An empirical model of the Earth’s horizontal wind fields: HWM07


Received 11 August 2008; revised 3 September 2008; accepted 15 September 2008; published 11 December 2008.

The new Horizontal Wind Model (HWM07) provides a statistical representation of the horizontal wind fields of the Earth’s atmosphere from the ground to the exosphere (0–500 km). It represents over 50 years of satellite, rocket, and ground-based wind measurements via a compact Fortran 90 subroutine. The computer model is a function of geographic location, altitude, day of the year, solar local time, and geomagnetic activity. It includes representations of the zonal mean circulation, stationary planetary waves, migrating tides, and the seasonal modulation thereof. HWM07 is composed of two components, a quiet time component for the background state described in this paper and a geomagnetic storm time component (DWM07) described in a companion paper.


1. Introduction

The empirical model described in this paper is the provisional (for reasons discussed in section 6) successor to the Horizontal Wind Model series (HWM87, HWM90, HWM93) [Hedin et al., 1988, 1991, 1996]. The HWM series includes the only parameterized global wind models of the lower and middle atmosphere (HWM93 and HWM07), as well as the only global empirical wind models of the thermosphere. These empirical models seek to describe the dynamics of the atmosphere, not explain it [Rishbeth, 2007]. For geophysical research applications, climatological empirical models provide a trade-off between the use of stored or precomputed patterns and full-scale theoretical calculations. HWM has been extensively utilized by the upper atmospheric research community.

Over the years a number of middle and lower atmospheric climatologies have also been developed. These were recently intercompared by the SPARC (Stratospheric Processes and their Role in Climate) program, as summarized by Randel et al. [2004]. Most are zonally averaged and tabular, covering the altitude range from the ground to approximately 120 km. Perhaps the most widely known of these are the COSPAR International Reference Atmosphere CIRA-86 [Rees et al., 1990 and references therein] and the Upper Atmosphere Research Satellite (UARS) Reference Atmosphere Project (URAP) wind climatology [Swinbank and Orland, 2003]. Wang et al. [1997] developed an empirical model of horizontal winds between 90 and 120 km from Wind Imaging Interferometer (WINDII) satellite instrument observations. Wang et al. [1997] also reviewed other climatological wind models based on rocketsonde measurements [Koshelkov, 1985], medium frequency (MF), and meteor radar observations [Miyahara et al., 1991; Portnyagin and Solovieva, 1992a, 1992b], and the combination of rocketsonde and some radar data [Koshelkov, 1990]. A number of monthly mean climatologies of global-scale vertically propagating waves such as migrating tides, nonmigrating tides, and stationary planetary waves have since been developed for the height range of 60 to 100 km [e.g., Talaat and Lieberman, 1999; Zhang...
et al., 2003; Manson et al., 2004; Oberheide et al., 2006, 2007; Wu et al., 2008a, 2008b]. One limitation of several of these climatologies is that they have been derived from a single instrument or type of instrument. The combination of many diverse ground- and satellite-based data sets can mitigate instrumental biases and thereby provide more accurate atmospheric specifications [Daley, 1991].

[4] As an empirical spectral climatology the HWM represents the predominant repeatable oscillations of the atmosphere (day to day and year to year). Purely random or stochastic variability from gravity waves and migrating planetary waves cannot be represented deterministically by the empirical models. In some instances, however, cause and effect relationships can be parameterized with geophysical indices such as $F_{10.7}$ or $Ap$. With its diverse observational database, HWM provides a framework for the statistical intercomparison and validation of various upper atmospheric measurements and theoretical models. HWM also provides background wind fields for studies of wave propagation [e.g., Drob et al., 2003], first principles ionospheric models [e.g., Huba et al., 2000], ionospheric data assimilation systems [e.g., Pi et al., 2003; Schunk et al., 2004], and studies of large and small-scale ionospheric processes such as dynamo electric fields [e.g., Alken et al., 2008]. HWM will continue to have numerous research and operational applications.

[5] The technical objective of the work presented is to address known HWM93 discrepancies by improving the mathematical formulation and assimilating new data sets. The component of HWM07 representing geomagnetically quiet conditions is described in this paper. The component representing storm-induced thermospheric disturbance winds, DWM07, is described in a companion paper by Emmert et al. [2008]. HWM07 provides an accurate functional global specification of thermospheric winds from the ground to space. The paper is organized in the following manner: the observational database (section 2), mathematical formulation (section 3), model parameter estimation procedure (section 4), overall results with comparisons to data (section 5), scientific discussion (section 6), and conclusions (section 7).

2. Observational Database

[6] The foundation of any reliable atmospheric specification, empirical or otherwise, is an adequate observational database. For the spatiotemporal domain considered here, no single observational data set is capable of providing complete coverage of the general circulation patterns of the atmosphere. Only through the amalgamation of many diverse data sets is a thorough survey of the wind fields possible. The present observational database includes the historical data sets previously incorporated in HWM93, as well as a significant number of new research measurements. In addition, the numerical results from a multiyear Thermosphere-Ionosphere-Mesosphere-Electrodynamics General Circulation Model (TIME-GCM) simulation [e.g., Roble et al., 1988; Roble and Ridley, 1994; Crowley et al., 1999, 2006a, 2006b] contributed to the reformulation of the new HWM empirical representation by providing supplemental synthetic data constraints in data-sparse regions.

2.1. HWM Historical Data Sets

[7] The historical data sets used to build the previous HWM models are of considerable scientific value. A brief synopsis of these data sets is presented here; additional details are given by Hedin et al. [1988, 1991, 1996]. The original HWM87 model [Hedin et al., 1988] was derived entirely from Atmospheric Explorer (AE-E) and Dynamics Explorer (DE 2) satellite observations. As a result, the model only described thermospheric winds above 220 km, with no altitude dependence. A rudimentary parameterization of geomagnetic activity level was included, but data coverage was insufficient to parameterize solar cycle effects. The HWM90 model [Hedin et al., 1991] extended HWM downward to 100 km and included height dependence, as well as solar cycle effects. To accomplish this, the model incorporated ground-based 6300 Å Fabry-Perot interferometer (FPI; ~250 km) and incoherent scatter radar (ISR; 90 to 400 km) observations. Early MF/meteor radar and rocketsonde data were also included. In HWM93 Hedin et al. [1996] extended the model to the ground by adding monthly average zonal mean and tidal amplitudes derived from MF/meteor radars (75 to 110 km), a very extensive rocket and radiosonde database (20 to 100 km), and gradient winds derived from CIRA-86 and MSISE-90 (0 to 120 km) [Hedin et al., 1991]. In the 0–35 km altitude region, HWM93 utilized a 4-year statistical summary of NCEP specifications [Wu et al., 1987].

[8] A number of technical challenges were overcome in the construction of the original HWM models. These included limitations in available computer memory, rudimentary visualization software, and sparse observational data coverage. The latter presented the greatest challenge. The AE-E Neutral Atmosphere Temperature (NATE) instrument [Spencer et al., 1973] only provided in situ cross track winds between ±20° latitude. The DE 2 zonal and meridional wind components were derived from two separate instruments: The Wind and Temperature Spectrometer (WATS) [Spencer et al., 1981], which measured in situ zonal winds, and FPI [Hays et al., 1981], which remotely measured meridional winds near 250 km. All three of these instruments only made measurements at a single height. Information on the vertical structure could only be obtained statistically via the varying altitude provided by the elliptical orbits. Such analysis is complicated by incomplete parameter space coverage resulting from the convolution of altitude sampling with that of local time, season, and latitude. The ISR measurements did provide contemporaneous vertically resolved information, but are limited to daytime conditions in the lower thermosphere or to line-of-sight (LOS) observations along the direction of the magnetic meridian in the upper thermosphere. Despite the limitations all these data sets made valuable scientific contributions to the understanding of the dynamics of the upper atmosphere. In the historical HWM database there are a total of 1.2 × 10⁶ observations from 35 different instruments spanning a total of 15 years.

2.2. New Data Sets

[9] We have since revised and added numerous ground- and satellite-based observational data sets to the database. There are now a total of 60 × 10⁶ observations from 35 different instruments spanning 50 years. This is approxi-
mately a 50 fold overall increase in the number of observations. Table 1 provides a synopsis of the new additions, as well as those data sets carried over to construct the new model. A visual comparison of the spatiotemporal coverage of the observations used to generate the HWM93 and HWM07 models is displayed in Figure 1. Figure 1a shows the latitude versus altitude cross section of the measurements for HWM93, while Figures 1b shows the corresponding cross section of a fraction of the measurements (<1% of the data points) used to construct HWM07 (see section 4). Figures 1c and 1d show the latitude versus local time cross section for the HWM93 and HWM07 data sets, respectively. Observations from the Wind Imaging Interferometer (WINDII) [Shepherd et al., 1993] onboard the UARS is the most extensive thermospheric data set that was added. From a circular orbit at 585 km, WINDII measured height

Figure 1. Comparison of HWM93 and HWM07 spatiotemporal coverage of available data sets; (a) latitude versus altitude cross section of the measurements used to construct HWM93, (b) the corresponding cross section of a fraction of the measurements (<1% of the data points) used to construct HWM07 (see section 4), (c) latitude versus local time cross section for the HWM93 data sets, and (d) a corresponding fraction of the available measurements from the HWM07 database. The data sets below 100 km have been omitted for clarity.
profiles of Doppler shifts in various airglow emissions, from which height-resolved vector horizontal winds were derived. The most frequently observed optical emission lines were the 557.7 nm OⅡ green line (90−300 km altitude during the day, 90−110 km at night) and the 630.0 nm OⅠ red line (daytime 125−300 km, nighttime 225−300 km). The height resolution is 3 km below 120 km, and 5 km above 120 km. The data cover latitudes up to 72°, except for nighttime upper thermospheric measurements, which only extend to 42°. For HWM07, we used green and red line observations from data version 5.11, which covers the period November 1991 to August 1997 and includes 595,000 height profiles measured on 1109 different days.

The WINDII data are supplemented by new and extensive ground-based FPI measurements from 11 different locations. These additions, available from the CEDAR database (http://cedarweb.hao.ucar.edu), improve latitudinal coverage, particularly in the southern hemisphere. Several of the data sets cover one or more solar cycles. Additional details regarding these instruments, reduction methods, and coverage are provided by Emmert et al. [2006a] and the CEDAR database.
The UARS High Resolution Doppler Imager (HRDI) [Hays et al., 1993] measurements are another important addition to the HWM database; adding spaced based winds in the 50 to 115 km region. Together with WINDII data, the HRDI measurements revolutionized the understanding of the dynamics of the mesosphere and lower thermosphere (MLT); including knowledge about the zonal mean circulation, interannual variability, and the semiannual, migrating, and nonmigrating tidal oscillations [e.g., Burragge et al., 1993, 1996; Lieberman et al., 1993; McLandress et al., 1996; Huang and Reber, 2003; Wu et al., 2008a, 2008b].

Ground-based instruments also provided extensive and complementary knowledge. New ground-based MLT data sets added to the HWM database include MF/meteor radars [Vincent and Lesicar, 1991; Palo et al., 1997; Murayama et al., 2000], Na density-temperature-wind LIDAR [She et al., 2004] and rocket born trimethyl aluminum (TMA) release [Larsen, 2002] measurements. Last, a 5-year archive of four times daily NOAA NCEP Global Forecast System (GFS) [Kanamitsu, 1989] and NASA GEOS-4 (Goddard Earth Observing System version 4) [Bloom et al., 2005] analysis fields (derived from numerous meterological data sources) provides information for HWM in the lower and middle atmosphere (0–55 km).

Even with the wealth of newly available data sets, several notable gaps in the observational data coverage exist. Owing to the multidimensional nature of the model space these may not readily apparent in Figure 1. The most significant are (1) nighttime wind profiles between 120 and 200 km, (2) high-latitude observations poleward of 70° from 95 to 350 km with the exception of FPI measurements at 250 km, and (3) mid- to high-latitude southern hemisphere summer winds above 200 km. The consequences and management of these gaps are discussed in subsequent sections.

2.3. TIME-GCM Data Set

To gain physical insight during the reformulation and production of the new HWM, a multiyear synthetic database was generated using the TIME-GCM theoretical general circulation model. The TIME-GCM was developed at NCAR by Roble and Ridley [1994], and later converted to a workstation environment by Crowley et al. [1999]. The TIME-GCM model includes fully time-dependent space weather drivers representing high-latitude electrodynamic inputs, solar irradiance, and tidal lower-boundary conditions. The cross-polar cap potential difference is obtained using a modified form of Weimer’s [1995] potential model, which requires Interplanetary Magnetic Field and solar wind parameters as input. This potential difference is then applied to TIME-GCM via Heelis et al.’s [1982] potential model as described by Roble and Ridley [1987]. The auroral particle precipitation parameterization was also described by Roble and Ridley [1987]. The solar flux specification uses the EUVAC [Richards et al., 1994] model driven by the \(F_{10.7}\) index. The upward propagating tidal specification at the TIME-GCM lower boundary uses a tidal climatology for the diurnal and semidiurnal tides [Fesen et al., 1986]. The synthetic database obtained from TIME-GCM is composed of a continuous 4-year simulation, with global 3-D output from approximately 30 to 500 km at 1-h intervals.

Similar models such as the Canadian Middle Atmosphere Model (CMAM) have also been used by others to perform Observational System Simulation Experiments (OSSE); for example in order to analyze the influence of data gaps in satellite observations on the aliasing of zonal mean winds and migrating tides in the lower thermosphere [e.g. McLandress and Zhang, 2007]. The TIME-GCM includes enough dynamic complexity to evaluate the robustness of the new HWM formulation as it relates to the available observational coverage. Initially candidate HWM formulations were fit to evenly distributed random statistical samples of the TIME-GCM histories. The results were then compared to the original global TIME-GCM output fields. This facilitated the discovery and rectification of a HWM93 vertical derivative constraint artifact in the 160–220 km region (see section 5). Furthermore, the TIME-GCM output indicates that the altitude at which the thermospheric wind velocity profile becomes constant depends on solar activity, a feature not captured by the HWM models (including HWM07).

3. Mathematical Formulation

Textbooks on the topic of data assimilation for geophysical state specification include Daley [1991] and Wunsch [1996]. Aspects of spectral aliasing and the generation of spurious harmonics during the analysis of sparse asymptotically sampled data sets has been explored in some detail by Salby [1982], Forbes et al. [1997], and McLandress and Zhang [2007]. The overarching principle for success in these enterprises is summarized by Daley [1991]; the spectral formulation must not be so simple that the data are underfit, resulting in a crude specification unrepresentative of the data, or so complex that data are overfit, providing a representation wrought with spurious artifacts.

In order to represent the predominant seasonal and diurnal variations of the general circulation of the atmosphere the HWM93 formulation was based on a set of truncated Vector Spherical harmonics (VSH [e.g., Morse and Feshbach, 1953; Swartztrauber, 1993]) that are Fourier modulated in time. To represent geomagnetic activity and solar flux variability effects, low-order polynomials multiplying select terms were included. The vertical variations were represented by cubic spline interpolation of a number of nodes on which the horizontal basis functions are defined. Above the last node at 200 km, the vertical wind profile was represented by a modified Bates-Walker function [Hedin et al., 1991] such that the winds become nearly constant with altitude above approximately 300 km. In order to implement this, a VSH expansion of the vertical derivative at 200 km was required. Unfortunately, this expansion, combined with incomplete data coverage between 120 and 220 km resulted in spurious artifacts in the region.

During HWM93 model generation the statistical significance of the various harmonics was evaluated by considering the ratio of the estimated parameter values to their respective estimated uncertainties. Any harmonics where this ratio was found to be less than approximately 0.3 were subsequently omitted from the generalized VSH expansion. Although this is a sound strategy for avoiding overfitting and reducing the number of unknown parameters
to be estimated, it later made the bookkeeping necessary to routinely upgrade the model difficult. Furthermore, the HWM93 model formulation required that the unknown parameters be estimated by a nonlinear least squares inversion technique. Despite these drawbacks the underlying concept of Fourier-modulated VSH functions with cubic splines in the vertical is far from obsolete. A similar approach was employed in the VSH model of Killeen et al. [1987] to represent a set of theoretical TIME-GCM temperature and wind fields.

[20] From these concepts and lessons learned, the spectral formulation of HWM07 includes several orthogonal Fourier-modulated VSH basis sets at a number of vertical levels

\[
\begin{align*}
\psi_{1}^{l,n,s}(\alpha, \theta, \tau) &= \frac{-\cos(g \alpha)}{\sqrt{n(n+1)}} \frac{P_{n}^{l}(\theta)}{d\theta} \cos(\tau \theta) + \sin(\tau \theta), \\
\psi_{2}^{l,n,s}(\alpha, \theta, \tau) &= \frac{\sin(g \alpha)}{\sqrt{n(n+1)}} \frac{P_{n}^{l}(\theta)}{d\theta} \cos(\tau \theta) + \sin(\tau \theta), \\
\psi_{3}^{l,n,s}(\alpha, \theta, \tau) &= \frac{-\sin(g \alpha)}{\sqrt{n(n+1)}} P_{n}^{l}(\theta) \cos(\theta) \cos(\tau \theta) + \sin(\tau \theta), \\
\psi_{4}^{l,n,s}(\alpha, \theta, \tau) &= \frac{-\cos(g \alpha)}{\sqrt{n(n+1)}} P_{n}^{l}(\theta) \cos(\theta) \cos(\tau \theta) + \sin(\tau \theta),
\end{align*}
\]

(2)

In this equation, \(u(\tau, \lambda, \delta, \theta)\) represents the zonal wind field at a given level as a function of the year (\(\tau\)), longitude (\(\lambda\)), local solar time (\(\delta\)), and latitude (\(\theta\)) respectively. The seasonal, local time, longitudinal, and latitudinal spectral wave numbers are \(s, l, m, n\) respectively. This notation differs slightly from the notation of classical tidal theory [e.g., Forbes, 1995] in which \(s\) represents the longitudinal wave number. The first summation in equation (1) represents the seasonal modulations of the zonal mean circulation, the second summation the migrating solar tidal components and seasonal modulations thereof, while the last summation the stationary planetary waves with their respective seasonal variations. The modulated VSH basis functions \(\psi_{i}^{l,n,s}\) are

\[
\begin{align*}
U^{l,n,s}(\tau, \lambda, \delta, \theta, z) &= \sum_{j=1}^{M} \beta_{j}(z) U_{j}(\tau, \lambda, \delta, \theta),
\end{align*}
\]

(3)

where \(\beta_{j}(z)\) is the amplitude of the \(j\)th weighting kernel with \(U_{j}(\lambda, \delta, \theta)\) the corresponding horizontal wind vector from (1) and the parity relationship. Figure 2 shows the vertical B-spline weighting functions and intervals for the portion of the model above 100 km. There are a total of 30 kernels defined by evenly spaced nodes at 5 km intervals from 0 to 110 km with unevenly spaced nodes at 117.5, 125, 135, and 150 km.

[22] Two hybrid basis functions (red curves in Figure 2) were created to represent the transition to an asymptotic wind profile in the upper thermosphere. These functions are defined so that they approach 0 or 1 over an exponential scale-height of 60 km, while satisfying second-order continuity and unit normalization with the remaining vertical functions. By contrast, the HWM93 transition to a modified Bates-Walker profile at 200 km, which is continuous only up to the first derivative. Relaxation to a constant thermospheric wind profile is a reasonable assumption given the fact that the effective viscosity increases dramatically with the mean free molecular path (and therefore with altitude), thereby inhibiting the growth of vertical sheers for all but daytime solar maximum conditions when ion drag becomes important [e.g., Kohl and King, 1967; Rishbeth, 1972; Hedin et al., 1991]. Furthermore, this is currently the only viable option given that the amount of data above 300 km is insufficient to identify any systematic height variations in the vertical gradients. As compared to HWM93 the new formulation increases the altitude at which the transition to an asymptotic profile occurs, providing greater flexibility in describing the vertical gradients between 200 and 300 km, which can now be resolved with the addition of WINDII profiles.

[23] On the basis of fitting candidate formulations to the newly available data sets, the maximum spectral orders of \(S = 2\) (annual and semiannual), \(L = 3\) (diurnal, semidiurnal, and terdiurnal), \(M = 2\) (planetary waves), and \(N = 8\) (latitude) were selected for the provisional HWM07. No statistically significant migrating tidal amplitudes (>1 m/s) were observed in the data below about 55 km, so no tidal harmonics are included in the first 12 levels. As a result there are either 376 or 796 unknowns per model level, with a total of 18,840 unknown model parameters over the entire model domain.

4. Parameter Estimation

[24] The principal computational challenge to constructing the HWM is the estimation of the unknown spectral coefficients \(\{C_{l,n,s}, B_{l,n,s}, C_{l,n,s}, \ldots, B_{l,n,s}\}\) from the sparse, multidimensional observational data sets. One of the reasons for overhauling the HWM93 formulation is to make it easier to routinely and reliably estimate additional unknown model parameters from newly available data sets.
The formulation given by equations (1)–(3) expresses the global time-dependent variations of atmospheric wind fields as a manageable set of linear basis functions. A generalized linear expression (\( \mathbf{Gm} = \mathbf{d} \)) for the new HWM formulation and estimation procedure can be written as

\[
\mathbf{G} = \begin{bmatrix}
\beta_1 \psi_1 & \beta_2 \psi_1 & \beta_3 \psi_1 & \beta_4 \psi_1 \\
\beta_1 \psi_2 & \beta_2 \psi_2 & \beta_3 \psi_2 & \beta_4 \psi_2 \\
\beta_1 \psi_3 & \beta_2 \psi_3 & \beta_3 \psi_3 & \beta_4 \psi_3 \\
\beta_1 \psi_4 & \beta_2 \psi_4 & \beta_3 \psi_4 & \beta_4 \psi_4 \\
\beta_1 \psi_5 & \beta_2 \psi_5 & \beta_3 \psi_5 & \beta_4 \psi_5 \\
\beta_1 \psi_6 & \beta_2 \psi_6 & \beta_3 \psi_6 & \beta_4 \psi_6 \\
\beta_1 \psi_7 & \beta_2 \psi_7 & \beta_3 \psi_7 & \beta_4 \psi_7 \\
\beta_1 \psi_8 & \beta_2 \psi_8 & \beta_3 \psi_8 & \beta_4 \psi_8 \\
\end{bmatrix}
\]

\[
\mathbf{d} = \begin{bmatrix}
\mathbf{d}_1 \\
\mathbf{d}_2 \\
\mathbf{d}_3 \\
\mathbf{d}_4 \\
\mathbf{d}_5 \\
\mathbf{d}_6 \\
\mathbf{d}_7 \\
\mathbf{d}_8 \\
\end{bmatrix}
\]

The vectors \( \mathbf{d}_i \) represents a subset of zonal, meridional, or line of sight (LOS) wind observations in the altitude interval \( i \), partitioned into 27 altitude intervals: The first interval \( (i = 0) \) is \( z > 150 \), and the remaining intervals are defined by the nodes of the weighting kernels. Within each interval there are four nonzero weighting kernels, as shown in Figure 2. The rows of \( \mathbf{G} \) consists of elements from four submatrices \( \beta_1 \psi_j \ldots \beta_4 \psi_j \). The vector \( \beta_j \) contains values of the \( j \)th altitude weighting kernel corresponding to vector of observations \( \mathbf{d}_i \). The columns of the submatrix \( \beta_j \psi_j \) correspond to the weighted spectral modes of the VSH basis functions of the \( j \)th model level. The product \( \beta_j \psi_j \) represents the multiplication of each column of \( \Psi_j \) by the vector \( \beta_j \) without summing. In general, the horizontal parameterization of the model level can be different. For all model levels at or above 55 km there are 796 columns in \( \mathbf{g}_j \) corresponding to the summed functions in equation (2). For all model levels below 55 km, the migrating tidal terms are omitted, and \( \mathbf{g}_j \) contains 376 columns. The elements of each row in \( \mathbf{G} \) are multiplied by the model parameter vectors \( \mathbf{m}_j \ldots \mathbf{m}_{j+3} \), where \( \mathbf{m}_j \) contains the unknown spectral coefficients \{\( C_{l,j} \), \( B_{l,j} \), \( C_{l,j} \), \ldots \} for the \( j \)th model level.

**Figure 2.** Vertical cubic B-spline basis functions \( \beta_j \) and corresponding data intervals \( \mathbf{d}_i \) for the new HWM model. The last two basis functions are constructed to approach either 0 or 1, subject to continuity and derivative constrains with the remaining functions.
As there are far more wind observations than model parameters the generalized linear expression given in (4) represents an overdetermined inverse problem [Menke, 1989]. An estimate of the unknown model parameters \( \hat{m} \) can be obtained via a weighted least squares inversion given by

\[
\hat{m} = \left( G^T \Sigma_d^{-1} G \right)^{-1} G^T \Sigma_d^{-1} d,
\]

where \( \Sigma_d \) is the data covariance, here constructed as a diagonal matrix of the measurement uncertainties. There are far too many observations and unknown model parameters over the entire model domain to compute (5) with conventional linear algebra packages. To handle this problem, random data sampling combined with a novel sequential estimation procedure is used. The unknown model parameters in equation (4) are estimated sequentially with the help of a standard iterative least squares estimation procedure [Rogers, 2000]

\[
\hat{x} = x_a + (K^T S_e^{-1} K + S_a^{-1})^{-1} K^T S_e^{-1} (y - Kx_a) \quad (6)
\]

\[
\hat{S} = (K^T S_e^{-1} K + S_a^{-1})^{-1}. \quad (7)
\]

The vector \( \hat{x} \) represents an estimate of a subset of the 18,840 unknown model parameter \( m_j \ldots m_{j+3} \), \( x_a \) is the a priori model parameter estimates from a previous iteration and/or data subset, \( y \) is a vector subset of the 60 x 10^6 observations \( d_1 \ldots d_{i+n} \), \( K \) is the corresponding subset of the forward spectral matrix \( G \), \( \hat{S} \) is the covariance matrix of the parameter estimate, \( S_a \) is the covariance matrix of the a priori parameter estimates, and \( S_e \) is the covariance matrix of observations \( y \).

To capture the vertical covariance among the various terms in equations (1)–(3) as provided by the observational data, the simultaneous estimation of a reasonable number of vertical levels is required. The submatrices for the first estimation step are indicated by the shaded regions in equation (4). For HWM07, the parameters for 12 vertical basis functions, corresponding to the observations from 9 altitude intervals are chosen for each sequential estimation step; \( \Delta z = 45 \) km in the lower atmosphere. This number of levels is chosen on the basis of the number of unknown model parameters and data that can be processed in each step, given the available computational resources. For each application of equations (6) and (7) there are anywhere from 4464 to 8244 unknown model parameters. Approximately 66% of the elements of forward matrix \( K \) are zero. This also has the desirable effect of introducing a lower limit of 25 m/s; a value representative of the geomagnetically quiet conditions data points above 110 km for which the 3-h \( v \) is greater than 12 are also rejected. The storm time winds are provided by DWM07, which is described in the companion paper by Emmert et al. [2008].

Depending on the altitude region, approximately 150,000 to 2500,000 observations are selected for each individual sequential estimation step. The resulting observational subset is large enough to ensure that the linear system (\( d = Kx \)) being estimated in equation 5 and 6 is overdetermined by a factor of 15 to 30. A new subset of observations is selected for each iterative step during the downward pass of a cycle, with the same random data subsets being used for the returning upward pass. The uncertainties reported with the observations is floored at a lower limit of 25 m/s; a value representative of the geophysical variance of stochastic phenomenology not resolvable by the formulation. Any reported measurement uncertainties between 25 m/s and 150 m/s are left unchanged. For these reasons, not to mention computational tractability, we therefore do not include off diagonal terms in the data covariance matrices (\( S_e, \Sigma_d \)).

In order to improve model parameter convergence and robustness a preliminary sequential pass is performed using a data set of evenly sampled global HWM93, TIME-GCM, NCEP-GFS, and GEOS-4 output fields. This preconditioning the first guess to reasonable values rather than zero. This also has the desirable effect of introducing a limited set of empirical and theoretical synthetic data to...
provide soft constraints for those regions of the model space where no direct observations exist. To reflect the theoretical nature of the synthetic data the corresponding observational uncertainties are set to 75 m/s. Numerical experiments indicate that the choice of initial conditions and synthetic data sets have no statistically significant effect on the results in data-dense regions; however outside of those regions the inclusion of these synthetic and empirical data sets mitigates the occurrence of spurious artifacts. With the inclusion of the synthetic data as soft constrains it is therefore possible to increase the spectral resolution needed to faithfully represent geophysical variations in the data-dense regions without adversely effecting the model results in data-sparse regions. Without this approach, the absence of any significant nighttime data in the lower thermosphere between 110 and 200 km and at high latitudes between 55 and 200 km would normally limit our ability to include higher-order spectral harmonics that are needed to accurately represent observed variations in the daytime thermosphere. On the basis of these conclusions, during subsequent estimation the synthetic TIME-GCM output is continually introduced in those regions of the model space containing no observational data. As compared to the other data sets the amount of TIME-GCM data used is limited in terms of quantity and spatiotemporal coverage.

5. Model Example/Comparisons

Highlights of the most important results of the provisional HWM07 are presented in this section, leaving more detailed validation of the model and evaluation of statistical performance measures to subsequent publications. Important morphological improvements include: significant change in the vertical structure of the winds between 120 and 250 km; increased amplitudes of the polar stratospheric zonal wind jets, stationary planetary wave structures, tro-
pospheric flows, and migrating tides. These improvements are facilitated by the increased information content of newly available observations in conjunction with the increase in the total number of unknown model parameters, including additional vertical model levels.

Figure 3 shows a comparison of HWM07 (Figure 3 (left)) and HWM93 (Figure 3 (right)) wind fields for Northern Hemisphere summer (day 181) from 0 to 300 km at local noon (1200 UT at 0°C176E). Several differences are obvious. For example, a slight glitch in the HWM93 wind fields near 200 km, which is caused by a derivative constraint, is no longer present in HWM07. The magnitude of the eastward Southern Hemisphere winter time stratospheric jet is increased from 80 m/s to over 100 m/s. Similar increases are true for the opposite hemisphere and solstice. Differences in the tidal structure between 90 and 150 km can also be seen; particularly in the zonal wind near the equator at 120 km and the Southern Hemisphere above 150 km. The meridional wind components are different as well. At this longitude and local time the thermospheric wind velocities above 150 km are also somewhat reduced, with the exception of the high-latitude Northern Hemisphere zonal wind component above 200 km.

Figure 4 compares HWM07 (Figure 4 (left)) and HWM93 (Figure 4 (right)) wind fields for spring equinox conditions (day 91) from 0 to 300 km at local noon (1200 UT at 0°C176E). Significant differences are notable again. For this example the thermospheric wind velocities above 150 km are generally reduced, particularly the high-latitude Northern Hemisphere zonal component. The classical alternating asymmetric migrating diurnal (1,1) meridional wind pattern [e.g., Burrage et al., 1993] between 80 and 130 km near the equator with its corresponding zonal wind structure, both absent in HWM93, can be seen in the HWM07 wind fields. The dominant migrating diurnal tide between 85 and 115 km has maximum meridional wind velocities approaching 60 m/s at ±22° with an average vertical wavelength of approximately 20 km. In addition, the
spurious 80 m/s eastward zonal wind enhancement located at the equator near 130 km in HWM93, which is related to an anomalous semidiurnal tidal component, was removed from HWM07. Furthermore, the weak equinox stratospheric westerlies, as well as the average topospheric jet streams, are better resolved by the new model.

[35] More important than how the models intercompare is how HWM07 directly compares to the observational data sets. The ability of HWM07 to faithfully represent the observed variations of the atmosphere is markedly improved over that of HWM93. Figure 5 illustrates improvement in the upper thermospheric vertical structure. It shows average zonal winds from WINDII and WATS as a function of height, along with corresponding HWM93 and HWM07 profiles. Although there is almost no overlap between the WINDII and DE 2 WATS data, their vertical structure is mutually consistent within the statistical uncertainties. HWM07 represents the gradients, whereas HWM93 exhibits unsupported variations between 150 and 250 km, due to a lack of daytime data in this region and a derivative constraint at 200 km. The HWM93 vertical structures between 150 and 200 km are completely inconsistent with the WINDII observations.

[36] Figures 6 and 7 further illustrate these improvements; they show a comparison of the HWM93 and HWM07 models with the bin averaged zonal and meridional WINDII observations for equatorial latitudes and December solstice conditions for different daytime local time sectors be-
The HWM07 model (blue line) is in excellent agreement with the seasonally averaged data. Spurious artifacts in the HWM93 model (red line) are readily apparent. There is a slight vertical phase shift on the order of a few kilometers between HWM07 meridional winds and the WINDII observations. The cause is unknown at this time.

In the MLT, the specifications of the diurnal and semidiurnal migrating tidal components have been completely revised. Their amplitude and phases are now quantitatively consistent with published results of observations from the WINDII, HRDI, and ground-based instruments. This is highlighted in Figure 8, which shows the seasonally averaged wind data in 1-h time-of-day bins at 95 km (Figure 8 (top and top middle)) and 85 km (Figure 8 (bottom middle and bottom)) between 38° and 48°N from WINDII (blue line), HRDI (green line), Fort Collins Lidar (orange line), and Wakkanai MF Radar (red line) instruments as compared to the HWM93 (dotted black line) and HWM07 (solid line) model output. The wind observations were averaged in overlapping 2-h bins with the error bars representing the standard deviation of the mean. The WINDII, HRDI, Fort Collins Lidar, and HWM07 models are in qualitative agreement. Figure 8 (top) illustrates that the observed midlatitude semidiurnal zonal wind oscillations at 95 km given by HWM07 and the WINDII, HRDI, and Fort Collins Lidar observations, is almost twice that of HWM93 and the Wakkanai MF radar observations. The zonal mean wind component appears to be greater in the Northern Hemisphere during summer solstice as described by both models and all four data sets. For the meridional component both models and the observational data sets are however in reasonable agreement. At 85 km there is better overall agreement between HWM07 and all four data sets. The HWM93 and Wakkanai radar observations show slightly smaller amplitudes and a phase shift. A 30 m/s eastward zonally average

Figure 6. Average WINDII zonal winds as function of height (black line) and corresponding results from HWM93 (red line) and HWM07 (blue line). The results represent quiet time (Kp < 3), December solstice (November–February), conditions in the indicated local time and latitude bins. The models were evaluated for the conditions of each observation then binned and averaged in the same way as the data. Error bars denote the estimated uncertainty of the mean.
There is a wind component seen during Northern Hemisphere Winter solstice at 85 km that is also absent in HWM93.

6. Discussion

[38] We have described the composite data sets, mathematical formulation, and parameter estimation procedure of the provisional HWM07 empirical wind model. The ability of HWM07 to faithfully represent the observed atmospheric variability is markedly improved over that of HWM93. The new HWM spectral formulation is organized in such a way as to be generalizable to any maximum wave number (i.e., spectral resolution) given suitable data, as well as to eventually include nonmigrating tidal components. To account for the observed seasonal variations, the HWM migrating tidal expression now includes completely coupled annual and semiannual modulation terms. Any aliasing between the zonal average and migrating tidal components [e.g., Forbes et al., 1997] should be minimal because all of the various tidal components including the zonal mean, plus the seasonal modulations thereof, are estimated simultaneously [e.g., Drob et al., 2000].

[39] The resulting specification represents the combined statistical average of observed satellite- and ground-based tidal amplitudes while simultaneously accounting for spatiotemporal sampling effects. In the absence of a priori information to determine if wind velocities measured by space-based optical means are overestimated, or conversely if the wind velocities measured by ground-based MF radars are underestimated, this approach provides a reliable means for obtaining an impartial statistical consensus. An alternate approach was taken for the empirical model of prevailing winds between 80 and 100 km developed by Portnyagin et al. [2004] whereby the HRDI measurements were shifted by $\frac{C_0}{6}$ m/s and then divided by a scaling factor of 1.7 in order to obtain agreement with the measurements from 46 ground-based MF/meteor radars. A methodology for the possibility of estimating time independent zero wind offsets (but not amplitude scaling factors) for each instrument and look direction while also simultaneously accounting for the geophysical variations was presented by Drob et al. [2000]; this may be considered in future HWM updates.

[40] Although HWM07 shows dramatic improvement over HWM93, there are several factors that motivate categorization of HWM07 as provisional. The most important is

Figure 7. Same as for Figure 6 but for the meridional component.
the potentially poor representation of thermospheric high-latitude circulation: neither HWM93 nor HWM07 appear to sufficiently describe the ion-convection-driven patterns that are observed in high-latitude wind measurements [e.g., Hernandez et al., 1991; Thayer and Killeen, 1993; Emmert et al., 2006b]. Ion drag is a dominant momentum source for the high-latitude thermosphere and ion convection patterns are organized by the geomagnetic field. Thus even under geomagnetically undisturbed conditions high-latitude winds are better organized in magnetic coordinates rather than geographic coordinates [e.g., Thayer and Killeen, 1993; Emmert et al., 2006b]. In contrast, quiet time low- and midlatitude winds are primarily driven by solar-induced pressure gradients, better represented in geographic coordinates. This presents a challenge for representing wind patterns with sparse data and relatively low global resolution. In geographic coordinates, higher-order harmonics are required to adequately represent high-latitude patterns [Thayer, 1990], with sufficient data coverage needed to avoid spurious artifacts. A novel approach, whereby the quiet time effects of high-latitude thermospheric ion neutral coupling is represented in (1) by also including a set of VSH harmonics expressed in rotated dipole magnetic coordinates is being studied.

Figure 8. Average winds in 1-h time-of-day bins at (top and top middle) 95 and (bottom middle and bottom) 85 km between 38 and 48°N from WINDII (blue line), HRDI (green line), Fort Collins Lidar (orange line), and Wakkanai MF Radar (red line) as compared to the HWM93 (dotted black line) and HWM07 (solid black line) model output.
Other known limitations of the provisional HWM07 are as follows. First, the model contains no dependence on solar irradiance variations; HWM93, by contrast, included a fairly simple dependence on solar cycle. Solar irradiance effects are generally small on the dayside [e.g., Emmer, 2001], but are substantial on the nightside [e.g., Biondi et al., 1999; Emmer et al., 2003]. Second, variations in the interplanetary magnetic field (IMF) significantly alter the quiet time high-latitude patterns [e.g., Emmer et al., 2006b], but HWM does not account for IMF effects. At present this is moot given that the model does not yet properly represent the overall high-latitude circulation. Finally, other than the stationary waves given by the third sum in (1), HWM07 does not include any representation of the nonmigrating tidal effects. It is well known that there are a number of nonmigrating tidal modes with significant amplitude in the MLT region [e.g., Talaat and Lieberman, 1999; Forbes et al., 2003; Oberheide et al., 2007]. This includes the diurnal eastward-propagating tide with zonal wave number 3 (DE3), which is sometimes the strongest oscillatory mode observed in MLT zonal winds [Oberheide et al., 2006].

Last, the HWM07 upper thermospheric winds start to approach a constant value above approximately 275 to 325 km. As mentioned in section 3, the present parameterization is motivated by geophysical considerations and data coverage limitations. Wang et al. [2008] showed that quiet time winds during solar minimum conditions in TIME-GCM simulations generally become asymptotic above about 250 to 300 km. This is consistent with the HWM07 vertical structure shown in Figure 4, supporting our current choice of parameterizations. However, the multyear synthetic TIME-GCM output fields described in section 2 indicate that the altitude at which the thermospheric winds become asymptotic is a function of solar irradiance. Additional flexibility in the HWM transition and scale heights is therefore warranted, provided it can be resolved by observational data.

Despite these caveats, HWM can support various scientific research activities in the various topic areas of atmospheric and space science. For example, the model provides a good representation of the atmospheric wind fields for idealized process studies of upward wave propagation and forcing of the ITM system from below. Furthermore, although operational Numerical Weather Prediction models may eventually span from the ground to the thermosphere [e.g., Akmaev et al., 2008] within the next decade, this will not make statistical empirical models obsolete. The development, validation, and day-to-day operations of such data assimilation systems will require accurate high fidelity background fields for comparison. Empirical models such as HWM also provide a means to control the dynamical stability or drift of these models via nudging techniques [e.g., Stauffer and Seaman, 1990]. Furthermore, by reformulating or totally removing the terms involving seasonal time scales, the HWM spectral function fitting scheme could provide a good framework for the thermospheric variational data analysis component of an operational numerical weather prediction system.

The flexibility of the new HWM formulation makes it easier to improve the model as new data sources become available. For example, additional data sets (e.g., from the NASA Thermosphere Ionosphere Energetic and Dynamics Mission) to resolve higher-order harmonics could provide a better representation of the abrupt onset of tidal maxima that occurs during equinoct. The pressing need for new observational techniques and missions to measure winds in sparse regions has been clearly identified during this survey. Other than a few sporadic rocket flight there are no nighttime wind observations between 130 and 200 km. Furthermore, although WINDII provided continuous profiles over the 90 to over 275 km altitude range, there are no direct measurements of continuous wind profiles that cover the altitude range from 250 to 500 km. Such measurements are important for understanding the role of ion neutral coupling, including geomagnetic forcing and solar cycle variations. New direct wind measurements in the equatorial regions above 300 km would also be useful. Observations of neutral winds for the specification of thermospheric polar region dynamics are also needed. The current lack of observations significantly limits our ability to validate theoretical understanding of the vertical structure and dynamics of the thermosphere.

7. Summary

The Horizontal Wind Model synthesizes the information content of many diverse observations and past calculations into summary patterns for future use. By overhauling the model parameterization and adding a significant number of new data sets the HWM93 model has been transformed into the provisional HWM07. The new formulation represents the zonal mean, migrating tides, stationary planetary waves, and seasonal modulations thereof via Fourier-modulated Vector Spherical Harmonics. The vertical variations are represented by cubic B-spline weighting kernels, with two specialized hybrid basis functions at the top of the model. There are 18,840 unknown model parameters that are estimated with a novel sequential estimation process from 60 × 10⁶ available data points, from 35 different instruments that span a period of over 50 years. Model output from a 4-year TIME-GCM simulation provided soft synthetic data constraints in several data-sparse thermospheric regions, permitting an increase in resolution in data-dense regions. The ability of HWM07 to represent the available observational data sets, and thus the true behavior of the atmosphere, is dramatically improved over that of HWM93; in particular the low- to midlatitude tidal oscillations between 80 and 150 km. By itself, HWM07 only represents geomagnetically quiet conditions; a component, DWMO7, representing storm-induced thermospheric disturbance winds is described in a companion paper [Emmer et al., 2008].

Acknowledgments. The effort to improve the thermosphere portion of the model, including its overall mathematical formulation, was supported by NASA Living With a Star (LWS) program grant 04-000-0098. Additional support to improve the 0–120 km portion of the model for ground-based detection of explosions was provided by U.S. Army Space and Missile Defense Command (SMDC). Supplemental support was provided by the Office of Naval Research (ONR). We wish to thank the many National Science Foundation CEDAR database contributors, as well as the NASA satellite data providers not listed as coauthors, including the NASA-GSFC GEOS-4 and NOAA-NCEP data providers. G.C. acknowledges support from LWS contract 04-000-0098 and Navy contract N0017306C6014. M.L. acknowledges support by NASA grant NNG05WC40G and NSF grant ATM-0541593. The Resolute FPI obser-
viation is supported by a NSF award ATM0404790 to NCAR. R.J.N. acknowledges NSF grant ATM0518555.

[47] Zuyin Pu thanks the reviewers for their assistance in evaluating this paper.

References


Burnside, R. G., and C. A. Tepley (1989), Optical observations of thermo-

Crickmore, R. I., J. R. Dudeney, and A. S. Rodger (1991), Vertical thermo-


J. R. Bowman and M. S. O’Brien, Science Applications International Corporation, 10260 Campus Point Drive, San Diego, CA 92121, USA.

G. Crowley, Atmospheric and Space Technology Research Associates, 12703 Spectrum Drive, Suite 101, San Antonio, TX 78249, USA.

D. P. Drobn, J. T. Emmert, and J. M. Picone, Space Science Division, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC 20375, USA. (douglas.drobn@nrl.navy.mil)

P. Hays, R. J. Niciejewski, and W. Skinner, Space Physics Research Laboratory, Department of Atmospheric, Oceanic, and Space Sciences, College of Engineering, University of Michigan, 2455 Hayward Street, Ann Arbor, MI 48109, USA.

G. Hernandez, Department of Earth and Space Sciences, University of Washington, Johnson Hall 070, Box 351310, 4000 15th Avenue NE, Seattle, WA 98195, USA.

M. J. Jarvis, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

S. Kawamura and Y. Murayama, National Institute of Information and Communications Technology, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan.

M. Larsen and J. W. Meriwether, Department of Physics and Astronomy, Clemson University, 118 Kinard Laboratory, Clemson, SC 29634, USA.

I. M. Reid and R. A. Vincent, School of Chemistry and Physics, University of Adelaide, SA 5005, Australia.

C. Y. She, Physics Department, Colorado State University, 1875 Campus Delivery, Fort Collins, CO 80523, USA.

G. G. Shepherd, Centre for Research in Earth and Space Science, York University, 249 Petrie Science Building, Toronto, ON M3J 1P3, Canada.

D. P. Sipler, Haystack Observatory, Massachusetts Institute of Technology, Off Route 40, Westford, MA 01886, USA.

C. A. Tepley, Arecibo Observatory, Cornell University, HC-03 Box 53995, Arecibo 00612, Puerto Rico.

Q. Wu, High Altitude Observatory, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA.