**Annales
Geophysicae**

A global model of the ionospheric F2 peak height based on EOF analysis

M.-L. Zhang, C. Liu, W. Wan, L. Liu, and B. Ning

Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

Abstract. The ionospheric F2 peak height hmF2 is an important parameter that is much needed in ionospheric research and practical applications. In this paper, an attempt is made to develop a global model of hmF2. The hmF2 data, used to construct the global model, are converted from the monthly median hourly values of the ionospheric propagation factor M(3000)F2 observed by ionosondes/digisondes distributed globally, based on the strong anti-correlation existed between hmF2 and M(3000)F2. The empirical orthogonal function (EOF) analysis method, combined with harmonic function and regression analysis, is used to construct the model. The technique used in the global modelling involves two layers of EOF analysis of the dataset. The first layer EOF analysis is applied to the hmF2 dataset which decomposed the dataset into a series of orthogonal functions (EOF base functions) E_k and their associated EOF coefficients P_k. The base functions E_k represent the intrinsic characteristic variations of the dataset with the modified dip latitude and local time, the coefficients P_k represents the variations of the dataset with the universal time, season as well as solar cycle activity levels. The second layer EOF analysis is applied to the EOF coefficients P_k obtained in the first layer EOF analysis. The coefficients A_k, obtained in the second layer EOF analysis, are then modelled with the harmonic functions representing the seasonal (annual and semi-annual) and solar cycle variations, with their amplitudes changing with the $F_{10.7}$ index, a proxy of the solar activity level. Thus, the constructed global model incorporates the geographical location, diurnal, seasonal as well as solar cycle variations of hmF2 through the combination of EOF analysis and the harmonic function expressions of the associated EOF coefficients. Comparisons between the model results and observational data were consistent, indicating that the modelling

technique used is very promising when used to construct the global model of hmF2 and it has the potential of being used for the global modelling/mapping of other ionospheric parameters. Statistical analysis on model-data comparison showed that our constructed model of hmF2, based on the EOF expansion method, compares better with the observational data than the model currently used in the International Reference Ionosphere (IRI) model.

1 Introduction

Ionospheric modelling has been one of the leading ways to study the ionosphere. An ionospheric model can be either a theoretical model (also named physical model or first principle model) which is derived from various laws of physics and based on the numerical solution of the equations describing the spatial and temporal distribution of medium parameters or an empirical (or semi-empirical) model which is derived from the observational results. Empirical ionospheric models are not only important for ionospheric research, but also very useful and much needed in a wide range of practical applications, such as radio and telecommunication, satellite tracking, Earth observation from space, etc.

In many empirical ionospheric models, such as the International Reference Ionosphere (IRI) (Bilitza, 1990; 2001) and NeQuick (Radicella and Leitinger, 2001; Leitinger et al., 2005) models, the calculation of the electron density profile usually uses the critical points such as F2, F1 and E layer peaks as anchor points, using parameters of foF2, hmF2, foF1, hmF1, foE and hmE (the critical frequencies and peak heights of the F2, F1 and E layers, respectively) as inputs. They are based on the global or regional "maps" of the ionospheric peak parameters. Since the ionospheric electron

Correspondence to: M.-L. Zhang
(zhangml@mail.iggcas.ac.cn)

density has its maximum values in the F2 layer, this layer is the most important region of the ionosphere which is primarily responsible for the reflection of radio waves in high-frequency communication and broadcasting. For this reason, F2 peak is the key anchor point to determine the ionospheric electron density profiles. Therefore, the F2 layer peak height hmF2 is one of the most important parameters in ionospheric empirical modelling. Due to the lack of observational data of hmF2, in practice, people usually obtained hmF2 based on its strong anti-correlations with the ionospheric propagation factor M(3000)F2 whose value is provided usually by the CCIR (International Radio Consultative Committee) or now called ITU (International Telecommunication Union) M(3000)F2 model (CCIR report, 1967). The CCIR M(3000)F2 model predicts M(3000)F2 based on the 12-month running average sunspot number Rz12, then hmF2 is calculated based on this CCIR M(3000)F2 model values using Eqs. (1–6) listed in Sect. 2. However, recently some authors (Adeniyi et al., 2003; Obrou et al., 2003; Zhang et al., 2004; 2007) found that in the equatorial and low-latitude regions, the values of hmF2 converted from the CCIR M(3000)F2 model value using Eqs. (1–6) have remarkable discrepancies with the observational hmF2. They revealed that the discrepancies stemmed from the inaccuracy of CCIR M(3000)F2 model. Because when the measured M(3000)F2 values are used, the hmF2 converted using Eqs. (1–6) agrees very well with the observational result derived from the manually edited traces of ionograms using ionogram inversion programs. Therefore, there is a necessity to update the existing M(3000)F2 model or construct directly a global model of hmF2. This is an urgent main task of the IRI working group community. To this goal, recently, some new modeling techniques have been proposed to model these parameters. For example, Oyeyemi et al. (2007) proposed a new modelling technique based on the application of neural network to model the M(3000)F2 parameter, whereas Liu et al. (2008) attempted to model the M(3000)F2 parameter based on the empirical orthogonal function (EOF) analysis of the observational dataset. Furthermore, recently Gulyaeva et al. (2008) derived a numerical model of hmF2 from the topside database of about 90 000 electron density profile provided by ISIS1, ISIS2, IK19 and Cosmos-1809 satellites for the period of 1969–1987. In this paper, we pursue constructing a global model of hmF2 using a modelling technique which is based on the empirical orthogonal function (EOF) decomposition of the global hmF2 dataset and the modelling of its associated EOF coefficients. In the following sections, we will first mention the dataset used for our model construction in Sect. 2, and then in Sect. 3 we will describe the modelling technique used. The modelling results with some discussions will be shown in Sect. 4 and the last section (Sect. 5) is the summary and conclusion.

2　Data and transformation equations

The fact that hmF2 is a parameter which is not easy to obtain from measurement makes it difficulty to have an observational hmF2 dataset with enough spatial coverage and history length of data accumulation that can be used for global modelling. However, it has been shown (Shimazaki, 1955; Wright and Mcduffie, 1960) that hmF2 is strongly anti-correlated to the ionospheric propagation factor M(3000)F2 and, fortunately, M(3000)F2 can be routinely scaled from ionograms recorded by ionosondes/digisondes distributed globally and its data has already been accumulated for a very long time. This enables us to construct a database of hmF2 from the global M(3000)F2 database for our modelling study. The original empirical formula between hmF2 and M(3000)F2, i.e., hmF2=1490/M(3000)F2-176, was derived by Shimazaki (1955). However, it was found, by later researchers, that a correction term ΔM should be added and the formula now has the format (Wright and Mcduffie, 1960; Bradley and Dudeney,1973; Eyfrig, 1973; Bilitza et al., 1979):

$$hm\text{F2} = \frac{1490}{\text{M(3000)F2} + \Delta M} - 176 \qquad (1)$$

The correction term ΔM in Eq. (1) accounts for the delay-effect caused by the ionizations in the E layer that is related to the foF2/foE ratio. There are various expressions of ΔM derived by different authors (Bradley and Dudeney, 1973; Eyfrig, 1973; Bilitza et al., 1979). The expression of ΔM we used in our calculation of hmF2 is the one derived by Bilitza et al. (1979), which has being used in the IRI model since its first release in 1978 (Rawer et al., 1978):

$$\Delta M = \frac{F_1(R_{12}) \cdot F_2(R_{12}, \Phi)}{fo\text{F2}/fo\text{E} - F_3(R_{12})} + F_4(R_{12}) \qquad (2)$$

where

$$F_1(R_{12}) = 0.00232 \cdot R_{12} + 0.222 \qquad (3)$$

$$F_2(R_{12}, \Phi) = 1 - R_{12}/150 \cdot \exp(-\Phi^2/1600) \qquad (4)$$

$$F_3(R_{12}) = 1.2 - 0.0116 \cdot \exp(R_{12}/41.84) \qquad (5)$$

$$F_4(R_{12}) = 0.096 \cdot (R_{12} - 25)/150 \qquad (6)$$

Where R_{12} is the 12-month running mean of the sunspot number, Φ is the magnetic dip latitude which is related to the magnetic inclination I as $2tg\Phi=tgI$.

The hmF2 data, used in our present study, is calculated from M(3000)F2 based on Eqs. (1–6). The use of Eqs. (1–6) is justified by many works (e.g., Adeniyi et al., 2003; Obrou et al., 2003; Zhang et al., 2004; 2007) that showed when the measured M(3000)F2 values are used as input, the hmF2 value obtained with Eqs. (1–6) agrees very well with the observational ones derived from the manually edited traces of ionograms using ionogram inversion programs. As

for the accuracy of Eq. (1), it is estimated by Dudeney (1983) that the uncertainty of hmF2 calculated from observed M(3000)F2 is within 4–5%, i.e., about 15–20 km. Thus, the relationship between M(3000)F2 and hmF2 can be taken as true. In the present study, data of M(3000)F2, foF2 and foE used to construct the hmF2 database were downloaded from the Space Physics Interactive Data Resource (SPIDR) website http://spidr.ngdc.noaa.gov/. Figure 1 shows the global distribution of the stations used for the present hmF2 global modelling study. In the present work, we used the monthly median data to do the modelling. Therefore, the magnetic activity effects are ignored in the model.

3 Modelling technique description

3.1 Fundamental of EOF analysis method

The modelling technique we used to model hmF2 is mainly based on the empirical orthogonal function (EOF) expansion or decomposition of the dataset. EOF analysis method was actually invented by Pearson (1901). This method has been widely and successfully used by meteorologists and oceanographers to analysis the spatial and temporal variations of physical fields since Lorenz (1956) introduced it into meteorology. In the ionospheric field, it was Dvinskikh (1988) who first introduced the EOF analysis into the empirical modelling of the ionospheric parameters. His pioneer work was followed by other works (e.g., Singer and Taubenheim, 1990; Bossy and Rawer, 1990; Singer and Dvinskikh, 1991; Dvinskikh and Naidenova, 1991). It has been shown by many researchers that the EOF analysis is a powerful method in the ionospheric data analysis and empirical modelling (e.g., Dvinskikh, 1988; Singer and Dvinskikh, 1991; Dvinskikh and Naidenova, 1991; Daniell et al., 1995; Marsh et al., 2004; Matsuo et al., 2002, 2005; Mao et al., 2005, 2008; Materassia and Mitchell, 2005; Zhao et al., 2005; Zapfe et al., 2006, Liu et al., 2008; Wan et al., 2009, and many more).

EOF analysis involves a mathematical procedure that transforms a dataset into a number of uncorrelated components called principal components, with any two components orthogonal to each other. Thus, this method sometimes is also named Principal Component Analysis (PCA) or Natural Orthogonal Component (NOC) algorithm. The underlying physical meaning of EOF analysis is that the variation of a physical field variable is mainly controlled by some independent processes that can be separated. This method decomposes a dataset into a series of eigen function (or base functions) and its associated coefficients, with the base functions orthogonal to each other. As we know, people usually use the Fourier or Spherical Harmonic analysis method to decompose a dataset. However, the base function set used in Fourier or Spherical Harmonic analysis is predesigned artificially. In EOF decomposition, the orthogonal base functions are not
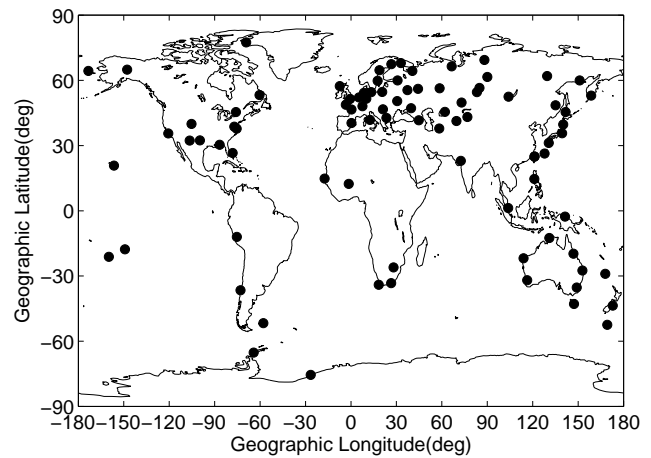

Fig. 1. Global distribution of the stations used for the present modelling study.

artificially designed in advance, but are naturally determined by the experimental dataset to be decomposed. Therefore, they possess the inherent characteristics of the original data, and the eigen series will converge very quickly. This makes it possible to use only a few orders of EOF components to represent most of the variance of the original dataset. Hence, the EOF expansion has advantages in data analysis and representation. In the following, we will briefly describe the fundamentals of the EOF analysis method. For further details, readers are referred to Dvinskikh (1988), Storch and Zwiers (1999) and Xu and Kamide (2004).

Let $Y_{ij}=Y(t_i, x_j)$ represent the value of a variable (e.g., hmF2) at the spatial point x_j(e.g., longitude and latitude) at time t_i, $i=1, 2, \ldots, m$ and $j=1, 2, \ldots, n$. Then $Y_{m \times n}$ is a matrix of m rows and n columns. Suppose there are totally r independent physical processes affecting the variation of the variable, then Y can be decomposed into a series of base function $E^k(x_j), (j = 1, 2, ..., n)$ and its associated coefficients $A^k(t_i), (i = 1, 2, ..., m)$.

$$Y = \sum_{k=1}^{r} Y^k(t,x) = \sum_{k=1}^{r} A^k(t) E^{kT}(x) \tag{7}$$

Where $E^k = [e_1^k, e_2^k, ..., e_n^k]^T$ and $A^k = [a_1^k, a_2^k, ..., a_m^k]^T$. The superscript T over a matrix means the transpose of the matrix. The base functions $E^k (k = 1, 2, ..., r)$ are uncorrelated, that is, they are orthogonal over space:

$$E^{kT} E^l = \sum_{j=1}^{n} e_j^k e_j^l = \begin{cases} 1 \ k = l \\ 0 \ k \neq l \end{cases} \tag{8}$$

e_j^k and a_i^k are obtained by minimizing

$$\delta = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[Y_{ij} - \sum_{k=1}^{r} a_i^k e_j^k \right]^2 \tag{9}$$
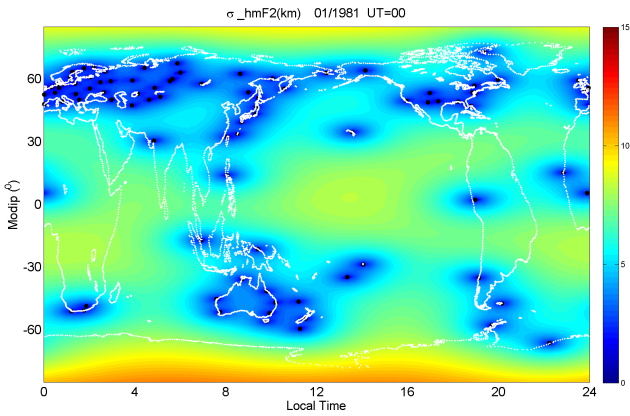
Fig. 2. A sample map of the standard deviation obtained when doing the kriging interpolation.

with respect to e_i^k and e_j^k. In fact, $E^k (k=1,2,...,r)$ is the orthogonal eigenvectors of the symmetric matrix $S = Y^T Y$, which can be obtained by solving the equation

$$SE^k = \lambda_k E^k \tag{10}$$

Where λ_k is the eigen value. Once E^k is found, the associated coefficient A^k can be computed by

$$A^k = Y E^k \tag{11}$$

It can be proved that

$$\delta = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[Y_{ij} - \sum_{k=1}^{r} a_i^k e_j^k \right]^2 = \sum_{i=1}^{m} \lambda_i - \sum_{i=1}^{r} \lambda_i \tag{12}$$

The percentage variance of the dataset captured by the first r components is $\sum_{i=1}^{r} \lambda_i \Big/ \sum_{i=1}^{m} \lambda_i \times 100\%$.

3.2   Data preprocessing and decomposition

Before we can apply the EOF decomposition to the dataset of hmF2, we need to do some data preprocessing work. Since the time periods with available data are different for different stations, we must normalize the data of each station to the same time period. In this study, we choose the years of 1975–1985 covering both ascending and descending phases of the solar cycle activities as the time period used for global modelling. This time period was chosen because data during this period are available for most of the stations used. To fill the missing data for any station during this period, we preprocessed the data as follows. First, for each single station, with its observational values of M(3000)F2, foF2 and foE as input, M(3000)F2 is converted into hmF2 using Eqs. (1–6). When foE is not available, the value obtained from IRI model is used as input. With the converted hmF2 dataset, single station model of hmF2 is constructed for each individual station using the modelling technique described in Liu et al. (2008).

Then, the missing data (if any), for any chosen station during the time period of 1975–1985, are filled with the single station model values.

After normalizing all the chosen stations' data to the time period of 1975–1985 as described above, data at evenly distributed grids ($5° \times 5°$ in a latitudinal range of 85° S–85° N and longitudinal range of 0–360° E) were then obtained by interpolation using Kriging method. To test the validity of the Kriging method in our data preprocessing, the accuracy of Kriging maps is estimated. Figure 2 is a sample but typical map of the standard deviation obtained when doing the Kriging interpolation. In Fig. 2, the black points represent stations whose data is used in interpolation. As can be seen from the figure, the standard deviations of the Kriging maps, at the large areas where observations are sparse, are less than 10 km, which is a very general result in our data preprocessing. This means the accuracy of the Kriging maps obtained in our data preprocessing is quite acceptable. The prepared gridded data are the dataset to which the EOF decomposition is going to be applied.

The first step in our modelling is to do the EOF analysis on the gridded dataset prepared as described above. Two layers EOF decomposition are applied. In the first layer, the prepared gridded dataset are decomposed into the EOF base functions $E_k(\mu, LT)$, which represent the variation of hmF2 with the modified dip latitude (short:Modip) μ and the local time (LT), and the associated EOF coefficients P_k (UT,m), which represent the variations of hmF2 with the universal time (UT) and seasonal as well as solar cycle variations (denoted all by the variable m) as the following

$$hm\text{F2}(\mu, LT, UT, m) = \sum_{k=1}^{N} E_k(\mu, LT) \cdot P_k(UT, m) \tag{13}$$

Here the coordinate (μ, LT) instead of geographical latitude/longitude or magnetic latitude/longitude is used since it was found (Rawer, 1963) that features of the ionospheric parameters, in particular those of the F2 layer, are mostly well organized under the coordinate (μ, LT) which is also well confirmed by recent research works (e.g., Azpilicueta et al., 2006) and by our experience when we tried to decompose the dataset using other coordinates such as the geographical latitude/longitude one. This is due to the fact that the ionosphere is controlled by both the orientation of the Earth's rotation axis and the configuration of the geomagnetic field. Therefore, its variation depends on both the geographical and geomagnetic latitudes, which is embedded in the modip μ defined as $tg\mu = I / \cos^{1/2}\varphi$, where I is the magnetic inclination and φ the geographical latitude. The ionosphere has also longitudinal variation. In our modelling, the longitudinal variation is embedded in the LT (of E_k) and UT (of P_k) variations since $LT=UT$+Longitude/15.

In the second layer of EOF decomposition, the coefficients $P_k(UT,m)$ obtained in the first layer EOF analysis are decomposed once again into a series of base function F_k^j (UT)
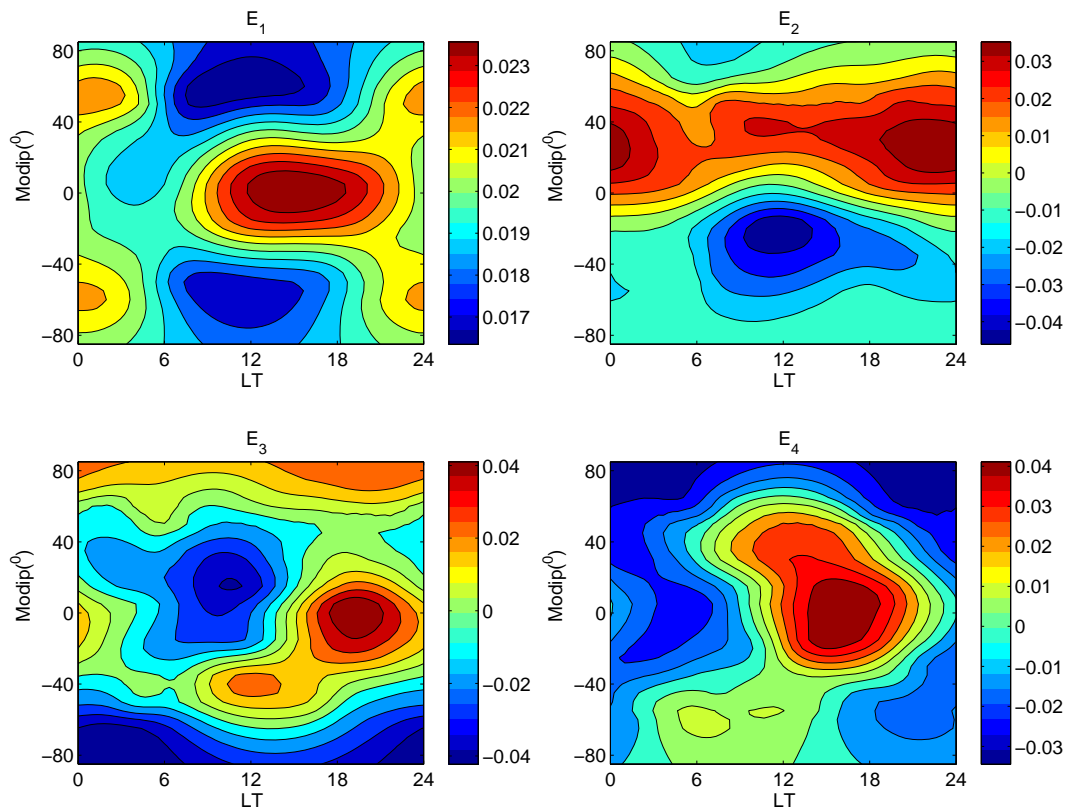
Fig. 3. Distribution of the first four orders of the base functions (E_1–E_4) obtained with the 1st layer EOF decomposition of hmF2.

representing the variation with the universal time (UT) and the associated coefficients $A_k^j(m)$ representing the seasonal as well as solar cycle variations

$$P_k(UT,m) = \sum_{j=1}^{N1} F_k^j(UT) \cdot A_k^j(m) \tag{14}$$

3.3 Modelling the associated EOF coefficients $A_k^j(m)$

The second step in our modelling is to model the EOF coefficients $A_k^j(m)$ with the following harmonic functions representing the seasonal (annual and semi-annual) and solar cycle variations. The reason justifying the usage of the following harmonic functions with their amplitudes changing with the solar flux index $F_{10.7}$ to model the coefficients $A_k^j(m)$ will be given in Sect. 4 when we discuss the results obtained by the EOF decompositions.

$$A_k^j(m) = B_{k1}^j(m) + B_{k2}^j(m) + B_{k3}^j(m) \tag{15}$$

$$B_{k1}^j(m) = c_{k1}^j + d_{k1}^j F_{10.7}(m) \tag{16}$$

$$B_{k2}^j(m) = (c_{k2}^j + d_{k2}^j F_{10.7}(m))\cos\frac{2\pi m}{12} + (s_{k2}^j + t_{k2}^j F_{10.7}(m))\sin\frac{2\pi m}{12} \tag{17}$$

$$B_{k3}^j(m) = (c_{k3}^j + d_{k3}^j F_{10.7}(m))\cos\frac{2\pi m}{6} + (s_{k3}^j + t_{k3}^j F_{10.7}(m))\sin\frac{2\pi m}{6} \tag{18}$$

where m is the month representing the seasonal variation, and $F_{10.7}$ index is a proxy most commonly used to represent the solar activity levels which was originally called the Covington index (Covington, 1948). It is the solar radio flux density measured at a wavelength of 10.7 cm and expressed in units of 10^{-22} Watts/m^2/Hertz. For more details on the proxy solar radio flux, please refer to Tobiska (2001) and references therein. With these equations, the coefficients c_{k1}^j, d_{k1}^j, c_{k2}^j, d_{k2}^j, s_{k2}^j, t_{k2}^j, c_{k3}^j, d_{k3}^j, s_{k3}^j, t_{k3}^j are determined by the least-square fitting approaches.

3.4 Model construction

Our model constructions are done in a reversal procedure based on the obtained EOF base functions E_k and F_k^j as well as the coefficients c_{k1}^j, d_{k1}^j, c_{k2}^j, d_{k2}^j, s_{k2}^j, t_{k2}^j, c_{k3}^j, d_{k3}^j, s_{k3}^j, t_{k3}^j obtained. Specifically, to calculate the model value of hmF2 at a given geographical location, one first uses Eqs. (15–18), with the given month m and the corresponding solar flux index $F_{10.7}(m)$, to calculate the modelled second layer EOF coefficients A_k^j. Then with the calculated modelled coefficients
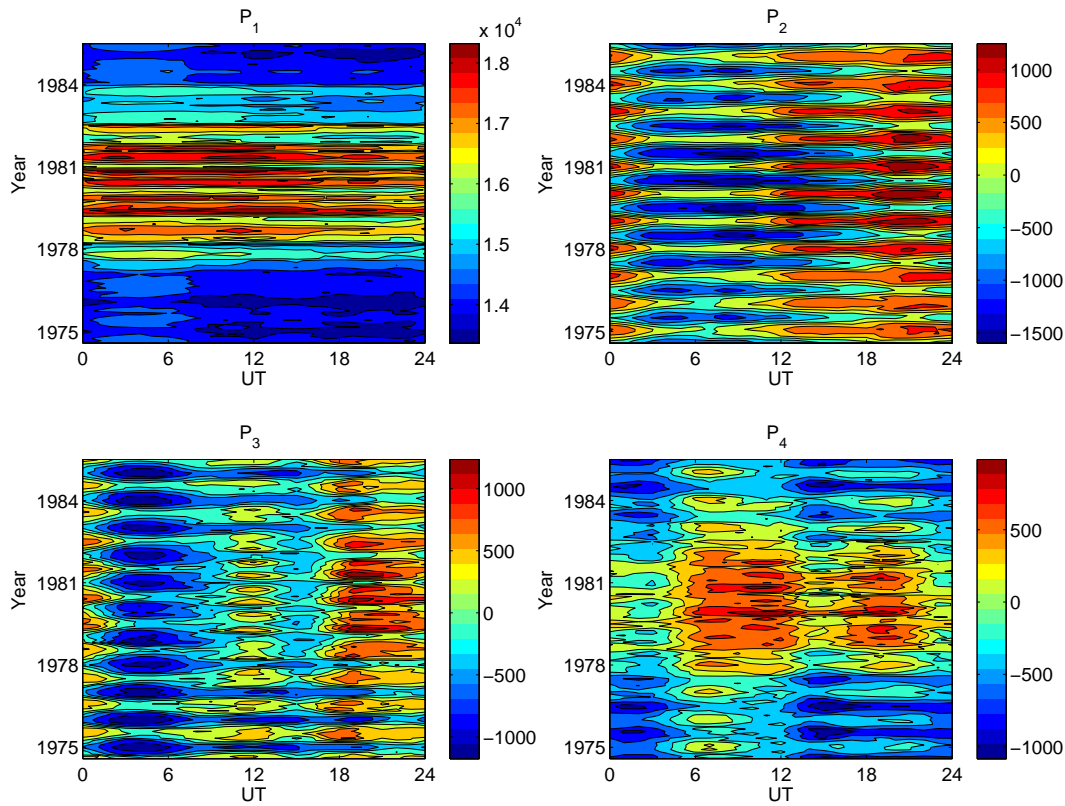
Fig. 4. Distribution of the coefficients (P_1–P_4) corresponding to the first four orders of base functions (E_1–E_4) shown in Fig. 3 obtained with the first layer EOF decomposition of hmF2.

A_k^j and the base functions F_k^j obtained in the second layer EOF decomposition, the modelled first layer EOF coefficients P_k are calculated with Eq. (14). At last, with the calculated modelled P_k and the base functions E_k obtained in the first layer EOF decomposition, the modelled value of hmF2 is calculated using Eq. (13).

4 Results and discussion

Figure 3 shows the contour plots of the first four orders of the base functions E_k, obtained with the first layer EOF decomposition of hmF2 dataset, versus the modified dip latitude (μ) and the local time (LT). It can be seen that apparently, the base functions E_k we obtained showed some typical features that have been observed in many data. For example, the distribution of the first order of base function E_1 manifests mainly a typical phenomenon related to the equatorial ionization anomaly. As is well known, the equatorial ionization anomaly is a phenomenon characterized by a structure with two crests of ionization (best represented by foF2) at about ±17° dip latitude on each side of the magnetic equator and a trough in between. It is formed as a consequence of the so called "fountain" effect. The influence of this equatorial fountain effects on the F2 peak height hmF2 is that it will

produce a latitudinal distribution of hmF2 with higher value near equatorial and low latitudes but with lower value outside. This latitudinal distribution structure of hmF2 is exactly what we see in the distribution of the first order of the base function E_1. The second base function E_2 mainly reflects the north-south asymmetry which is closely related to the seasonal change of the solar zenith angle. Because we will see, in Fig. 4, the associated EOF coefficient P_2 corresponding to this base function shows mainly an annual variation pattern. As for the third base function E_3, the most remarkable feature to notice is the evening enhancement of hmF2 in the equatorial and low-latitude region. This feature is apparently a result of the regeneration/enhancement of the fountain effect during the post-sunset hours due to the evening pre-reversal enhancement (PRE) of the F2 region plasma's electromagnetic $\boldsymbol{E} \times \boldsymbol{B}$ drift caused by the enhanced zonal electric field near the magnetic equator (Fejer et al., 1995, and references therein). Another feature worth noticing is the enhancement/abatement of E_3 in the north/south auroral zones. This feature can be explained by combining it with the pattern of P_3 shown in Fig. 4. P_3 is positive/negative during the north winter/summer seasons. Therefore, the contribution of the component $E_3 \times P_3$ is positive/negative during the north winter/summer seasons. This is apparently a phenomena caused by the difference of sunlit hours in the North/South
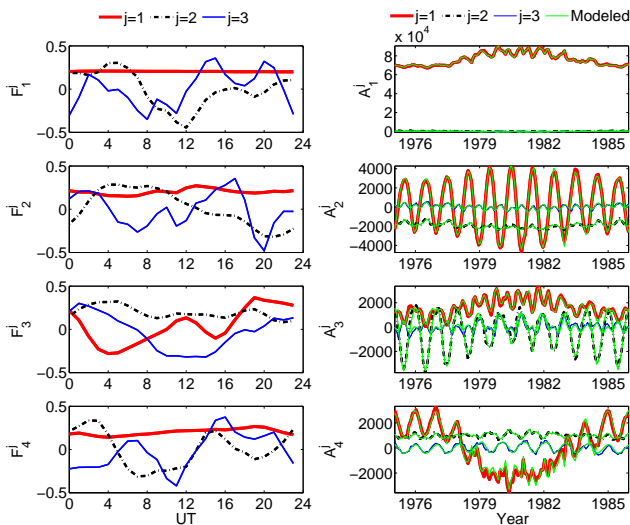
Fig. 5. Distribution of the base functions (F_1^j–F_4^j) (left panels) and their corresponding coefficients (A_1^j–A_4^j) (right panels) obtained with the second layer EOF decomposition of *hm*F2.

Fig. 6. Scatter plots of model value versus observational data of *hm*F2.

Hemispheres during winter/summer seasons. Therefore, as is demonstrated here, the EOF decomposition analysis method, to some extent, is able to separate the variance of a dataset into components caused by sources due to different physical processes or mechanisms. This demonstrated that EOF analysis is a powerful and advantage method to analyze and organize the ionospheric data.

Figure 4 shows the distribution of the first four orders of the associated coefficients P_k obtained in the first layer EOF decomposition of *hm*F2 dataset. They correspond to the first four base functions shown in Fig. 3. It can be seen that P_1 shows a variation pattern with a very strong dependence on the solar cycle activity. It also shows some seasonal variations. The coefficients (P_2–P_4) corresponding to the other orders of base functions show most obviously the annual and semi-annual variations, besides, they also show dependence on the solar activity levels.

Figure 5 shows the results obtained in the second layer EOF analysis, that is, the EOF decomposition of P_k. The left panels are the distributions of the base functions F_k^j obtained, the right panels are for the corresponding associated coefficients A_k^j. The different curves are for the first three components (j=1, 2, 3). For A_k^j, the results modelled with Eqs. (15–18) are also plotted (thin green curves overlaying on each corresponding decomposed ones). As we can see from the right panels, the second layer EOF coefficients A_k^j mainly contains the annual and semiannual variation components and their amplitudes also change with the solar cycle activity levels. This justifies the modelling of the coefficients A_k^j with the harmonic functions representing the seasonal (annual and semi-annual) variations as well as their so-
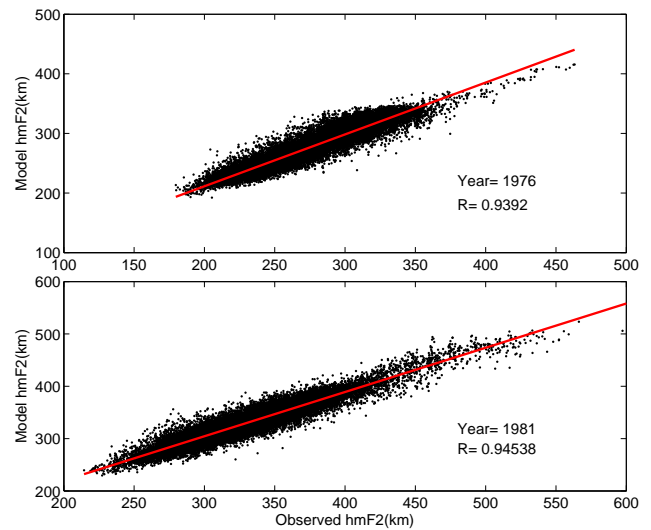
lar cycle activity dependence represented by the $F_{10.7}$ index as expressed by Eqs. (15–18). As can be seen from the figure, the modelled coefficients A_k^j replicated the original ones very well.

As an example for showing the validity of the constructed global model of *hm*F2, Fig. 6 shows the scatter plots of the model value versus observational data of *hm*F2. Please pay attention that the observation presented here is the true observational data from all individual stations available, the kriging and single stations' screen points are not included here. The upper panel is for the case of the low solar activity year (1976) and the lower panel for the case of the high solar activity year (1981). It can be seen that the model values calculated with our constructed model based on the EOF decomposition and the observational data show a high linearity and the correlation coefficients R between the model values and the observational data are very high (R is 0.939 and 0.945 for the years of 1976 and 1981, respectively). The high linearity and correlation coefficients between the modelled *hm*F2 values and the observational data imply that the constructed model is able to reproduce the observational data quite well.

Figure 7a–b shows some sample plots demonstrating the comparison between the observational data of *hm*F2 and the model values given by our EOF based *hm*F2 model as well as those given by the IRI model for the low solar activity year 1965 (Fig. 7a) and the high solar activity year 1970 (Fig. 7b) for 10 stations representative of high, middle and low latitudes from both North and South Hemispheres. Their geographical coordinates are listed in Table 1. In Fig. 7a–b, from top to bottom, stations are ordered from north high, middle and low latitudes to the south low, middle and high latitudes. In these plots, the thick green curves represent our EOF *hm*F2 model results; the thick black curves represent the
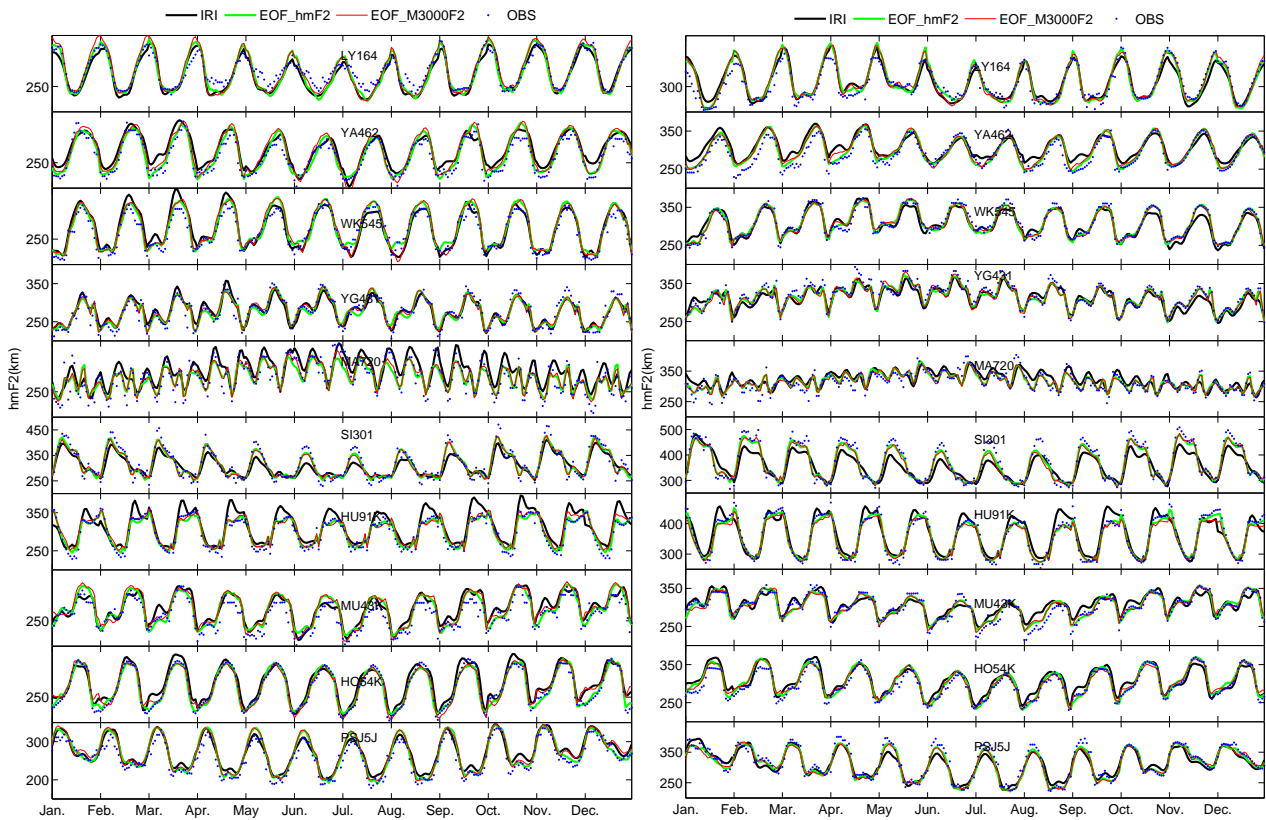
Fig. 7. Sample plots of model values and observational data of *hm*F2 for the years of **(a)** 1965 and **(b)** 1970. From top to bottom panels, stations are ordered from north high, middle and low latitudes to the south low, middle and high latitudes.

Table 1. Geographical coordinates of the stations labeled in Fig. 7a–b.

No	Code	GLAT(° N)	GLON(° E)
1	LY164	64.7	18.8
2	YA462	62.0	129.6
3	WK545	45.4	141.7
4	YG431	31.2	130.6
5	MA720	20.8	203.5
6	SI301	1.3	103.8
7	HU91K	−12.0	284.7
8	MU43K	−32.0	116.4
9	HO54K	−42.9	147.3
10	PSJ5J	−51.7	302.2

IRI model results, whereas the blue points are the observational data. As an additional comparison, results provided by the EOF-based M(3000)F2 model (Liu et al., 2008) are also shown (thin red curves). They are obtained by converting the modelled M(3000)F2 to *hm*F2 using Eqs. (1–6). As can be seen from these plots, the EOF model results (both of *hm*F2

and M(3000)F2) in general reproduced quite well the variation behaviour of the observational data. Compared with the results given by the IRI model, the EOF model results compare obviously better than the IRI model results with the observational data. As for the EOF model results, it can be seen from Fig. 7a–b that, in general, the results based on the M(3000)F2 model are very similar to those provided by the *hm*F2 model. However, a careful inspection of the details of the plots indicates that the *hm*F2 model results are somewhat improved over the M(3000)F2 model results.

To estimate the accuracy of the model, we made a statistical analysis on differences between the model values and the observational data by calculating the root-mean-squared-error (RMSE) of the model:

$$\text{RMSE} = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (hm\text{F2}_{\text{mod } el} - hm\text{F2}_{\text{obs}})^2} \qquad (19)$$

Where N_p is the total number of the data points.

The calculated *RMSEs* for our EOF-based *hm*F2 model are 11.8 km for the low solar activity year (1965) and 13.4 km for the high solar activity year (1970), respectively. Those for the results based on the EOF M(3000)F2 model are 13.3 km

and 14.3 km for the years 1965 and 1970, respectively. As a comparison, the RMSEs for the IRI model results are 18.9 km and 22.6 km for the year 1965 and 1970, respectively. Therefore, in general, our constructed global model of hmF2 based on the EOF decomposition has a higher accuracy than the model currently used in IRI. The accuracy of the EOF-based hmF2 model is also slightly higher than that of the global EOF-based M(3000)F2 model (in terms of the converted hmF2).

5 Summary and conclusion

In the present study, an attempt was made to construct the global model of the ionospheric F2 peak height parameter hmF2 based on the EOF decomposition of the dataset and the modelling of the associated EOF coefficients with harmonic functions representing annual and semi-annual seasonal variations. Solar cycle dependence is also taken into account in the model by including the changes of the harmonic amplitudes with the solar irradiance flux index $F_{10.7}$. Comparisons between the model predictions and the observational data are in agreement. Statistical analysis on the differences between model values and observational data showed the constructed model of hmF2 based on the EOF expansion agrees better with the observational data than the model currently used in IRI. The accuracy of the hmF2 model developed in this study is slightly higher than that of the global EOF-based M(3000)F2 model (in terms of converted hmF2) previously developed by our team (Liu et al., 2008). From the point of view of practical application, the hmF2 model is more preferable than the M(3000)F2 global model, since in many applications it is the parameter hmF2, rather than M(3000)F2, that is really required. It is also due to the fact that the conversion from M(3000)F2 to hmF2 involves the foF2/foE ratio. It would be much more convenient to obtain hmF2 directly from a reliable hmF2 global model than converting to hmF2 from an M(3000)F2 model.

References

Adeniyi, J. O., Bilitza, D., Radicella, S. M., and Willoughby, A. A.: Equatorial F2-peak parameters in the IRI model, Adv. Space.

Res., 31(3), 507–512, doi:10.1016/S0273-1177(03)00039-5, 2003.

Azpilicueta, F., Brunini, C., and Radicella, S. M.: Global ionospheric maps from GPS observations using modip latitude, Adv. Space Res., 38(11), 2324–2331, doi:10.1016/jasr.2005.07.069, 2006

Bilitza, D.: International Reference Ionosphere 1990, Rep. NSSDC/WDC-R&S 90-22, World Data for Rockets and Satell., Nat. Space Sci. Data Cent., Greenbelt, Md., 1990.

Bilitza, D.: International Reference Ionosphere 2000, Radio Sci., 36(2), 261–275, doi:10.1029/2000RS002432, 2001.

Bilitza, D., Sheikh, N. M., and Eyfrig, R.: A global model for the height of the F2-peak using M3000 values from the CCIR numerical map, Telecommun. J., 46, 549–553, 1979.

Bossy, L. and Rawer, K.: Discussion of a new method for mapping ionospheric characteristics, Adv. Space Res., 10(11), 65–74, doi:10.1016/0273-1177(90)90307-L, 1990.

Bradley, P. A. and Dudeney, J. R.: A simple model of the vertical distribution of electron concentration in the ionosphere, J. Atmos. Terr. Phys., 35, 2131–2146, 1973.

CCIR: Comite Consultatif International des Radiocommunications, Reports 340, 340-2 and later supplements, Geneva, 1967.

Covington, A. E.: Solar noise observations on 10.7 centimeters, Proc. IRE, 36, 454–457, 1948.

Daniell, R. E., Brown Jr., L. D., Anderson, D. N., Fox, M. W., Doherty, P. H., Decker, D. T., Sojka, J. J., and Schunk, R. W.: Parameterized ionospheric model: A global ionospheric parameterization based on first principles models, Radio Sci., 30(5), 1499–1510, doi:10.1029/95RS01826, 1995.

Dvinskikh, N. I.: Expansion of ionospheric characteristics fields in empirical orthogonal functions, Adv. Space Res., 8(4), 179–187, doi:10.1016/0273-1177(88)90238-4, 1988.

Dvinskikh, N. I. and Naidenova, N. Ya.: An adaptable regional empirical ionospheric model, Adv. Space Res., 11(10), 7–10, doi:10.1016/0273-1177(91)90312-8, 1991.

Dudeney, J. R.: The accuracy of simple methods for determining the height of the maximum electron contration of the F2-layer from scaled ionospheric characteristics, J. Atmos. Terr. Phys., 45(8–9), 629–640, 1983.

Eyfrig, R.: Eine Bemerkung zur Bradley-Dudeney'schen Modell Ionosphaere, Kleinheubacher Berichte, 17, 199–202, 1973.

Fejer, B. G., de Paula, E. R., Heelis, R. A., and Hanson, W. B.: Global equatorial ionospheric vertical plasma drifts measured by the AE-E satellite, J. Geophys. Res., 100(4), 5769–5776, 1995.

Gulyaeva, T. L., Bardley, P. A., Stanislawskw, I., and Juchnikowski, G.: Towards a new reference model of hmF2 for IRI, Adv. Space. Res., 42, 666–672, 2008.

Leitinger, R., Zhang, M.-L., and Radicella S. M.: An improved bottomside for the ionospheric electron density model NeQuick, Ann. Geophys., 48(3), 525–534, 2005.

Liu, C., Zhang, M.-L., Wan, W., Liu, L., and Ning, B.: Modelling M(3000)F2 based on empirical orthogonal function analysis method, Radio Sci., 43, RS1003, doi:10.1029/2007RS003694, 2008.

Lorenz, E. N.: Empirical orthogonal functions and statistical weather prediction, Technical report, Statistical Forecast Project Report 1, Dept. of Meteor., MIT, 48pp, 1956.

Mao, T., Wan, W., and Liu, L.: An EOF-based empirical model of TEC over Wuhan, Chin. J. Geophys., 48(4), 751–758, 2005.

Mao, T., Wan, W., Yue, X., Sun, L., Zhao, B., and Guo, J.: An empirical orthogonal function model of total electron content over China, Radio Sci., 43, RS2009, doi:10.1029/2007RS003629, 2008.

Marsh, D. R., Solomon, S. C., and Reynolds, A. E.: Empirical model of nitric oxide in the lower thermosphere, J. Geophys. Res., 109, A07301, doi:10.1029/2003JA010199, 2004.

Materassia, M. and Mitchell, C. N.: A simulation study into constructing of the sample space for ionospheric imaging, J. Atmos. Sol. Terr. Phys., 67, 1085–1091, doi:10.1016/j.jastp.2005.02.019, 2005.

Matsuo, T., Richmond, A. D., and Nychka, D. W.: Modes of high-latitude electric field variability derived from DE-2 measurements: Empirical Orthogonal Function (EOF) analysis, Geophys. Res. Lett., 29(7), 1107, doi:10.1029/2001GL014077, 2002.

Matsuo, T., Richmond, A. D., and Lu, G.: Optimal interpolation analysis of high-latitude ionospheric electrodynamics using empirical orthogonal functions: Estimation of dominant modes of variability and temporal scales of large-scale electric fields, J. Geophys. Res., 110, A06301, doi:10.1029/2004JA010531, 2005.

Obrou, O. K., Bilitza, D., Adeniyi, J. O., and Radicella, S. M.: Equatorial F2-layer peak height and correlation with vertical ion drift and M(3000)F2, Adv. Space Res., 31(3), 513–520, doi:10.1016/S0273-1177(03)00024-3, 2003.

Oyeyemi, E. O., Mckinnell, L. A., and Poole, A. W. V.: Neural network-based prediction techniques for global modelling of M(3000)F2 ionospheric parameter, Adv. Space Res., 39(5), 643–650, doi:10.1016/j.asr.2006.09.038, 2007.

Pearson, K.: On lines and planes of closest fit to systems of points in space, Phil. Mag., 6, 559–572, 1901.

Radicella, S. M. and Leitinger, R.: The evolution of the DGR approach to model electron density profiles, Adv. Space. Res., 27(1), 35–40, doi:10.1016/S0273-1177(00)00138-1, 2001.

Rawer, K.: Meteorological and Astronomical Influences on Radio Wave Propagation, ed. B. Landmark, Pergamon Press., Oxford, pp. 221–250, 1963.

Rawer, K., Ramakrishnan, S., and Bilitza, D.: International Reference Ionosphere, International Union of Radio Science, Special Report, Brussels, Belgium, 1978.

Shimazaki, T.: World-wode daily variations in the height of the maximum electron of the ionospheric F2-layer, J. Radio Res. Lab. 2(7), 85–97, 1955.

Singer, W. and Dvinskikh, N. I.: Comparison of empirical models of ionospheric characteristics developed by means of different mapping methods, Adv. Space Res., 11(10), 3–6, doi:10.1016/0273-1177(91)90311-7, 1991.

Singer, W. and Taubenheim, J.: Application of the expansion into empirical orthogonal functions to ionospheric characteristics, Adv. Space Res., 10(11), 59–64, doi:10.1016/0273-1177(90)90306-K, 1990.

Storch, H. V. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, 1999.

Tobiska, W. K.: Validating the solar EUV proxy, $E_{10.7}$, J. Geophys. Res., 106(A12), 29969–29978, 2001.

Wan, W., Ding, F., Zhang, M.-L., Liu, L., and Ning, B.: Modelling the Global Ionospheric TEC with Statistical Eigen Mode Analysis, Adv. Space Res., in review, 2009.

Wright, J. W. and Mcduffie, R. E.: The relation of hmaxF2 to M(3000)F2 and hpF2, J. Radio Res. Lab (Tokyo), 7(32), 498–520, 1960.

Xu, W.-Y. and Kamide, Y.: Decomposition of daily geomagnetic variations by using method of natural orthogonal component, J. Geophys. Res., 109, A05218, doi:10.1029/2003JA010216, 2004.

Zapfe, B. D., Materassi, M., Mitchell, C. N., and Spalla, P.: Imaging of the equatorial ionospheric anomaly over South America – A simulation study of total electron content, J. Atmos. Sol. Terr. Phys., 68, 1819–1833, doi:10.1016/j.jastp.2006.05.025, 2006.

Zhang, M. L., Shi, J. K., Wang, X., Wu, S. Z., and Zhang, S. R.: Comparative study of ionospheric characteristic parameters obtained by DPS-4 digisonde with IRI2000 for low latitude station in China, Adv. Space Res., 33(6), 869–873, doi:10.1016/j.asr.2003.07.013, 2004.

Zhang, M. L., Shi, J. K., Wang, X., Shang, S. P., and Wu, S. Z.: Ionospheric behavior of the F2 peak parameters *fo*F2 and *hm*F2 at Hainan and comparisons with IRI model predictions, Adv. Space Res., 39(5), 661–667, doi:10.1016/j.asr.2006.03.047, 2007.

Zhao, B., Wan, W., Liu, L., Yue, X., and Venkatraman, S.: Statistical characteristics of the total ion density in the topside ionosphere during the period 1996–2004 using empirical orthogonal function (EOF) analysis, Ann. Geophys., 23, 3615–3631, 2005, http://www.ann-geophys.net/23/3615/2005/.