



# Spatial and temporal variation of total electron content as revealed by principal component analysis

Elsayed R. Talaat<sup>1,a</sup> and Xun Zhu<sup>1</sup>

<sup>1</sup>The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA

<sup>a</sup>now at: Heliophysics Division, NASA Headquarters, Washington, D.C. 20546, USA

Correspondence to: Xun Zhu (xun.zhu@jhuapl.edu)

Received: 14 July 2016 – Revised: 30 September 2016 – Accepted: 6 November 2016 – Published: 29 November 2016

**Abstract.** Eleven years of global total electron content (TEC) data derived from the assimilated thermosphere–ionosphere electrodynamics general circulation model are analyzed using empirical orthogonal function (EOF) decomposition and the corresponding principal component analysis (PCA) technique. For the daily averaged TEC field, the first EOF explains more than 89 % and the first four EOFs explain more than 98 % of the total variance of the TEC field, indicating an effective data compression and clear separation of different physical processes. The effectiveness of the PCA technique for TEC is nearly insensitive to the horizontal resolution and the length of the data records. When the PCA is applied to global TEC including local-time variations, the rich spatial and temporal variations of field can be represented by the first three EOFs that explain 88 % of the total variance. The spectral analysis of the time series of the EOF coefficients reveals how different mechanisms such as solar flux variation, change in the orbital declination, nonlinear mode coupling and geomagnetic activity are separated and expressed in different EOFs. This work demonstrates the usefulness of using the PCA technique to assimilate and monitor the global TEC field.

**Keywords.** Ionosphere (ionospheric disturbances)

## 1 Introduction

The ionosphere is highly variable and has a complex system of drivers including variable solar radiation plus geomagnetic activity from the upper atmosphere and momentum and energy fluxes associated with neutral wind dynamics from the lower atmosphere. While magnetospheric forcing dominates the variability at high latitudes in the ionosphere, photochem-

istry and neutral dynamics play dominant roles in the ionospheric structure and variability at mid- and low latitudes. One critical quantity describing the ionosphere and its variability is the total electron content (TEC). A variety of in situ and remote-sensing techniques has been employed to study the Earth's ionosphere in terms of the electron density. The method and platform used for measurement determine resolution in time and space, with the measurement often being distributed unevenly. Unless assimilated general circulation models are used, no one method effectively allows for the sampling of large areas of the ionosphere with high and uniform resolution in both time and space. Since ionospheric electronic densities respond to a complex set of highly variable driving mechanisms, the global characterization of the response to solar variability posts a significant challenge.

The X-ray and ultraviolet solar irradiance which creates the ionosphere varies on all timescales: with an 11-year solar cycle associated with the 22-year magnetic cycle of the solar dynamo, with a quasi-27-day period due to active regions rotating with the sun, and on the order of minutes to hours as eruptions occur on the disk of the sun (Lean, 1987; Tobiska, 1993). The ionosphere varies on all these timescales in response to the solar inputs, while the geographic relationship of the Earth's orbit, rotation and seasonal tilt creates the solar zenith angle dependence that yields the observed diurnal, seasonal and annual variations in ionospheric density. This is further complicated by the tilt of the magnetic field of Earth, the magnetospheric inputs that drive the ionosphere at high latitudes, and the neutral dynamics generated by solar and magnetospheric forcing, indicating significant longitudinal and hemispherical asymmetries in space. Thus, the multi-scale variability and complexity in both time and

space induced by different kinds of physical processes are the main characteristics of the ionosphere.

In this paper, we perform a set of multi-year assimilated runs of the thermosphere–ionosphere electrodynamics general circulation model (TIEGCM) (Roble et al., 1988; Richmond et al., 1992) driven by a lower boundary condition of tidal forcing derived from the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument (Talaat and Lieberman, 1999; Talaat et al., 2001). To effectively decompose and understand the complicated temporal and spatial variations of the ionosphere, we apply the empirical orthogonal function (EOF) and the associated principal component analysis (PCA) to the TEC fields derived from the TIEGCM runs. Section 2 describes the EOF and PCA notations and their application to different TEC fields. Section 3 shows how the EOF and PCA are able to reveal major physical processes associated with different EOFs derived from data fields. Section 4 provides a final summary and conclusions.

## 2 EOF and PCA of the assimilated TEC

### 2.1 A brief review and notations of the EOF and PCA

The empirical orthogonal function (EOF) and the corresponding principal component analysis (PCA) are the standard statistical techniques for analyzing atmospheric data since being introduced by E. N. Lorenz in 1956 (Wilks, 2006). The PCA technique decomposes a given spatial–temporal field such as TEC into a set of base functions called EOFs that is internally determined from the data sets. Let  $\mathbf{u}_n = u_{n,ij} = u(t_n; x_i, y_j)$  denote the time series of the measured or the assimilated TEC field at  $N$  time steps over  $K = I \times J$  spatial grids ( $n = 1, 2, \dots, N$ ;  $i = 1, 2, \dots, I$ ;  $j = 1, 2, \dots, J$ ), where  $t_n$ ,  $x_i$  and  $y_j$  denote time, longitude and latitude, respectively. Then, the PCA expresses the spatial–temporal field  $\mathbf{u}_n$  by its time-averaged field  $\bar{\mathbf{u}} (= N^{-1} \sum_{n=1}^N \mathbf{u}_n)$  and a set of EOFs ( $\mathbf{v}_m$ ,  $m = 1, 2, \dots, K$ ) that do not vary with time (Preisendorfer, 1988, Wilks, 2006):

$$\mathbf{u}_n = \bar{\mathbf{u}} + \mathbf{u}'_n = \bar{\mathbf{u}} + \sum_{m=1}^K E_{m,n} \mathbf{v}_m, n = 1, 2, \dots, N, \quad (1)$$

where the principal component (PC)  $E_{m,n}$  is the time series of the decomposition coefficient that characterizes the temporal variation of the original field  $\mathbf{u}_n$  being projected onto the  $m$ th EOF  $\mathbf{v}_m$ . Furthermore, different PCs are orthogonal to each other in time (Preisendorfer, 1988).

Equation (1) is a reconstruction formula and can be considered as a linear transform that projects the original perturbation field  $\mathbf{u}_n - \bar{\mathbf{u}}$  onto a set of base functions  $\{\mathbf{v}_m\}$ . Since EOFs are derived from the data rather than prescribed base functions such as spherical harmonic functions that are often used for analyzing global data, the most information on the

original field that represents the physics and its variations can be expressed by a highly truncated Eq. (1) with  $M \ll K$ ,

$$\mathbf{u}_n \approx \bar{\mathbf{u}} + \sum_{m=1}^M E_{m,n} \mathbf{v}_m, n = 1, 2, \dots, N. \quad (2)$$

The effectiveness of the PCA can be quantitatively measured by the ratio of the percentage of the total variance in  $\mathbf{u}_n - \bar{\mathbf{u}}$  explained by individual EOFs:

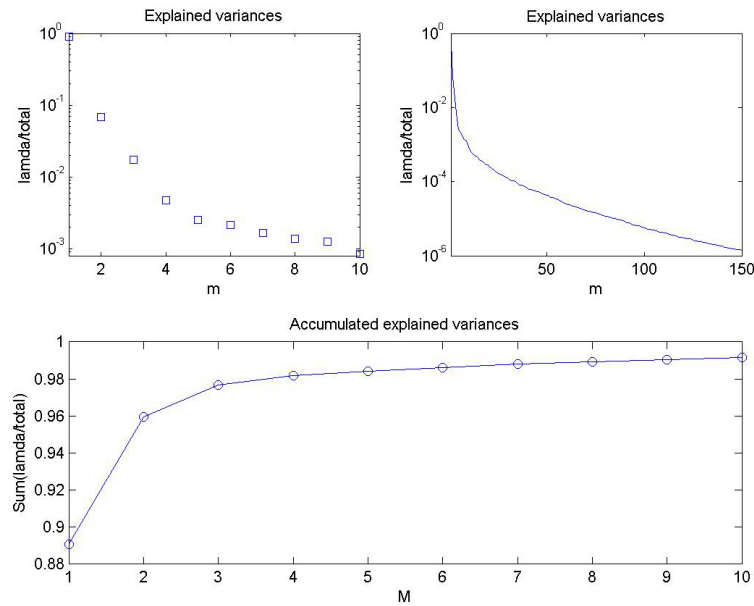
$$r_m^2 = \frac{\text{Var}(E_{m,n} \mathbf{v}_m)}{\text{Var}(\mathbf{u}_n - \bar{\mathbf{u}})} \times 100\%. \quad (3)$$

The value of  $r_m^2$  is proportional to the magnitude of the  $m$ th eigenvalue ( $\lambda_m$ ) of the covariance matrix of  $\mathbf{u}_n - \bar{\mathbf{u}}$  that decreases with  $m$  (Wilks, 2006). Since different EOFs are orthogonal to each other in space, the explained variance by the truncated PCA expression Eq. (2) for different  $M$  is given by

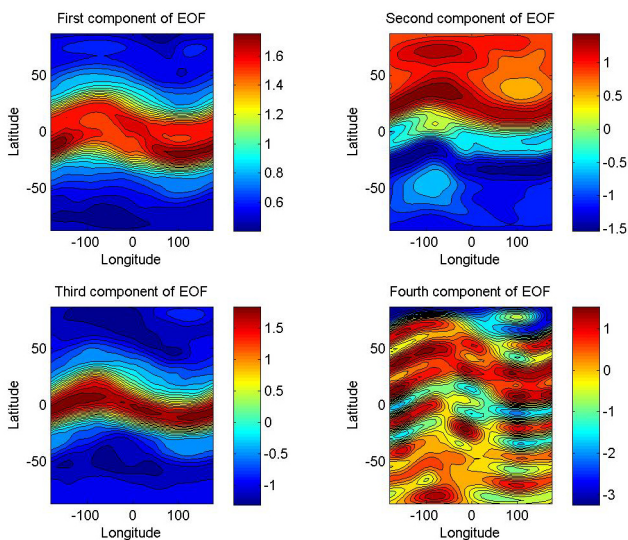
$$R_M^2 = \sum_{m=1}^M r_m^2. \quad (4)$$

### 2.2 EOF and PCA applied to daily averaged TEC

We ran the TIEGCM with realistic solar inputs and tidal forcing over the 11-year time period from 1990 to 2010. The horizontal resolution of the TEC field is  $5^\circ \times 2.5^\circ$ , which corresponds to  $K = 72 \times 71 = 5112$  spatial grids. In Fig. 1, we show the plots of  $r_m^2$  vs.  $m$  and  $R_M^2$  vs.  $M$  for the daily averaged global TEC over an 11-year period with 4180 days (1999–2010). The most striking feature is that the first EOF has already explained 89% of the total variance. The combination of the first four EOFs explains more than 98% of the total variance. The fact that  $R_M^2 > 98\%$  for  $M = 4$  shows the effectiveness of the PCA on data compression or information representation in this particular example. Here, the effectiveness of the data compression of PCA is mainly due to the following two reasons: (i) the base functions have been internally derived from data rather than externally prescribed and (ii) there exist substantial correlations among data on different spatial and temporal grids. In Fig. 2, we show the first four EOFs for the 11-year daily averaged global TEC as functions of longitude and geographic latitude. Note that the first component explains the majority (with  $r_1^2 = 89.1\%$ ) of the spatial variance of TEC. It represents the combined effects of solar radiation that makes the TEC decrease with latitude and geomagnetic field that produces a zonal wave number 1 structure due the offsetting of the geomagnetic poles. Note that the electrons produced by the photoionization are symmetric with respect to geographic poles, whereas those produced by the energetic particle precipitation are symmetric with respect to the geomagnetic poles. The second EOF (with  $r_2^2 = 6.9\%$ ) captures the interhemispheric asymmetry of the TEC, mainly caused by the seasonal variation of the inclination angle of the Earth's orbit. The third EOF (with  $r_3^2 = 1.7\%$ ) represents the narrower or finer structure of TEC



**Figure 1.** Explained variance  $r_m^2$  vs.  $m$  and the accumulated explained variance  $R_M^2$  vs.  $M$  for the daily averaged global TEC over an 11-year time period (4180 days) from 1999 to 2010 and  $72 \times 71$  spatial grids.



**Figure 2.** First four EOFs for the 11-year daily averaged global TEC as functions of longitude and geographic latitude calculated by the spatial and temporal settings given in Fig. 1.

in the latitudinal direction that is mainly also caused by the solar radiation and cannot be fully represented by the first EOF. The fourth EOF (with  $r_4^2 = 0.5\%$ ) shows much smaller wave structure in both longitudinal and latitudinal directions.

Since the EOFs are determined internally from the data, it is worthwhile examining in this application how sensitive the derived  $R_M^2$  and EOFs are with respect to the data we used. Figures 3 and 4 show the same calculations as in Figs. 1 and 2

for a 6-year period with 2190 days from 1999 to 2004. Comparison between the two sets of figures shows that the change in the length of the spatial–temporal field only slightly modifies PCA output. Though the record of the data has been cut shorter by about one half, the first EOF still explains 79.4% and the combination of the first four EOFs explains 96.4% of the total TEC variance. More importantly, the spatial patterns of the first four EOFs, shown in Fig. 4, are also very similar to those shown in Fig. 2.

We have already indicated that it is the high correlation of the spatial–temporal variations of the field that leads to a few EOFs being able to explain the majority of the variance. Such a high correlation in data is also reflected by the fact that the first few EOFs capture relatively smooth or large-scale structures of the field as shown in Figs. 2 and 4 for the first three EOFs. Since the large-scale structures can also be effectively represented by a lower-resolution field, one would expect a similar PCA result for the same field with a lower spatial resolution. In Figs. 5 and 6, we show the same PCA as in Figs. 1 and 2 except for a lower horizontal resolution of  $10^\circ \times 5^\circ$  that corresponds to  $K = 36 \times 35 = 1260$  spatial grids. This leads to a reduction of about a factor 4 in spatial grids for the analysis. Now, the variances explained by the first four EOFs are 89.4, 6.8, 1.7 and 0.4%, respectively, which amounts to 98.3% of the total variance when four EOFs are combined (Fig. 5). This is almost identical to the standard-resolution PCA shown in Fig. 1. The spatial patterns of the first three EOFs shown in Fig. 6 are almost identical to those shown in Fig. 2. There exists some minor difference in the fourth EOF because some structural fea-

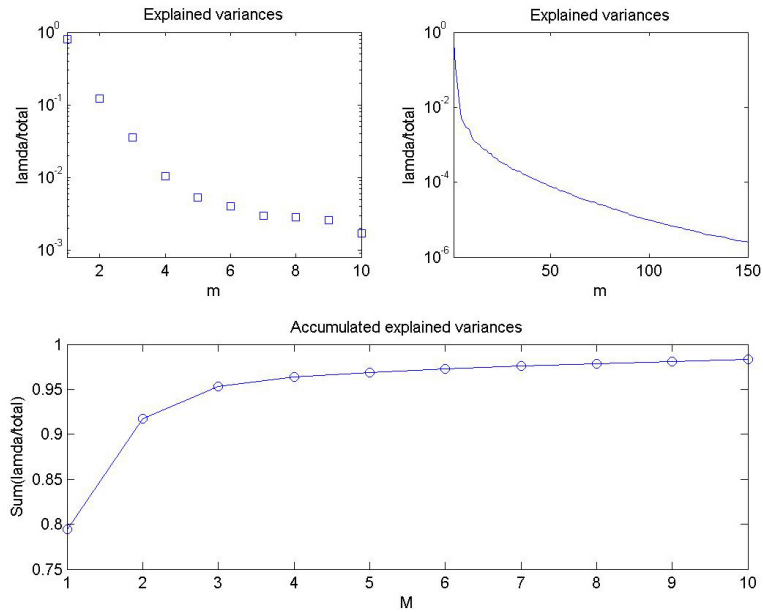


Figure 3. Same as Fig. 1 except for a 6-year time period (2190 days) from 1999 to 2004.

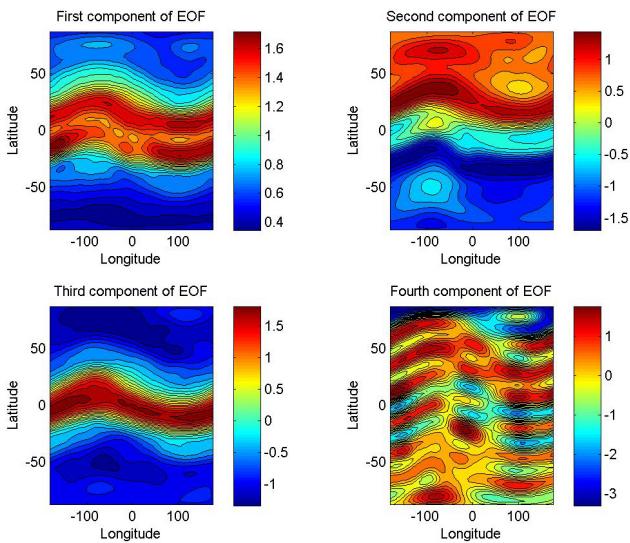


Figure 4. Same as Fig. 2 except for a 6-year time period (2190 days) from 1999 to 2004.

tures cannot be fully resolved by a lower-resolution analysis ( $r_4^2 = 0.4\%$  in Fig. 1 vs.  $r_4^2 = 0.5\%$  in Fig. 5).

2.3 EOF and PCA applied to hourly TEC

Next, we examine the TEC field that includes the local-time variations. The PCA is applied to the  $u(t_n; x_i, y_j)$  field with the lower-resolution spatial grid of Figs. 5 and 6 but with a temporal resolution of 2 h. The total number of time steps  $N = 50\ 160$ , which corresponds to 11.4 years of time interval starting from 1999. Figures 7 and 8 show plots of  $(r_m^2, R_M^2)$

and the first six EOFs. We first note that (Fig. 7) the contribution by the first EOF (48.8 %) is significantly lower than one shown in the daily averaged TEC analysis (Figs. 1, 3 and 5), indicating a weaker correlation or a richer structure in spatial and temporal variability of the TEC field. This is expected since unlike the daily averaged field, the local-time variation introduces additional and stronger longitudinal variations due to the local-time response of TEC that is directly related to the solar radiation forcing. The sum of the first six EOFs makes up approximately 93.6 % of the total variance. Another significant feature shown in Fig. 7 is that the second and third EOFs contribute almost equally (20.4 and 18.6 %) to the total variance. The spatial pattern of the first EOF in Fig. 8 is almost identical to that shown in Fig. 2, mainly representing the effects of daily averaged solar forcing and the dominant wave number 1 feature due to the off-setting geomagnetic field. The next two EOFs shown in Fig. 8 represent longitudinal variation of wave number 1 with approximately equal magnitudes and  $\pi/2$  phase shift. These two natural modes represent the major contributions from the first two components of the sinusoidal decomposition in longitude  $\sin(\pi x_i/180)$  and  $\cos(\pi x_i/180)$  caused by the local-time solar forcing, which are not present in the daily averaged TEC field. The fourth and fifth EOFs in Fig. 8 once again approximately duplicate the second and third EOFs in Fig. 2. In other words, because different EOFs are orthogonal, the increase in resolution in TEC data only adds new modes to PCA but has little effect on the existing modes in the lower-resolution PCA. The sixth EOF mainly represents the longitudinal variation of wave number 2. Note that the paired  $(r_6^2, r_7^2)$  and  $(r_8^2, r_9^2)$  shown in Fig. 7 are similar to  $(r_2^2, r_3^2)$  and have nearly equal magnitudes of the explained variance.



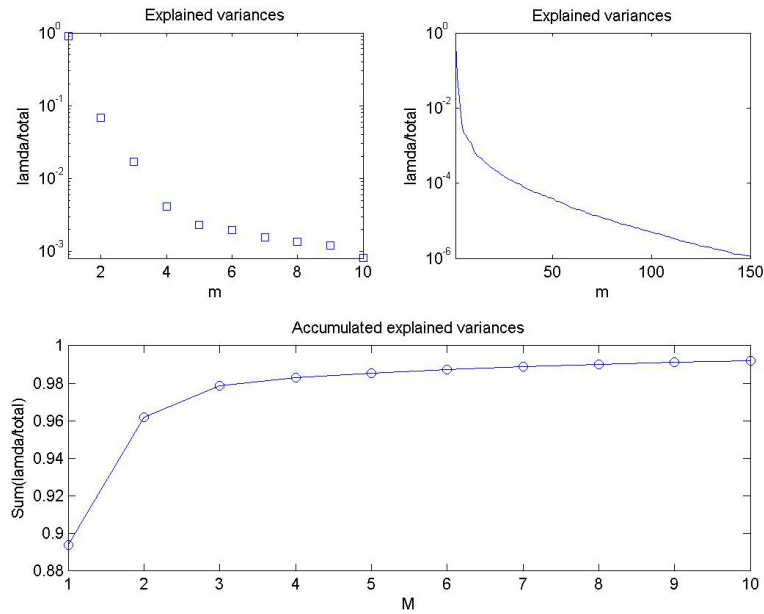


Figure 5. Same as Fig. 1 except for a spatial resolution of  $36 \times 35$  grids.

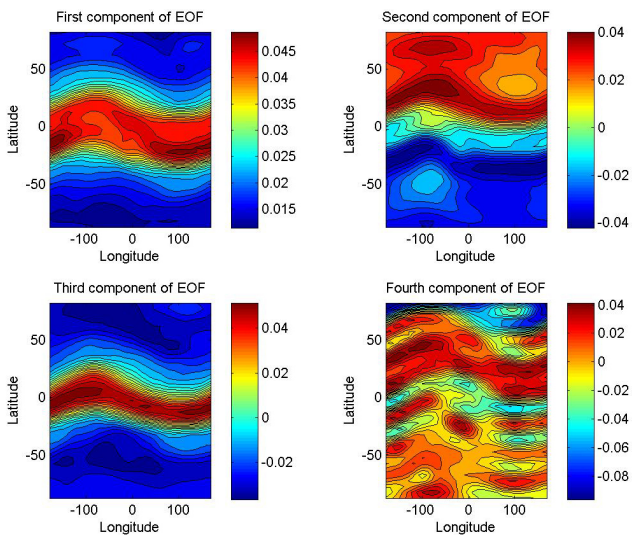


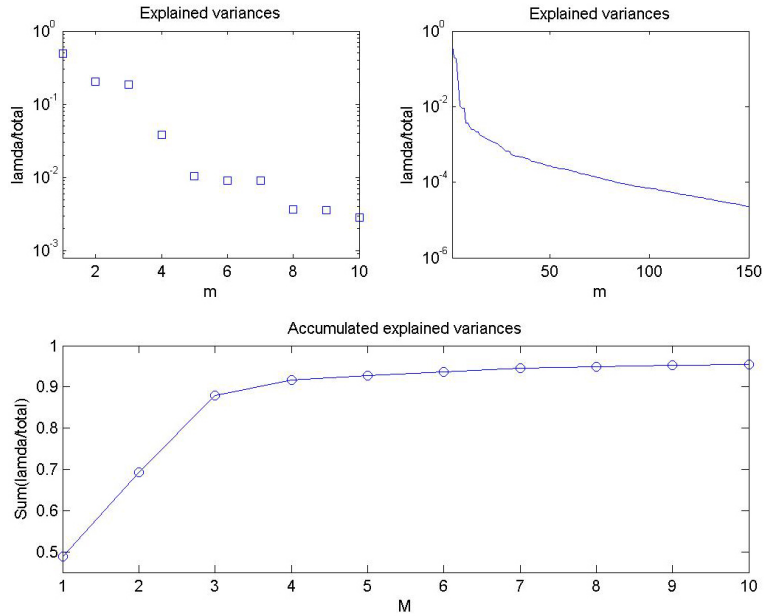
Figure 6. Same as Fig. 2 except for a spatial resolution of  $36 \times 35$  grids.

One expects that the paired EOFs ( $v_6, v_7$ ) and ( $v_8, v_9$ ) represent the high harmonics of the longitudinal variance associated with the local-time variation of the solar forcing. We have also performed sensitivity analyses for the case similar to Figs. 3 and 4 by applying PCA to shorter data records of 4 years that correspond to either solar maximum (1999–2003) or solar minimum (2006–2009). In terms of their spatial structure, the derived EOFs are almost identical to those shown in Fig. 8.

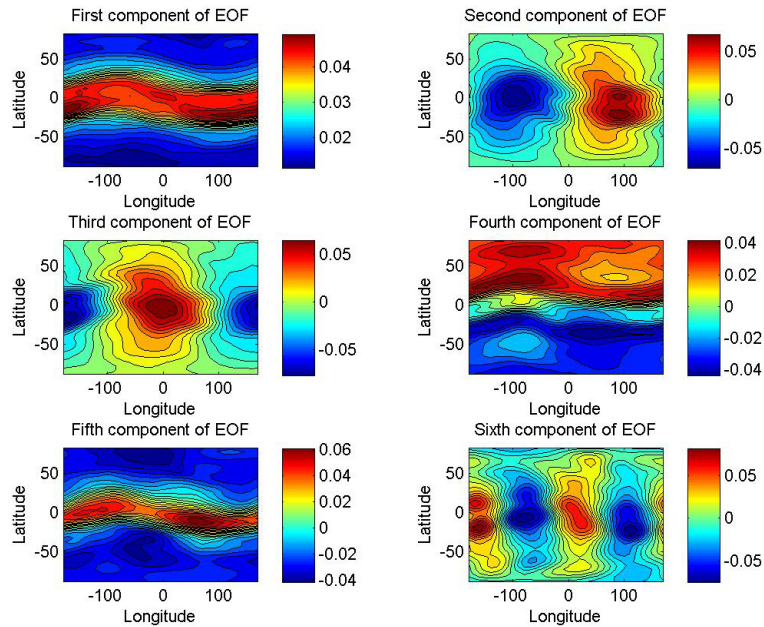
### 3 Physical processes as objectively revealed by EOFs and TEC

One major advantage of the PCA technique is the data compression, so the physical field is effectively projected onto a few modes that include the majority of the variance of the original field. We have already shown that the four EOFs in Fig. 2 and the six EOFs in Fig. 8 contain 98.2 and 93.6 % of the total variance, respectively. Based on their spatial patterns, we have also indicated in the above analysis that there exist clear physical mechanisms that drive the different EOFs solely derived from the data. To illustrate the roles played by different physical mechanisms in extracting EOFs, we show in Fig. 9 the time series of the decomposition coefficients  $E_{m,n}$ , i.e., PCs, of the PCA for the first four EOFs, shown in Fig. 2. The specific physical mechanisms are best described in by  $E_{1,n}$  and  $E_{2,n}$ . Since  $v_1$  and  $v_2$  in Fig. 2 mainly represent the effects of the solar radiation and the seasonal variation of the inclination angle of the Earth’s orbit, the two time series plotted in the upper panel of Fig. 9 show the responses of the 11-year solar cycle and annual oscillation. Note that the variance contributed from  $v_3$  and  $v_4$  is far less than those from  $v_1$  and  $v_2$  (Fig. 1). This is also reflected by the fact that the magnitudes of  $E_{3,n}$  and  $E_{4,n}$  are much smaller than those of  $E_{1,n}$  and  $E_{2,n}$ . Furthermore, since neither effect of the solar radiation nor the seasonal variation is orthogonal in space, the remaining EOFs, including  $v_3$  and  $v_4$  together with their coefficients  $E_{m,n}$ , will include both effects, as indicated above and also shown in the lower panel of Fig. 9.

To further demonstrate the benefit of the PCA technique in the current application, we show in Fig. 10 the time series of the longitudinally averaged TEC in the unit of



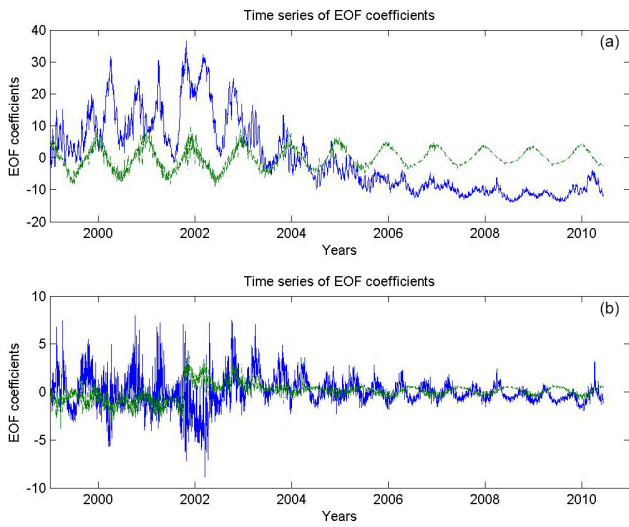
**Figure 7.** Explained variance  $\lambda_m^2$  vs.  $m$  and the accumulated explained variance  $R_M^2$  vs.  $M$  for the global TEC over an 11-year time period (4180 days) from 1999 to 2010 with a 2 h temporal resolution and  $36 \times 35$  spatial grids.



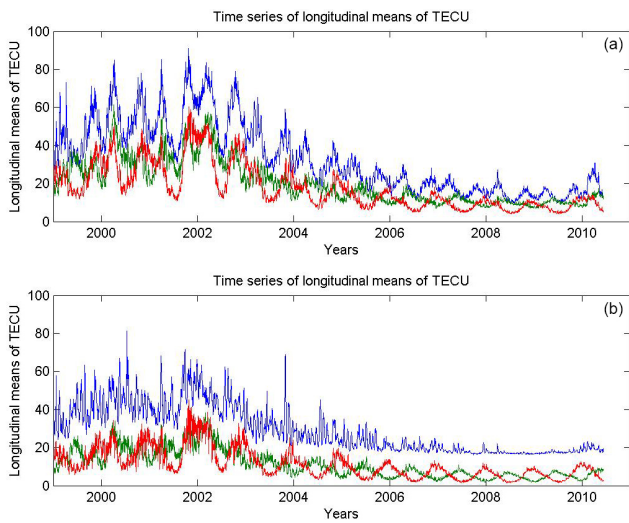
**Figure 8.** First six EOFs for the 11-year global TEC as functions of longitude and geographic latitude calculated by the spatial and temporal settings given in Fig. 7.

TECU ( $1 \text{ TECU} = 10^{16} \text{ electrons m}^{-2}$ ) at five different latitudes: Equator ( $0^\circ$ ),  $37.5^\circ \text{ S}$ ,  $37.5^\circ \text{ N}$ ,  $62.5^\circ \text{ S}$  and  $62.5^\circ \text{ N}$ . Also shown in the figure is the daily  $F_{10.7}$  that is the 10.7 cm solar radio flux (in units of  $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ ) and that measures the intensity of the solar activity. The overall correlation between  $F_{10.7}$  and TEC is overwhelming, suggesting that the solar radiation and/or solar wind are the ultimate

source for producing electrons in the Earth’s ionosphere. In comparison with Fig. 9, we note that though the plots show a certain latitudinal structure, namely that low-latitude TECs are generally higher than the high-latitude ones, all five time series have a similar magnitude and there is no clear separation between the effect of solar radiation and seasonal variations. Furthermore, there exist significant anticorrelations be-

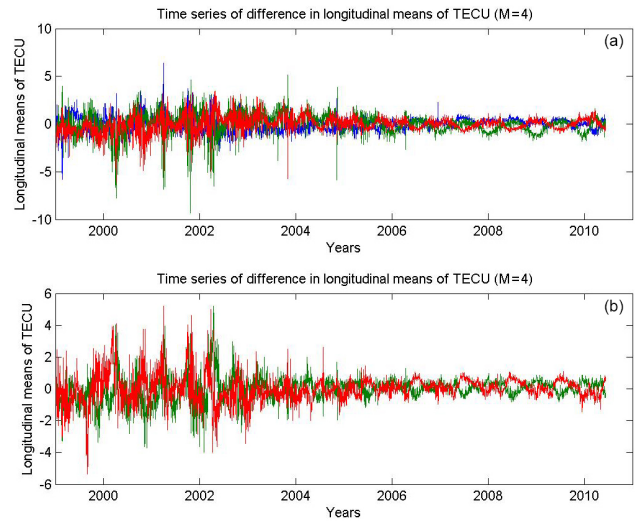


**Figure 9.** Time series of the principal components of the PCA for the first four EOFs shown in Fig. 2:  $E_{1,n}$ ,  $E_{2,n}$  – blue, green lines in panel (a);  $E_{3,n}$ ,  $E_{4,n}$  – blue, green lines in panel (b).

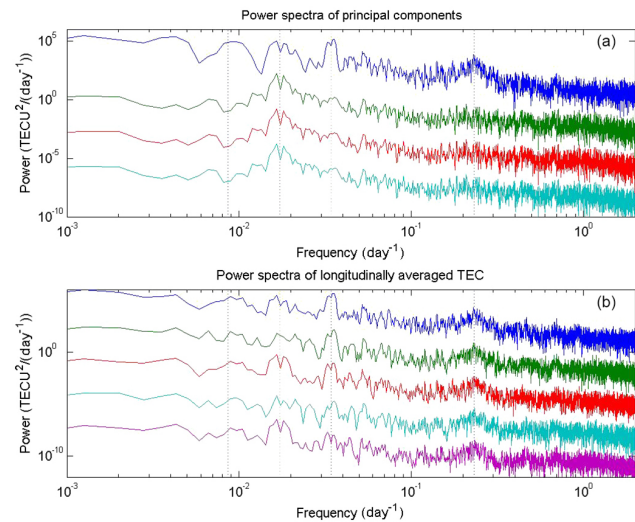


**Figure 10.** Daily mean time series of the longitudinally averaged TEC at five different latitudes:  $0^\circ$ ,  $37.5^\circ$  S,  $37.5^\circ$  N – blue, green, red lines in panel (a);  $62.5^\circ$  S,  $62.5^\circ$  N – green, red lines in panel (b). The blue line in panel (b) denotes  $F_{10.7}$ , which is the 10.7 cm solar radio flux (in units of  $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ ).

tween the southern latitude and northern latitude time series. In other words, the five time series of TEC shown in Fig. 10 are highly correlated so that the information in Fig. 10 contains a significant redundancy. Note that the time series of  $E_{m,n}$  coupled with the basis function  $\{v_m\}$  in the highly truncated expansion Eq. (2) represent the entire spatial–temporal field. In Fig. 11, we show the differences of the TEC time series shown in Fig. 10 and the ones calculated based on Eq. (2) with  $M = 4$ . Comparison between Figs. 10 and 11 shows that differences are small and close to irregular, suggesting that



**Figure 11.** Differences in TEC time series between those shown in Fig. 10 and the ones calculated based on highly truncated expansion Eq. (2) with  $M = 4$  at five different latitudes:  $0^\circ$ ,  $37.5^\circ$  S,  $37.5^\circ$  N – blue, green, red lines in panel (a);  $62.5^\circ$  S,  $62.5^\circ$  N – green, red lines in panel (b).

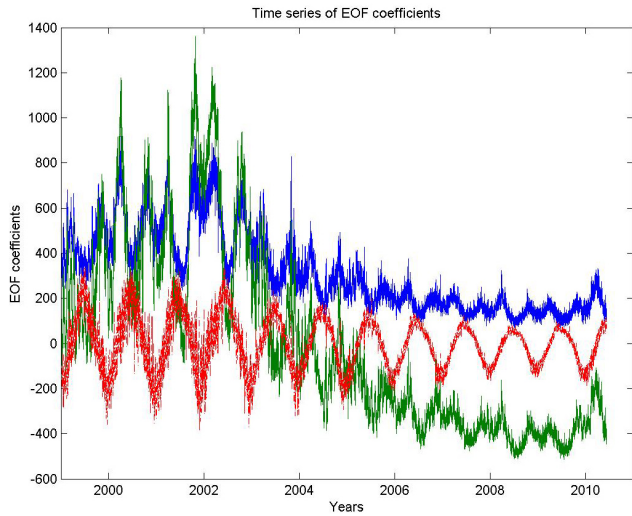


**Figure 12.** Fourier power spectra of the first four PCs shown in Fig. 9 (panel a) and the time series at five different latitudes shown in Fig. 10 (panel b). To avoid clustering, all the curves except the first PC in the upper panel and the one at the Equator in panel (b) have been consecutively shifted downward by 3 orders of magnitude.

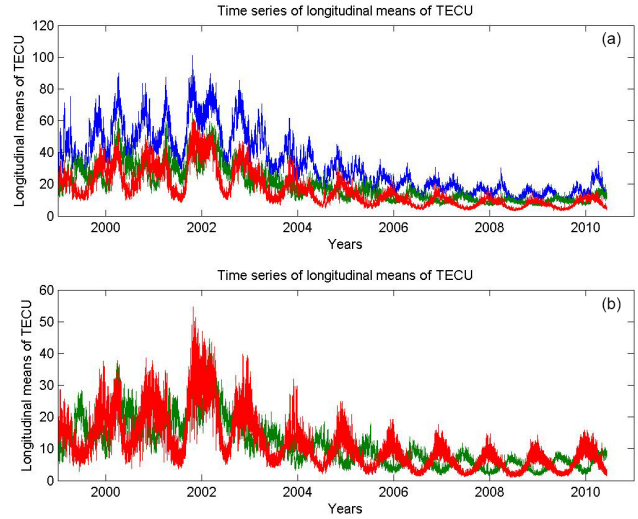
the first four EOFs together with the four PCs have included major physical processes that control the TEC distribution and variations.

In Fig. 12, we show the Fourier power spectra of the time series shown in Figs. 9 and 10. There are four vertical dashed lines denoting frequencies corresponding to 2-year, annual, semiannual and 27-day periods of oscillations. Now, it is

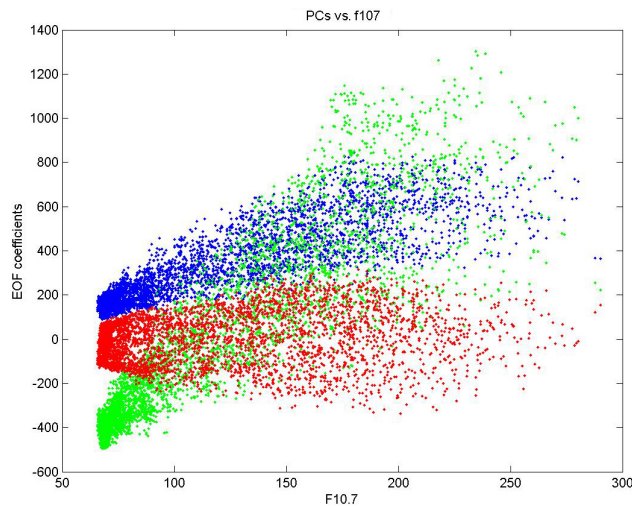




**Figure 13.** Time series of the principal components of the PCA for the first EOFs shown in Fig. 8:  $E_{amp}$ ,  $E_{1,n}$ ,  $E_{4,n}$  – blue, green, red lines.



**Figure 15.** Two-hour resolution time series of the longitudinally averaged TEC at five different latitudes:  $0^\circ$ ,  $40^\circ$  S,  $40^\circ$  N – blue, green, red lines in panel (a);  $60^\circ$  S,  $60^\circ$  N – green, red lines in panel (b).



**Figure 14.** Scatterplots between the daily averaged  $F_{10.7}$  and the daily averaged time series ( $E_{amp}$ ,  $E_{1,n}$ ,  $E_{4,n}$ ) shown in Fig. 13: ( $E_{amp}$ ,  $E_{1,n}$ ,  $E_{4,n}$ ) – (blue, green, red) dots.

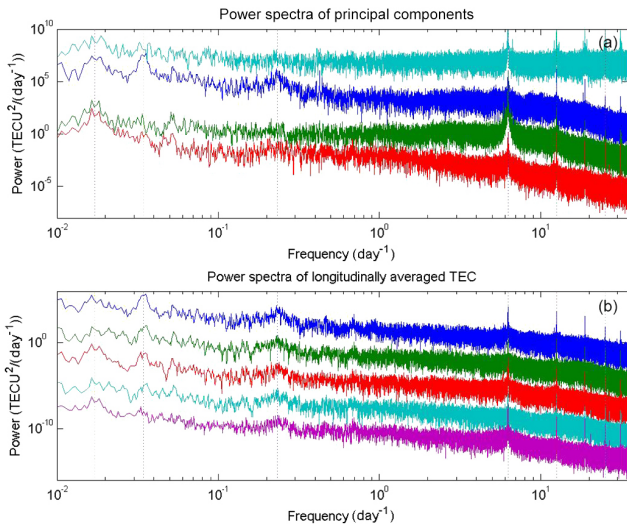
more clearly and quantitatively shown in the figure that the major spectral features among different EOFs are different, whereas those among the longitudinally averaged TEC at different latitudes are nearly the same. The first EOF spectrum peaks at all four frequencies, whereas the second EOF only shows one striking peak in the annual period that is dominant over the entire spectral domain. For the third and fourth EOFs, only two peaks at semiannual and 27-day periods are noticeable. Since different spectral characters often correspond to different driving mechanisms, Fig. 12 suggests that the PCA technique is able to differentiate directly between physical mechanisms and the data.

Figure 13 shows the time series of PCA for the case of Fig. 8, which includes local-time variation  $E_{amp}$ ,  $E_{1,n}$  and  $E_{4,n}$ , where  $E_{amp}$  is the amplitude of the time series for the paired EOFs ( $v_2$ ,  $v_3$ ):

$$E_{amp} = \sqrt{(E_{2,n})^2 + (E_{3,n})^2},$$

$$E_{phase} = \tan^{-1}(E_{3,n}/E_{2,n}). \tag{5}$$

Note that the paired EOFs ( $v_2$ ,  $v_3$ ) shown in Fig. 8 denote the fixed spatial patterns of the local-time variations with a  $\pi/2$  phase shift. As a result, the corresponding paired coefficients ( $E_{2,n}$ ,  $E_{3,n}$ ) are expected to have similar magnitudes and change rapidly with the local time in order to represent a moving wave structure. The amplitude defined in Eq. (5) combines the effects of both components and also eliminates the local-time variation. We also note from Fig. 13 that  $E_{1,n}$  and  $E_{amp}$  are highly correlated with a dominant period of about half a year though  $E_{1,n}$  is not correlated with either  $E_{2,n}$  or  $E_{3,n}$ . This is because the first three EOFs all represent the effect of solar radiation. On the other hand, the fourth EOF shows a clear signature of annual cycle that corresponds to the effect of the inclination angle of the Earth’s orbit. Such a difference in the effects of solar radiation on the different EOF components can also be seen from Fig. 14, which shows the scatterplots between the daily averaged  $F_{10.7}$  and the daily averaged time series ( $E_{amp}$ ,  $E_{1,n}$ ,  $E_{4,n}$ ) shown in Fig. 13. The figure shows that  $F_{10.7}$  is well correlated with both  $E_{1,n}$  and  $E_{amp}$ , whereas  $E_{4,n}$  appears to be insensitive to  $F_{10.7}$  variations. Figure 15 shows the time series of the longitudinally averaged TEC at five different latitudes: Equator ( $0^\circ$ ),  $40^\circ$  S,  $40^\circ$  N,  $60^\circ$  S and  $60^\circ$  N. Comparing this to Fig. 10, which also includes the local-time average, we note the significant increase in daily variability of TEC, sug-



**Figure 16.** Fourier power spectra of the first four PCs shown in Fig. 13 (panel a) and the time series at five different latitudes shown in Fig. 14 (panel b). To avoid clustering, the green and red curves in (a) have been moved down by 3 and 6 orders of magnitude, respectively. All the curves except the one at the Equator in (b) have been consecutively shifted downward by 3 orders of magnitude. The cyan curve in panel a denotes the spectrum of the phase  $10^4 E_{\text{phase}}$  defined in Eq. (5).

gesting that the longitudinal and local-time averages are not interchangeable for the TEC field.

In Fig. 16, we show the Fourier power spectra for the time series shown in Figs. 13 and 15. The six vertical lines denote the peak frequencies that correspond to the following periods: annual, semiannual, 27-day, 1-day, 0.5-day and 0.25-day. Panel a in the figure shows that  $E_{\text{amp}}$  has significant spectral peaks in all the noted frequencies, whereas  $E_{\text{phase}}$  does not show spectral peaks at the 27-day and semiannual periods. We also note that there are significant sub-harmonic peaks in the  $E_{\text{amp}}$  spectrum for a frequency greater than  $2\pi/1$ -day. The spectral peaks at  $3\pi/1$ -day and  $5\pi/1$ -day result from the sum of spectral peaks at  $2\pi/1$ -day plus  $\pi/1$ -day and  $3\pi/1$ -day plus  $2\pi/1$ -day, respectively. Again, the spectral features of  $E_{\text{amp}}$  are similar to that for the third EOF shown in Fig. 13, whereas the spectrum for the fourth EOF is dominated by the annual cycle, also consistent with the results shown in Fig. 13.

#### 4 Summary and conclusions

In this study, we apply the EOF and PCA to the global TEC data derived from TIEGCM forced under the realistic solar inputs from above the SABER-observed tidal waves from below. We demonstrate the effectiveness of the EOF decomposition of the ionospheric variations in both time and space. It is shown that for the daily averaged TEC field, the first EOF explains more than 89% and the first four EOFs ex-

plain more than 98% of the total variance of the TEC field. When PCA is coupled with the spectral analysis of the time series of the EOF coefficients, it is also shown that the EOF analysis is not only a data compression technique but also a powerful tool to objectively reveal the relative importance of individual physical mechanisms (such as solar flux variation, change in the orbital declination, nonlinear mode coupling and geomagnetic activity) that are responsible for the total TEC variance.

#### 5 Data availability

The TIEGCM output TEC fields that generated all the figures in this paper are available from Xun Zhu (xun.zhu@jhuapl.edu) upon request.

*Acknowledgements.* This research was supported by NASA Living With a Star Program under grants NNX09AJ61G and NNX13AF91G and the Heliophysics Supporting Research program under grant NNX16AG68G to the Johns Hopkins University Applied Physics Laboratory. Comments on the paper by two anonymous reviewers are greatly appreciated.

The topical editor, K. Shiokawa, thanks two anonymous referees for help in evaluating this paper.

#### References

- Lean, J.: Solar ultraviolet irradiance variations: A review, *J. Geophys. Res.*, 92, 839–868, 1987.
- Preisendorfer, R. W.: *Principal Component Analysis in Meteorology and Oceanography*, edited by: Mobley, C. D., Elsevier, Amsterdam, the Netherlands, 425 pp., 1988.
- Richmond, A. D., Ridley, E. C., and Roble, R. G.: A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophys. Res. Lett.*, 19, 601–604, 1992.
- Roble, R. G., Ridley, E. C., and Richmond, A. D.: A coupled thermosphere/ionosphere general circulation model, *Geophys. Res. Lett.*, 15, 1325–1328, 1988.
- Talaat, E. R. and Lieberman, R. S.: Nonmigrating diurnal tides in the mesosphere and lower thermosphere, *J. Atmos. Sci.*, 56, 4073–4087, 1999.
- Talaat, E. R., Yee, J.-H., and Zhu, X.: Observations of the 6.5-day wave in the mesosphere and lower thermosphere, *J. Geophys. Res.*, 106, 20715–20724, doi:10.1029/2001JD900227, 2001.
- Tobiska, W. K.: Recent solar extreme ultraviolet irradiance observations and modeling: A review, *J. Geophys. Res.*, 98, 18879–18893, 1993.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, Second Edition, Academic Press, New York, USA, 627 pp., 2006.