is relevant to this discussion is the approximately cosine dependence relative to geographic latitude due to changes in the solar zenith angle. The overall balance between the \([O]/[N_2]\), \(q\), and \(\beta\) is adjusted if this latitude varies. This will be better ordered in geographic coordinates than in geomagnetic ones, which is an exception to the more normal situation in which the electron density is better ordered in geomagnetic coordinates [Mendillo, 2006].

So what other parts of equation (2) could vary over the solar cycle? Figure 7 shows the variation of the recombination rate with temperature. Temperatures in the summer daytime at solar minimum are typically of the order of 800 K. Winter temperatures are typically 100–200 K colder. The higher summer temperatures at solar minimum mean less recombination, so there should be more electrons. At solar maximum, daytime summer temperatures are typically 1200–1300 K, whereas winter ones are 1000–1100 K. Thus, there is more recombination in the summer hemisphere than the winter one at solar maximum. At first we thought that this was the explanation for the solar cycle changes in the WAR, but it turned out that this effect could only cause a 5–15% change in \(N_mF_2\). The next possibility is \([O]/[N_2]\). When we ran NRL-MSIS [Picone et al., 2002; Hedin, 1983, 1987, 1991], we found no change in
the winter to summer ratio of \([O]/[N_2]\) over a solar cycle. However, we realized that there are potential problems with NRL-MSIS calculations of the \([O]/[N_2]\) contributions to the WAR variations over a solar cycle. The MSIS compositional description came primarily from a variety of satellites, most of which were not operated for long enough to provide consistent information over a large enough portion of the solar cycle (Atmosphere Explorer-E was an exception, but its measurements did not extend to the middle latitudes). Thus, MSIS may not contain an accurate representation of the solar cycle contribution of \([O]/[N_2]\) to the WAR.

Instead, we looked at the GUVI variations of \([O]/[N_2]\) in the south and the north from 2002 to 2007 (Figure 8). To simplify the plot, we used zonal averages of \([O]/[N_2]\) for each latitude and ran a 120 day running median at 45°N and S geographic to remove the yaw cycle from the data. The results are plotted in Figure 8. The southern results are plotted in red and the northern results in dark blue. The light blue dashed lines denote the solstices. Smallest values of \([O]/[N_2]\) were found in the summer, and the largest values occurred on the autumnal side of the solstice. Generally, the summer to winter variation of \([O]/[N_2]\) in 2002 was about 2.1 in June and 2.0 in December. In 2006, these ratios decreased to about 1.5 in June and about 1.6 in the December. Thus, there is a pronounced solar cycle change in the winter to summer ratios of \([O]/[N_2]\), with smaller values occurring at solar minimum than at solar maximum. Note that the behavior of \([O]/[N_2]\) has significant hemispheric differences: summer (June in the north and December in the south) values of \([O]/[N_2]\) are greater in the north than in the south, whereas winter values (December in the north and June in the south) tend to be greater in the north than in the south.

A calculation of the expected change in \(N_{mF_2}\) was made that was based on these numbers (a linear scaling factor was applied to account for the height difference between the \([O]/[N_2]\) measurement at ~ 150 km and the \(F_2\) peak at ~ 300 km). The result was calculated for January and plotted as a green line on top of Figure 9. This line fits the observed change in \(N_{mF_2}\) accurately, showing that the change in the WAR over the solar cycle is primarily the result of a change in the winter to summer ratio of \([O]/[N_2]\) in this period. Note that the green line also decreases to a value of 1 or below in 2006 (indicative of no winter anomaly), providing a further indication that no winter anomaly should be expected when \(F_{10.7}\) gets below about 90–100 solar flux unit (sfu).

5. Discussion

The winter anomaly has been observed since 1936 [Berkner et al., 1936]. It has been defined by there being greater electron densities at the \(F_2\) peak in winter than summer. It is anomalous because the solar zenith angle is greater in winter than in summer; thus, there is more ionization in summer than in winter. So, if everything else was equal, there should be greater electron densities in summer than in winter. Rishbeth and Setty [1962] resolved this conundrum by determining that the change in \([O]/[N_2]\) between winter and summer could more than balance the reverse effect due to the solar zenith angle.

However, important issues remained. One of which is the definition of winter and summer and how the winter anomaly can interact with the annual anomaly, which apparently has a different source [e.g., Zeng et al., 2008]. If winter and summer values of \(N_{mF_2}\) are those measured at one station, then the annual anomaly cannot be uniquely separated from the differences due to a seasonal anomaly. The results of Torr and Torr [1973] reflect this and are therefore somewhat confusing. Burns et al. [2012] chose instead to define the winter anomaly to be the difference between the winter value and the summer value at the same solstice.
Thus, annual anomaly effects can be mostly separated from the seasonal or winter anomaly effects. In practice, this works better for satellite data than for ionosonde data, as it is difficult to find northern and southern stations that are opposite each other over the equator. There is therefore some variation in the results shown in this paper due to changes in solar zenith angle.

The Torr and Torr [1973] paper did provide insight into how the winter anomaly varied with solar cycle, despite the confusion with the annual anomaly. They found that the winter anomaly was small to nonexistent at solar minimum and large at solar maximum. The work that has been done in this paper has confirmed that there are solar cycle variations in the winter anomaly and has also shown that much of the north-south variation that Torr and Torr found can be attributed to the interaction of season with the annual anomaly. We have also shown how the winter anomaly varies with solar cycle and determined a cutoff value of solar flux (90–100 sfu), below which the winter anomaly is not seen. Furthermore, we have shown that the solar cycle variation of the WAR is primarily the result of changes in the winter to summer ratio of [O]/[N2], with secondary effects due to changes in the recombination rate, which occur as the temperature changes over the solar cycle.

To a certain extent this just changes the problem to another one: why does the winter to summer ratio of [O]/[N2] change over a solar cycle? Hagan [1993] and Buonsanto and Witasse [1999] used Millstone Hill radar data to show that there were small changes in meridional winds in the daytime over a solar cycle. Seasonal changes in neutral composition are driven primarily by changes in the neutral winds [Mayr et al., 1978; Burns et al., 1989]. Burns et al. showed that the winds that blew through a constant pressure surface, which respond to changes in the latitudinal and longitudinal gradients of the horizontal winds, were the major source of changes in neutral composition. Changes of [O]/[N2] occur in both the winter and the summer. In summer, the [O]/[N2] is slightly less in solar maximum than in solar minimum, whereas they are considerably greater in winter in solar maximum than they are in solar minimum. However, given that meridional wind changes over a solar cycle are relatively small and are in the wrong direction to produce the solar cycle variations in NmF2, changes in wind cannot drive these changes.

Another issue is whether the composition results at ~150 km can be extended to the height of the F2 electron density peak. The first problem with attempting to do this is that most ionosondes do not provide a height for the F2 peak, so there is a significant uncertainty in knowing the height to which the GUVI data should be extended.

Extending the GUVI [O]/[N2] data from 150 km to 300 km is, in principle, easy for models, but it involves assumptions, which may weaken the conclusions drawn rather than strengthening them. The first is that, in the case of MSIS, the atmosphere can be adequately represented by the modified Bates’ [1959] profile that is in MSIS. A brief digression is needed to explain the theoretical basis of the Bates’ profile. The Bates’ profile solves the thermal conductivity equation assuming a hot top boundary and a cold lower boundary. For this equation to be successful, no other heating or cooling can occur within the domain of the model. However, other sources of heating and cooling do occur in the thermosphere and ionosphere. Compressional heating and cooling by expansion change the temperature in different ways at different locations. No cooling alters the profile by effectively raising the altitude at which radiational cooling provides a cool lower boundary. No cooling is solar cycle dependent, with little occurring at solar minimum, but with this process having a significant impact at solar maximum [Roble et al., 1987]. For these reasons, using MSIS to extend this compositional information upward has potential errors.

Furthermore, neither MSIS nor first-principles models current capture adequately the solar cycle variations seen in the GUVI compositional data (e.g., see Qian et al. [2013], which shows that a first-principles models produces a winter anomaly at solar minimum that is not seen in the data). Without accurate temperatures, using models to extend composition over 200 km upward vertically will probably result in an incorrect representation of the composition at the F2 peak.

There is a large difference between the southern and northern annual [O]/[N2] responses. The annual change in this parameter is much greater in the Northern Hemisphere than in the southern one, perhaps causing stronger winter increases in electron density in the north than in the south. These changes in composition represent an annual anomaly, but its source cannot be identified with the data used in the current study.

There is also a considerable longitudinal signal in this annual variation, but the variation with solar cycle is similar in each longitudinal sector. The conclusions of this paper are, therefore, not affected by longitudinal variations in composition.
An apparent contradiction occurs. The meridional winds appear to change little over a solar cycle and hence should not be the source of the changes in the summer-to-winter variations in electron density. But the winds also control the composition. The horizontal winds are not the main driver for composition changes, rather it is the vertical winds that blow through pressure surfaces. [Burns et al., 1989]. Downward winds enhance the relative abundance of the light species like O, whereas upward winds enhance heavy species like N₂. If the horizontal strength of the horizontal winds does not change over a solar cycle, vertical winds should also remain the same. The most likely answer is that the winds do not act alone to change composition; they also interact with molecular diffusion. At solar minimum, heavy species are concentrated at lower altitudes than at solar maximum, and thus, molecular diffusion should be stronger in the lower thermosphere, reducing the effect of the vertical winds and the summer to winter composition gradient. Obviously, this concept requires further study.

6. Conclusions

Ionsonde, GUVI, and COSMIC data were used to investigate the changes that occurred in the winter anomaly over a solar cycle. The following results were found:

1. The winter anomaly ratio (WAR—defined as the ratio between the winter value of NᵥF₂ and the summer value on the same day of the year) changes over a solar cycle. It is strong at solar maximum, reaching a maximum in the last solar cycle of at least 150% of the summer value. It is virtually nonexistent at solar minimum.

2. The changes in WAR with solar cycle are primarily caused by the varying winter-to-summer [O]/[N₂] over this period. These seasonal differences are considerably larger at solar maximum than they are at solar minimum.

3. Changes in the recombination rate make a secondary contribution to the changing WARs over a solar cycle.

4. MSIS does not include this solar cycle variation of winter-to-summer [O]/[N₂].

5. The greater winter increases in electron density in the Northern Hemisphere than in the Southern Hemisphere appear to be related to the greater annual variation of [O]/[N₂] in the north than in the south.

6. Further knowledge is needed about a number of features of the lower thermosphere, including the variation of O and N₂ with altitude and how molecular diffusion changes over the solar cycle.

Acknowledgments

Robert Lysak thanks Edward Llewellyn and an anonymous reviewer for their assistance in evaluating this paper.

References


