Single station TEC modelling during storm conditions

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

of

RHODES UNIVERSITY

by

Jean Claude Uwamahoro

October 2015

Abstract

It has been shown in ionospheric research that modelling total electron content (TEC) during storm conditions is a big challenge. In this study, mathematical equations were developed to estimate TEC over Sutherland (32.38^{o}S, 20.81^{o}E), during storm conditions, using the Empirical Orthogonal Function (EOF) analysis, combined with regression analysis. TEC was derived from GPS observations and a geomagnetic storm was defined for Dst $\leq$ -50 nT. The inputs for the model were chosen based on the factors that influence TEC variation, such as diurnal, seasonal, solar and geomagnetic activity variation, and these were represented by hour of the day, day number of the year, F10.7 and A index respectively. The EOF model was developed using GPS TEC data from 1999 to 2013 and tested on different storms. For the model validation (interpolation), three storms were chosen in 2000 (solar maximum period) and three others in 2006 (solar minimum period), while for extrapolation six storms including three in 2014 and three in 2015 were chosen. Before building the model, TEC values for the selected 2000 and 2006 storms were removed from the dataset used to construct the model in order to make the model validation independent on data. A comparison of the observed and modelled TEC showed that the EOF model works well for storms with non-significant ionospheric TEC response and storms that occurred during periods of low solar activity. High correlation coefficients between the observed and modelled TEC were obtained showing that the model covers most of the information contained in the observed TEC. Furthermore, it has been shown that the EOF model developed for a specific station may be used to estimate TEC over other locations within a latitudinal and longitudinal coverage of 8.7^o and 10.6^o respectively. This is an important result as it reduces the data dimensionality problem for computational purposes. It may therefore not be necessary for regional storm-time TEC modelling to compute TEC data for all the closest GPS receiver stations since most of the needed information can be extracted from measurements at one location.

Acknowledgement

I greatly appreciate my supervisor, Dr John Bosco Habarulema, for his rigorous supervision and useful suggestions. I know that without his critical comments and scientific contribution, this work could not be completed .

Many thanks to the National Astrophysics and Space Science Progamme (NASSP) for awarding me bursaries for my Honours and Masters degrees. I am also grateful to the South African National Space Agency (SANSA) for additional supports as well as for introductory training in scientific research in Space Physics.

A special thank-you to Laura Wentworth for her assistance in Python programming, also to all my friends who provided so much support throughout my studies: Lifa Mbuli and Tsige Atilaw for their help with Latex techniques, Matamba Tshimangadzo for her useful documentation on the storm database, and Nigussie Giday for advice and scientific guidance. Thanks to Ercha A for his help and suggestions in implementing the EOF technique.

I would like to express my sincere gratitude to my family, especially to Claudine Uwineza, Claudette Umugwaneza and Claudia Ujeneza, for encouragement, support, and for being with me during all my studies.

All who have contributed directly or indirectly to the realisation of this work and whose names are not mentioned here, please know that I am very grateful to you for all your help.

Thank you, God, for being with me during all my life and for helping me to accomplish this milestone.

Dedication

This thesis is dedicated to my loving wife Claudine Uwineza and my daughters Claudette Umugwaneza and Claudia Ujeneza.

Contents

Abstract

Aknowledgement i

Dedication ii

List of Figures iv

List of Tables vii

1 Introduction 1

1.1 Research Objective . . . 3

1.2 Thesis outline . . . 4

2 Theoretical background 5

2.1 Brief introduction to the ionosphere . . . 5

2.2 Ionospheric TEC variations . . . 6

2.2.1 Diurnal variation of TEC . . . 7

2.2.2 Seasonal variation of TEC . . . 8

2.2.3 TEC variation with latitude . . . 9

2.2.4 TEC variation with solar activity . . . 10

2.3 Geomagnetic storm . . . 11

2.3.1 Phases of a geomagnetic storm . . . 12

2.3.2 Causes of geomagnetic storms . . . 13

2.4 Geomagnetic and solar indices . . . 13

2.4.1 Disturbance storm-time index . . . 13

2.4.2 K index . . . 15

2.4.3 A index . . . 16

2.4.4 F10.7 index . . . 17

2.5 Summary . . . 17

3 Data sources, processing and modelling techniques 18

3.1 GPS overview . . . 18

3.1.1 GPS structure . . . 18

3.1.1.1 Space segment . . . 18

3.1.1.2 Control segment . . . 19
3.1.1.3 User segment . . . 21
3.1.2 GPS signals . . . 21
3.1.3 The effect of the ionosphere on GPS signals . . . 21
3.1.4 Deriving TEC from GPS measurements . . . 25
3.2 Data processing . . . 27
3.3 Modelling techniques . . . 28
3.3.1 Regression analysis . . . 28
3.3.2 Empirical Orthogonal Function (EOF) analysis . . . 29
3.3.2.1 Goals of EOF analysis . . . 30
3.3.2.2 Mathematical description of EOF analysis . . . 30
3.4 Summary . . . 35

4 Results and discussion **36**
4.1 Model construction . . . 36
4.1.1 TEC decomposition using EOF analysis . . . 36
4.1.2 Modelling the EOF coefficients $A_k(d)$ using the regression analysis . . . 41
4.2 Application of EOF model to selected storms . . . 43
4.2.1 TEC modelling during the storms of 2000 and 2006 (Interpolation) . . . 44
4.2.2 TEC modelling during the storms of 2014 and 2015 (Extrapolation) . . . 45
4.3 Statistical analysis . . . 48
4.4 Model validation over other stations . . . 51
4.5 Model improvement . . . 54
4.6 Summary . . . 60

5 Conclusions and future work **62**
5.1 Conclusions . . . 62
5.2 Challenges in modelling storm-time TEC using the EOF analysis . . . 64
5.3 Future work . . . 64

List of Figures

2.1 Diurnal variation of TEC over Sutherland (32.38^o S, 20.81^o E), South Africa, on 01-11-2014. Local time (LT) = UT + 2 hours. 8
2.2 Annual variation of TEC during 2003, over Sutherland (32.38^o S, 20.81^o E), South Africa. 9
2.3 Global TEC map for 09-03-2012, at 12:00 UT. Data used for the generation of the map was obtained from ftp://cddis.gsfc.nasa.gov/pub/gps/products/ionex. 10
2.4 TEC variation with solar activity over Sutherland (32.38^o S, 20.81^o E), in South Africa. 11
2.5 Motions of charged particles in the Earth's magnetosphere (Baumjohann *et al.*, 1997). 14
2.6 Four phases of a geomagnetic storm that occurred on 15-05-2005, as defined by Dst index. 15
2.7 Hermanus K-indices, on 17-03-2015. 16

3.1 GPS constellation of 24 satellites distributed in six orbits (Kaplan and Hegarty, 2005) 19
3.2 The GPS operational control segment consists of a master control station, an alternate master control station, 12 command and control antennas, and 16 monitoring stations (GPS.gov, 2015). 20
3.3 STEC to VTEC mapping (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006) 26
3.4 Satellite paths on 17-03-2015 over Sutherland (32.38^o S, 20.81^o E) (a) before selecting elevation angles greater than 20^o, and (b) after selecting elevation angles greater than 20^o. 27
3.5 Example of a scree plot 34

4.1 The 865 $\times$ 24 matrix of hourly TEC values. 37
4.2 Diurnal variation of the first twelve EOF base functions. 39
4.3 Diurnal variation of the first EOF base function and the average TEC. . . . 39
4.4 Long-term variations of the EOF coefficients A_k. 41
4.5 Variations of A_1 and F10.7 index for the storm periods of 1999 - 2013. 42

4.6 Comparison between observed and modelled TEC for storms that occurred during the high solar activity period on (a) 24-05-2000 (b) 16-07-2000 and (c) 11-12-08-2000 45
4.7 Comparison between observed and modelled TEC for storms that occurred during the low solar activity period on (a) 09-04-2006 (b) 14-04-2006 and (c) 15-12-2006 46
4.8 Comparison between observed and modelled TEC for storms that occurred (a) February 2014 (b) 27-08-2014 and (c) 12-09-2014 47
4.9 Comparison between observed and modelled TEC for storms that occurred (a) 04-01-2015 (b) 07-01-2015 and (c) 17-03-2015 48
4.10 RMSE values for selected storm periods. The RMSE of each storm period is shown by a bar that corresponds to the storm date. 49
4.11 Comparison between the modelled and observed values of TEC for twelve selected storms. Also shown, are the correlation coefficients between the observed and modelled TEC. The storm period is specified for each storm. . . . 50
4.12 Locations of different GPS receiver stations relative to IPPs over Sutherland (32.38^o S, 20.81^o E) after selecting satellites with elevation angles greater than 20^o. 52
4.13 Comparison between observed and modelled TEC during the storm period 18 - 24 February, 2014 over (a) ANTH, CTWN, PELB and SBOK (b) BETH, MFKG, ERAS and ULDI. 54
4.14 Comparison between observed and modelled TEC during the storm period from 21 - 26 January, 2004, over (a) ANTH, GEOA, SBOK and ELDN (b) BETH, MFKG, ERAS and ULDI. 55
4.15 Comparison between observed and modelled TEC during the storm period 06 - 10 March 2012, over (a) ANTH, CTWN, PELB and SBOK (b) BETH, MFKG, ERAS and ULDI. 56
4.16 RMSE values for selected storms (a) 18 - 24 February, 2014 (b) 21 - 26 January, 2004 (c) 06 - 10 March 2012. 57
4.17 Comparison between observed and modelled TEC for storms that occurred (a) 18 - 24 February, 2014 (b) 21 - 26 January, 2004 (c) 06 - 10 March 2012. Also shown, are the correlation coefficients between the observed and modelled TEC. 58
4.18 RMSE values for selected storms using modified indices. The RMSE values for each storm are shown by bars that correspond to the storm date. 59
4.19 Frequency of occurrence each set of indices was found to give the smallest RMSE values. 60
4.20 Comparison between observed and modelled TEC using modified indices as inputs: (a) storms of 11-12 August, 2000, (b) storm of 14-04-2006. 60

List of Tables

2.1 Storm classification by Loewe and Prölss (1997). 15
2.2 Conversion between K and a. 17

4.1 Proportion of variance accounted for by each EOF component. 38
4.2 Geographic latitudes (GLAT) and longitudes (GLON) of GPS receiver stations used in this project. 51

Chapter 1

Introduction

Magnetic storms cause disturbances in the Earth's magnetic field which later cause changes in the ionospheric electron density and hence in ionospheric total electron content (TEC) (Yizengaw and Essex, 2002). TEC is defined as the total number of electrons within an imaginary cylinder of cross-sectional area of 1 m^2 between a satellite and a receiver on the ground (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006; Habarulema *et al.*, 2009; Rao *et al.*, 2013). TEC undergoes dramatic changes during geomagnetic storms resulting into two major effects, classified as positive and negative storm effects, corresponding to the increase and decrease in TEC respectively (Habarulema *et al.*, 2013; Borries *et al.*, 2015). This important ionospheric parameter can be used for many purposes, including the study of the ionosphere-plasmasphere system and Global Navigation Satellite Systems (GNSS) applications (Stankov *et al.*, 2010). For example, one can get information about the ionisation level of the ionosphere by measuring changes in the wave parameters when the signal transmitted by a satellite interacts with the ionospheric plasma (Jakowski *et al.*, 2012). On the other hand, it is known that TEC perturbations have a significant impact on satellite applications such as satellite navigation, communication, space weather forecasting, global positioning systems (GPS) surveying and remote sensing systems, which rely on an electromagnetic signal that interacts with ionospheric plasma when it passes through the ionosphere (Hofmann-Wellenhof *et al.*, 1992; Habarulema *et al.*, 2009; Borries *et al.*, 2015). A typical example is that TEC, encountered between satellites and ground receivers, is one of the main sources of error for positioning applications, especially for single frequency users (Bergeot *et al.*, 2013). Thus, TEC modelling is of high importance for a better understanding of its response to variations of solar activity and geomagnetic storms.

In the literature much has been reported on efforts to model TEC during both quiet and storm conditions. Using Empirical Orthogonal Functions (EOF) analysis, TEC was modelled during magnetically quiet days over China (Mao *et al.*, 2005, 2008). A global model for TEC based on EOF analysis using the global ionosphere maps provided by Jet Propulsion Laboratory has been developed using data for the period 1999 - 2009 (A *et al.*, 2012). A comparison between modelled TEC using the International Reference Ionosphere (IRI) model

and TEC modelled by means of EOF model has shown that EOF model performs better than IRI model and its accuracy is high (Mao *et al.*, 2005, 2008; A *et al.*, 2012). However, when A *et al.* (2012) applied the global EOF model to modelling TEC during storm conditions, the model did not perform well. It was then suggested that the use of TEC data of high spatial-temporal resolution, specifically during storm conditions, could improve the modelling results. TEC was also predicted over Chumphon (10.72^o N, 99.37^o E), an equatorial latitude station in Thailand, using Neural network (NN) model, and the outcome was compared with TEC predicted by the IRI-2007. It has been shown that the NN model predicts TEC better when compared with the IRI-2007 model (Watthanasangmechai *et al.*, 2012). Furthermore, the most recently available IRI-2012 model was applied to TEC modelling during quiet and storm periods. GPS TEC over a low-latitude Singapore station (01.37^o N, 103.67^o E) was used and the IRI model was not able to predict the storm's impact. It has thus been suggested that to model TEC behaviour during storms, significant improvements in the IRI model are required (Kumar *et al.*, 2014).

Recently, the NN and the IRI models were applied in modelling GPS TEC over South Africa. A comparison of GPS TEC predicted by the NN model and GPS TEC predicted by the IRI-2001 over South Africa, showed that the NN model predicts GPS TEC with higher accuracy than IRI-2001 (Habarulema *et al.*, 2007). The NN and IRI-2007 models have also been applied in modelling TEC over Southern Africa during storm conditions. Although both models are good in following storm dynamics in TEC variations, the accuracy is still low and needs to be improved (Habarulema *et al.*, 2010).

A study of variations of the ionospheric critical frequency of the F2 layer (foF2) and GPS TEC over the Antarctic sector, has confirmed that diurnal, seasonal and solar activity variations of TEC are similar to those observed in the foF2 values (Mosert *et al.*, 2011). Therefore, it is important to mention reports on the modelling of foF2, an ionospheric parameter that is influenced by ionospheric dynamics in the same way as TEC. Using ionosonde data over Rome (41.8^o N, 12.5^o E), Italy, and over Grahamstown (33.32^o S, 26.50^o E) in South Africa, for a period of 50 days from November 12 to December 31, 1997, a comparison of the monthly mean values computed from the observed data, with the values predicted by the IRI model showed that the predicted values from the IRI model were in a good approximation of the monthly mean values of foF2. However, quantitative model/data comparisons did not produce very encouraging results. For the winter mid-latitude the model reproduces the magnitude of foF2 day-to-day variability and follows the positive phase during storms; however for the summer mid-latitude both day-to-day variability and the magnitude of the negative phase were underestimated (Fuller-Rowell *et al.*, 2000). The NN technique was also used to develop a near real-time global foF2 (NRTNN) empirical model. Comparisons between the observed and

predicted values of foF2 during two magnetic storms that occurred during 17-19 September 1979 and during 17-19 November 1989, showed good agreement (Oyeyemi *et al.*, 2006).

Modelling ionospheric parameters such as TEC and foF2 during storm conditions is a big challenge. Auroral precipitation and uncertainty in the magnitude and spatial distributions of the magnetospheric electric field have been mentioned as some of the causes that make difficult to model the response to a specific disturbance at a particular location in the mid-latitude region (Fuller-Rowell *et al.*, 2000). On the other hand, due to an incomplete understanding of ionospheric dynamics, and of sudden changes in the ionospheric electron density observed during storm conditions, especially when both positive and negative storm effects are observed during one storm period, it has been noticed that modelling storm-time TEC is a difficult task (Habarulema *et al.*, 2010).

This thesis therefore focuses on the improvement of TEC modelling during magnetic storms using South African GPS data. The task was to develop a model to accurately predict both TEC magnitude and its dynamics during magnetic storm conditions, i.e., a model that estimates TEC with high accuracy and capable to capture both positive and negative storm effects.

EOF analysis together with regression analysis were used to model TEC during storm conditions over Sutherland (32.38^o S, 20.81^o E), South Africa, during the period 1999 - 2013. EOF analysis was chosen because it is capable of identifying hidden patterns in the data and classifying them according to how much of the information stored in data they account for. In addition to this particular advantage of EOF analysis, it has been shown that the EOF model performs well when it is applied to modelling some ionospheric parameters, including TEC during magnetically quiet days (Mao *et al.*, 2005, 2008; A *et al.*, 2012). Regression analysis was introduced to model the EOF coefficients in terms of solar and geomagnetic indices in order to introduce the influence of solar and geomagnetic activities on TEC within the model.

1.1 Research Objective

The main objective of the project was to develop mathematical equations to estimate TEC over Sutherland (32.38^o S, 20.81^o E), during storm conditions.

To achieve this objective, a TEC database for storm days was built for the period 1999-2013, using GPS TEC over Sutherland. Magnetic storms that occurred during the period under study were identified by means of the disturbance storm time (Dst) index, the storm

criterion being Dst ≤ -50 nT. The fact that TEC varies with time of the day and seasons was taken into account to build the model as well as solar and magnetic activities which are considered as the main causes of TEC variability.

1.2 Thesis outline

The thesis is divided into five chapters:

The first chapter gives a brief description of the project, specifying the project objective, period under study, region of interest, modelling techniques and the outcome of the project.

The second chapter gives theoretical background on TEC, geomagnetic storms, solar and geomagnetic indices.

The third chapter provides details about data measurements and processing, and a description of the modelling techniques.

The fourth chapter presents the results and a discussion on TEC modelling during storm conditions.

Finally, conclusions and future work are presented in chapter 5.

Chapter 2

Theoretical background

In this chapter a brief introduction to the ionosphere is given and details about TEC and its variation are discussed. An introduction to magnetic storms, as well as a description of the solar and geomagnetic indices that were used in this project, are provided.

2.1 Brief introduction to the ionosphere

Depending on temperature variation with height, the Earth's atmosphere is divided into regions named, from the ground to the upper limit of the atmosphere, troposphere, stratosphere, mesosphere, thermosphere and exosphere (Rishbeth and Garriott, 1969; Hunsucker and Hargreaves, 2002). The ionosphere is the partially ionised region of the Earth's atmosphere, within the mesophere and thermosphere, which extends from about 50 km above the surface of the Earth to about 1000 km (Hunsucker and Hargreaves, 2002; Moeketsi, 2008). The ionisation process originates mainly from the extreme ultraviolet (EUV) radiation and the X-rays from the sun that interact with neutral atmospheric constituents (McNamara, 1991). When a photon from the sun hits a neutral atom, an electron can escape from the atom and the latter becomes a positive ion. This process is called photoionisation and only takes place during the day when the sun is above the horizon. The reverse process of photoionisation is called recombination, simply because electrons and positive ions combine to produce neutral atoms. While photoionisation is a process by which positive ions are produced in the ionosphere, recombination is a process by which electrons are lost in the lower and upper ionosphere and only takes place when photoionisation ceases at sunset (McNamara, 1991).

The ionosphere is subdivided into three regions, termed lower ionosphere, bottomside ionosphere and topside ionosphere. Since the EUV radiation that mainly causes the photoionisation is attenuated as it passes through the ionosphere towards the lower atmosphere, the intensity of EUV radiation decreases as the radiation penetrates deeper. On the other hand, the density of neutral atoms that can be photoionised decreases as the altitude increases, i.e. towards the topside ionosphere. A net result of these opposing effects is the formation of a

layer of maximum electron density at some altitude in the ionosphere, whereas above and below this altitude, layers with lower electron density are produced. Thus, the bottomside ionosphere is subdivided into three regions according to the altitude, namely D, E, and F, and the latter in turn contains two layers named F1 and F2 (McNamara, 1991).

- The D layer is the lowest ionospheric layer with an altitude range from 50 km to 90 km (McNamara, 1991; Moldwin, 2008). The main source of ionisation is cosmic rays and X-rays from solar flares. The electron density is lower in the D layer compared to other layers, since the D layer is weakly ionised, and at approximately 90 km, is typically 10^{10} m^{-3}, during the daytime (Ondoh and Marubashi, 2001; Moeketsi, 2008; Habarulema, 2010). During the nighttime, the D layer disappears due to the recombination process, and its electron density is recovered during the daytime due to photoionisation (Moldwin, 2008).

- The E layer is above the D layer at an altitude ranging from 90 km to 120 km. X-rays and UV solar radiation of molecular oxygen (O_2) are the major sources of ionisation (Moldwin, 2008). At an altitude of about 105 km, the typical peak electron density is 10^{11} m^{-3} (Moeketsi, 2008). Except for the remaining ionisation during the night, the E layer disappears at sunset and appears at sunrise (Davies, 1990; Moldwin, 2008).

- The F layer is the ionospheric region that starts from an altitude of 120 km, just above the E layer and contains two layers, F1 and F2, that are present during the daytime. The F layer formation is due mainly to the solar EUV radiation that ionises oxygen atoms, and the typical daytime maximum electron density is 10^{12} m^{-3} (Moeketsi, 2008; Moldwin, 2008). At night, since recombination is the dominating process, the F1 layer almost completely disappears, while F2 survives but gets depleted. For this reason, the F2 layer is the most important for the propagation of HF radio waves (McNamara, 1991).

2.2 Ionospheric TEC variations

As previously defined, TEC is a measure of the total number of electrons encountered inside an imaginary cylinder with a cross-sectional area of 1 m^2 between a satellite and a receiver on the ground. Mathematically, TEC is computed by integrating the electron density with respect to the altitude (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006):

$$TEC = \int_S^R N_e(l)dl \tag{2.1}$$

where R and S in the integration limits stand for receiver and satellite, respectively. $N_e(l)$ represents the electron density profile (number of electrons per volume unit) along the signal path and dl is an element of distance in such a way that

$$l = \int dl \tag{2.2}$$

represents the geometric distance measured along the straight line between the satellite and the receiver. TEC is expressed in TEC unit (TECU) and 1 TECU = 10^{16} electrons/m^2.

Since TEC depends on the electron density according to the Equation 2.1 and the electron density in turn depends on the solar radiation which is the main source of photoionisation, TEC variability is highly influenced by several factors that cause changes in the intensity of the solar radiation received on the Earth. Solar and geomagnetic activity, geographic location of a particular region on the Earth (latitude and longitude), and changes in solar radiation with time, are the primary causes of spatial and temporal variations of the ionospheric TEC (Huang and Roussel-Dupré, 2005). The spatial variation of TEC is generally influenced by various ionospheric latitude zones, while temporal variation is due to changes in solar radiation with time. Temporal variation includes diurnal variation i.e. over a period of 24 hours of a day, seasonal variation, and solar cycle variation observed over an 11-year period between high and low solar activity (Huang and Roussel-Dupré, 2005). The different factors that influence/cause TEC variation are summarised below.

2.2.1 Diurnal variation of TEC

High TEC values are usually observed around local midday, since the photoionisation rate is higher due to the maximum solar radiation intensity. Low TEC values are observed at night from sunset until just before sunrise due to the decrease in ionisation rate, since solar radiation intensity is low at night. The general trend of the diurnal variation of TEC is illustrated by the Figure 2.1 which shows an example of diurnal variation of TEC over Sutherland (32.38^o S, 20.81^o E), South Africa, during a magnetically quiet day on 01-11-2014.
Similar observations have been reported for different regions. A study of diurnal variation of TEC over a low latitude station, Agran (27.12^o N, 78.89^o E), India, using TEC data for a period from 01 August 2006 to 31 July 2009, showed that minimum values of TEC were observed around 05:00 LT of the morning while the diurnal peak in TEC was generally observed between 12:00 and 14:00 LT, in the afternoon (Chauhan *et al.*, 2011). Other authors described the general pattern of diurnal variation of TEC as follows: a short time depression in TEC early in the morning, followed by an increase in TEC in the morning until it reaches its maximum in the early afternoon, and then a decrease in TEC after the sunset (Huang and Roussel-Dupré, 2005; Ya'acob *et al.*, 2010).

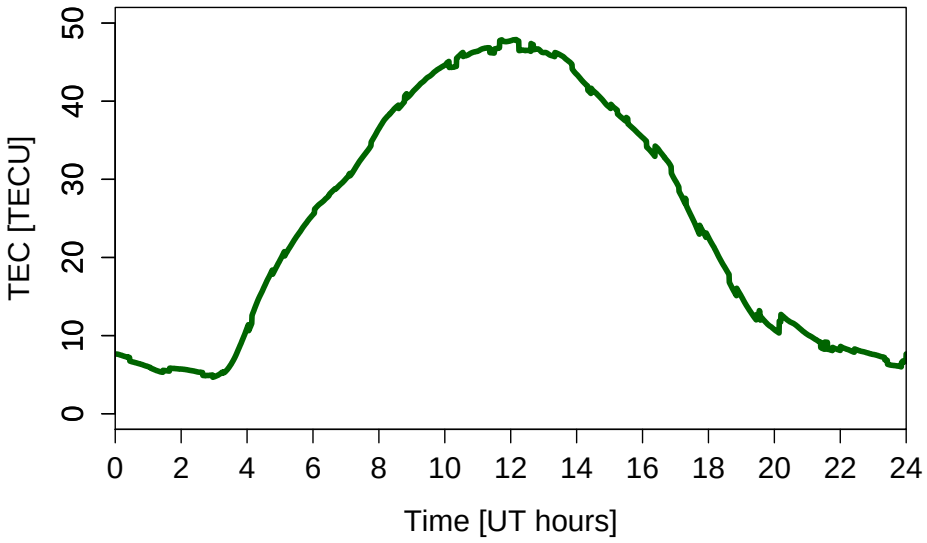


Figure 2.1: Diurnal variation of TEC over Sutherland (32.38^o S, 20.81^o E), South Africa, on 01-11-2014. Local time (LT) = UT + 2 hours.

2.2.2 Seasonal variation of TEC

Usually high TEC values are observed around the equinoxial months (March and September) and in summer while low TEC values are observed in winter (McNamara, 1991; Habarulema, 2010). During the equinoxes, the sun is at zenith over the equator, which means that it shines more directly on the Earth. Therefore, due to the high level of photoionisation, the ionospheric electron density increases and consequently, high TEC values are observed. Furthermore, it has been found that the zenith angle at noon in winter is always greater than the corresponding solar zenith angle in summer. This means that the sun is more overhead in summer than it is in winter and as consequence, higher electron density and TEC in summer than in winter. Although the seasonal variation of TEC has been described as above, it may not always be the case, since ionospheric variability also relies on the changes of the neutral atmosphere from which the ionosphere is created (McNamara, 1991).
Figure 2.2 illustrates the 2003 annual variation of TEC over Sutherland. Only TEC values at 10:00 UT (which corresponds to 12:00 South African LT) were chosen to represent diurnal values of TEC. It is noticeable that high TEC values were observed in March and in the October-November period, just close to the equinoxial months. From a semiannual view, the trend of TEC looks like a periodic function characterised by high TEC values at the equinoxes and low TEC values at the solstices (June and December).

A similar trend has been reported for other latitude regions. For example, for the low-

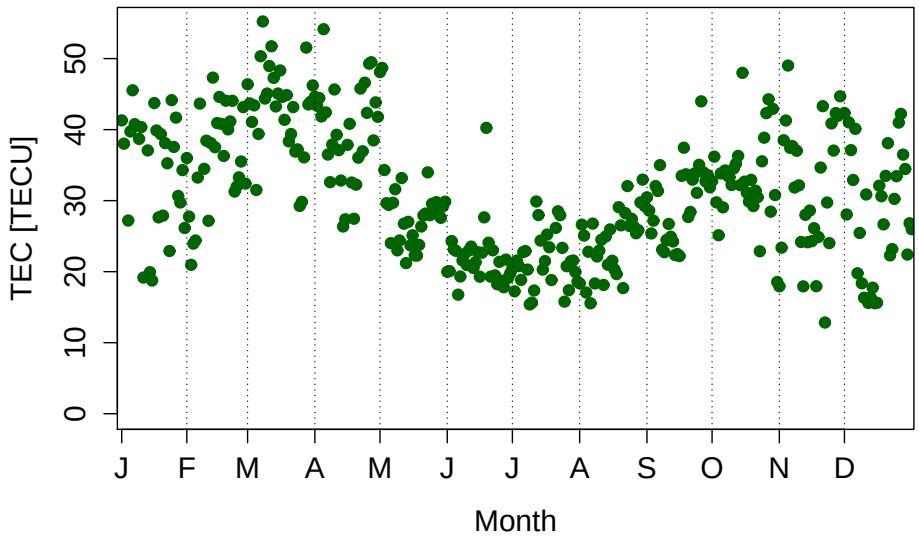


Figure 2.2: Annual variation of TEC during 2003, over Sutherland (32.38^o S, 20.81^o E), South Africa.

latitude stations, Agran (27.12^0 N and 78.89^0 E) and Rajkot (22.29^o N, 70.74^o E), India, it was reported that high TEC values were observed during the equinoxial months, low TEC value were observed in winter, while intermediate values were observed in summer (Bagiya *et al.*, 2009; Chauhan *et al.*, 2011).

2.2.3 TEC variation with latitude

High TEC values are usually found in low latitude regions, i.e. near the equator. This is due to the high level of photoionisation that is produced over the equatorial region, since the sunlight is more direct over this region than elsewhere. As ones moves from the equator towards the poles, the zenith angle increases and the solar radiation hits the Earth's atmosphere at an oblique angle, meaning that the rate at which the photoionisation occurs is less than over low latitudes (McNamara, 1991; Habarulema, 2010).

Figure 2.3 shows TEC variation over different latitude regions. It is clear that high TEC values are observed near the geomagnetic equator within a latitude range of 20^o above and below the geomagnetic equator. Peak TEC values are found in the equatorial ionisation anomaly region. A similar observation was also reported by Huang and Roussel-Dupré (2005). The equatorial ionisation anomaly is characterised by a trough in ionization concentration at the magnetic equator and an enhancement of ionization at approximately $\pm$ 15^o from the geomagnetic equator. It is caused by a vertical upward drift of plasma across the geomagnetic field lines at equatorial latitudes due to E x B forces and the subsequent downward diffusion

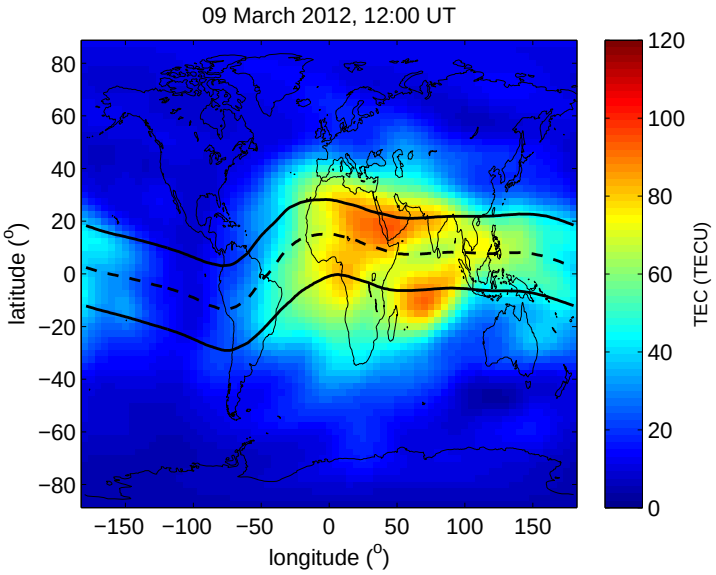


Figure 2.3: Global TEC map for 09-03-2012, at 12:00 UT. Data used for the generation of the map was obtained from ftp://cddis.gsfc.nasa.gov/pub/gps/products/ionex.

of plasma along the field lines to higher latitudes under the influence of gravity and pressure gradient forces (Balan *et al.*, 1997). .

2.2.4 TEC variation with solar activity

The sunspot number (SSN) and the solar radio flux index at a wavelength of 10.7 cm (F10.7) are highly correlated and both can be used as indicators of the general level of solar activity (McNamara, 1991). The period over which the sunspot number changes from a minimum to a maximum and then decreases to a minimum has been found to be approximately 11 years, and it is known as the solar or sunspot cycle (Herman and Goldberg, 1978; McNamara, 1991). The solar maximum occurs when the solar cycle is at its peak while the solar minimum occurs when the solar cycle is at its minimum. During the solar maximum the number of sunspots is high which indicates that there are many active regions (sunspots) on the sun's surface which could possibly release energy towards the Earth. Solar flares and CMEs that are released from the solar active regions are associated with the energy transported to the Earth. During the period of solar maximum, solar flares are frequent (Davies, 1990) and thus, we expect the ionisation rate to be higher during the solar maximum than during the solar minimum. As a consequence, enhanced ionospheric electron density and TEC are observed. The top panel of Figure 2.4 shows the solar activity as represented by the F10.7 index during the period from 1999 to 2014. It is noticeable that the solar maximum occurred in the period around 2000-2001 and then around 2012-2014, while the solar minimum period was between 2006-2010 as shown by F10.7 values which are higher during solar maximum period than during solar minimum. The bottom panel of the Figure 2.4 represents TEC variation over

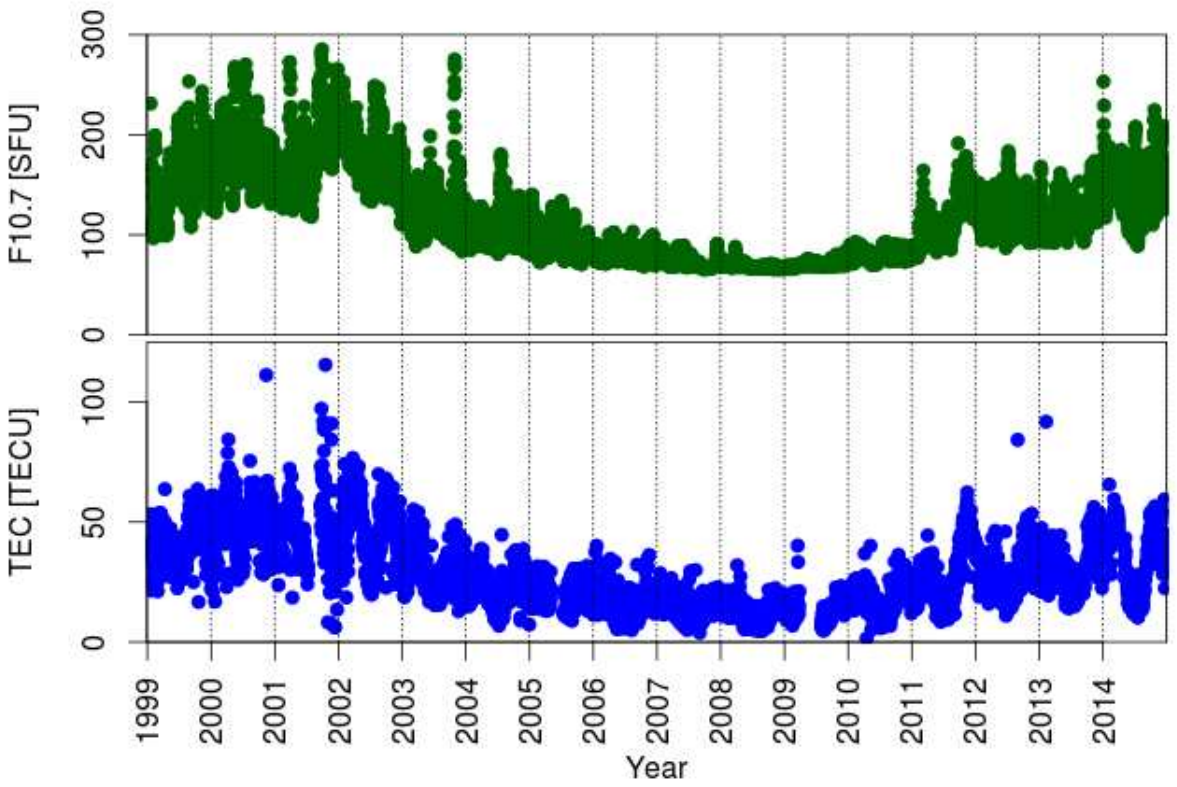


Figure 2.4: TEC variation with solar activity over Sutherland (32.38^o S, 20.81^o E), in South Africa.

the period 1999 - 2014, and it is clear that TEC values are high during the solar maximum period, whereas TEC values are low during the solar minimum. Note that only TEC values at 10:00 UT were used to represent TEC variation for the period 1999-2014.

2.3 Geomagnetic storm

The Earth's magnetic field controls the behaviour of electrically charged particles that lie within a region surrounding the Earth, called the Earth's magnetosphere (Campbell, 2001). Due to dynamic processes on the sun, it happens that a stream of charged particles, commonly known as the solar wind (mostly electrons and protons), released from the sun, perturbs the Earth's magnetic field when it hits the Earth's magnetosphere. This temporary disturbance caused by interaction between charged particles and the associated magnetic field (the inter-planetary magnetic field (IMF) or solar wind magnetic field) with the geomagnetic field is called a geomagnetic storm or simply a magnetic storm (Campbell, 2001, 1997).

Geomagnetic storms are generally observed when the solar wind magnetic field or IMF is southward directed for a prolonged period of time, a condition commonly known as the "southward IMF B_z component". When this happens, the IMF cancels the Earth's magnetic field at the Earth's magnetopause and the solar wind plasma enters the Earth's magnetosphere (Tsurutani and Gonzalez, 1997).

2.3.1 Phases of a geomagnetic storm

A geomagnetic storm can have four phases: the sudden storm commencement (SSC), an initial phase, a main phase and a recovery phase (Toffoletto, 2004).

A sudden change in the Earth's magnetic field associated with a wave shock due the arrival of the solar wind at the Earth's magnetosphere is known as the storm sudden commencement (SSC) (Campbell, 1997). However two cases need to be distinguished: after the shock, if the IMF associated with the disturbance remains northward behind the shock, normally no storm will follow and the shock is called a geomagnetic sudden impulse. If the IMF associated with the disturbance is southward oriented (as seen from negative values of B_z) behind the shock, a geomagnetic storm follows and the impulse is called a SSC (Campbell, 1997).

The initial phase is caused by an increase in ram pressure of the solar wind associated with the increase in density and speed at and behind the shock, as the interplanetary shock wave hits the magnetosphere (Tsurutani and Gonzalez, 1997).

The onset of a main phase of a storm is initiated by a physical process known as magnetic reconnection that takes place at the Earth's magnetopause. The magnetic reconnection between the IMF and the Earth's magnetic field is a mechanism during which the solar wind energy is transferred into the Earth's magnetosphere (Tsurutani, 2001). When the IMF B_z component is southward oriented, the solar wind and the Earth's magnetic fields are oppositely directed. When this phenomenon takes place, there is a cancellation of the magnetic field and hence a creation of a neutral region and magnetic reconnection takes place: plasma outside the magnetosphere recombines with the plasma inside the magnetosphere since the solar wind just opens up the magnetosphere like a can-opener, and matter squirts in. Due to the particles injected into the inner magnetosphere, the ring current is enhanced and as consequence, a depression in the magnetic field is observed (Baumjohann *et al.*, 1997).

The storm recovery phase is associated with the loss of the ring current particles from the magnetosphere via different physical mechanisms such as plasma convection, charge exchange

with particles of the neutral atmosphere, Coulomb collisions and wave-particles resonant interactions (Tsurutani, 2001).

2.3.2 Causes of geomagnetic storms

Coronal mass ejections (CMEs) from the sun and the associated interplanetary shock waves, and the so-called Corotating Interaction Regions (CIRs) that result from interactions between high-speed solar wind streams emanating from the coronal holes and slow-speed solar winds in the interplanetary medium, are the main causes of magnetic storms that occur in the Earth's magnetosphere (Srivastava and Venkatakrishnan, 2004). During the solar maximum, i.e. when the solar activity is high, CMEs associated with solar events such as solar flares and eruptive prominences, are the main causes of large geomagnetic storms (Veenadhari *et al.*, 2012). During the solar minimum, i.e. when the solar activity is low, or during the declining phase of the solar cycle, the CIRs are the main cause of moderate storms that occur when they impinge on the Earth's magnetosphere (Tsurutani *et al.*, 2006; Veenadhari *et al.*, 2012).

2.4 Geomagnetic and solar indices

Some details about solar and geomagnetic indices are provided here. In the following, only the Disturbance storm-time (Dst), K, A and F10.7 indices are described. Dst is used to define the storm period and strength, the local geomagnetic indices A and K, are commonly used to determine the level of geomagnetic activity at a particular observatory, and F10.7 is a good indicator of solar activity.

2.4.1 Disturbance storm-time index

As shown in Figure 2.5, charged particles trapped by the Earth's magnetic field execute three types of motions: spiral motion around the magnetic field lines, bounce motion back and forth along the magnetic field lines and drift motion across the field lines, in such a way that positive ions drift westward, while electrons move eastward around the Earth . The current associated with this drift motion of protons and electrons moving in opposite directions is called the ring current and flows in the westward direction, in the equatorial plane (Baumjohann *et al.*, 1997).

The Dst index is related to the ring current in such a way that when the ring current intensity increases, Dst decreases. During the main phase of magnetic storms, when particles are injected into the magnetosphere, the ring current increases. Since the ring current itself induces a magnetic field, this induced field opposes the Earth's magnetic field and consequently, the magnitude of the resulting field decreases. This is what is reported as the Dst index and the

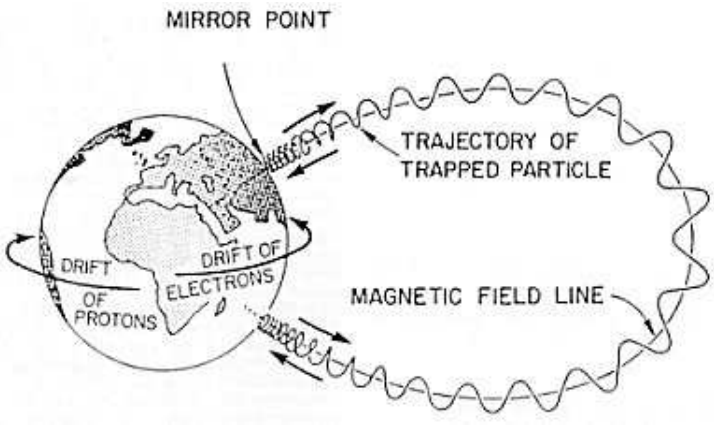


Figure 2.5: Motions of charged particles in the Earth's magnetosphere (Baumjohann *et al.*, 1997).

ring current behaves like a diamagnetic current (Tsurutani, 2001; Davies, 1990). Therefore, the Dst index is a measure of the decrease in the horizontal component of the Earth's magnetic field near the magnetic equator due to an increase in the magnetospheric ring current. The Dst index is normally expressed in nanotesla (nT) and four low-latitude magnetic observatories are used to measure the global average geomagnetic perturbation of the ring current (Campbell, 1997).

The Dst index is used to define the occurrence of a geomagnetic storm, its intensity and duration. Before the onset of a geomagnetic storm, Dst normally varies around 0 nT. The SSC is characterised by a positive sudden rise in Dst. During the initial phase, Dst is positive and relatively constant, whereas during the main phase Dst decreases until it reaches its minimum value. During the storm recovery phase, Dst slowly increases from its minimum value towards normal values around zero which corresponds to quiet geomagnetic conditions (Love and Gannon, 2010; Rathore *et al.*, 2014).

Figure 2.6 illustrates four phases of a geomagnetic storm that occurred on 15-05-2005. Note that not all magnetic storms necessarily have the four phases discussed above. Even if the minimum Dst value reached during a geomagnetic storm is used to classify its strength, there is no fixed minimum Dst value as a reference value to determine whether a storm has occurred or not. For example, Rathore *et al.* (2014) defined a magnetic storm as Dst$\leq$ -50 nT while Loewe and Prölss (1997) defined a magnetic storm as Dst $\leq$ -30 nT.

Table 2.1 shows a common classification of geomagnetic storms based on the minimum Dst value reached during the storm conditions (Loewe and Prölss, 1997). Based on this classification, the storm that occurred on 15-05-2005 is classified as a severe storm.

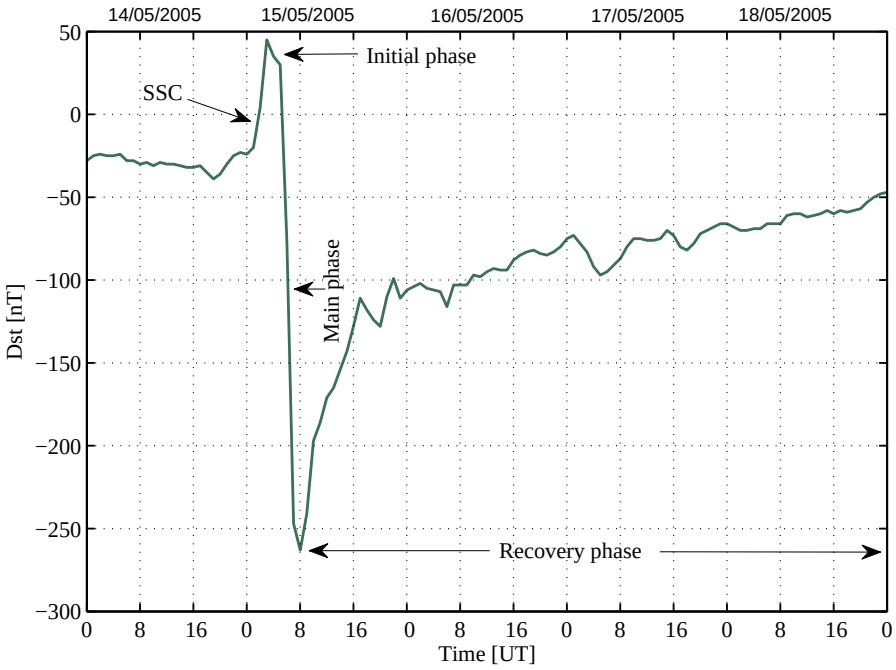


Figure 2.6: Four phases of a geomagnetic storm that occurred on 15-05-2005, as defined by Dst index.

Table 2.1: Storm classification by Loewe and Prölss (1997).

Storm class	Range of minimum Dst
Weak	-30 nT to -50 nT
Moderate	-50 nT to -100 nT
Strong	-100 nT to -200 nT
Severe	-200 nT to -350 nT
Great	$<$ -350 nT

2.4.2 K index

The K index is specified by an integer in the range 0 - 9, where 0 indicates very magnetically quiet and 9 indicates extremely magnetically disturbed conditions. According to the National Oceanic and Atmospheric Administration (NOAA) Space Weather Scale for Geomagnetic Storms, for a specific magnetic observatory, any value of K in the range 5 - 9 indicates a geomagnetic storm. The K index provides information on the state of the geomagnetic field at a particular observatory and is derived from the maximum fluctuations of the horizontal component of the geomagnetic field observed on a magnetometer at that particular observatory, during a three-hour interval: 0 - 3, 3 - 6,..., 21 - 24 UT (Davies, 1990; Reeve, 2010).

Whereas the local K index is an indicator of disturbances in the Earth's magnetic field

at a particular observatory, the planetary K index or simply Kp index is a global geomagnetic storm index based on 3 hour measurements of the K-indices. Kp index is derived by computing a weighted average of K indices from a network of geomagnetic observatories. Similar to K index, the Kp index ranges from 0 to 9 where a value of 0 means that there is very little geomagnetic activity and a value of 9 means extreme geomagnetic storm (Davies, 1990; Reeve, 2010).

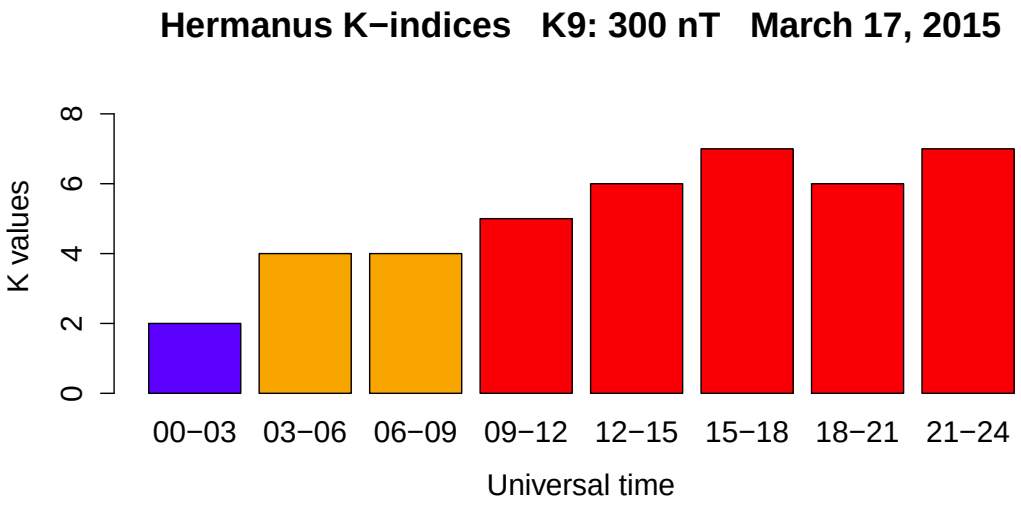


Figure 2.7: Hermanus K-indices, on 17-03-2015.

Figure 2.7 illustrates K-values observed at the South African National Space Science Agency (SANSA) Directorate, Hermanus, on 17-03-2015. It is clear that the Earth's magnetic field was magnetically disturbed, indicating a magnetic storm occurrence, as the observed K-values are greater than 4 from 09:00 UT.

2.4.3 A index

The A index was invented due to the need for an index that provides information about the daily average level of geomagnetic activity. However, the K index is not suitable for this purpose, because of its quasi-logarithmic nature and hence, it cannot be averaged (Davies, 1990; Reeve, 2010). Rather than averaging K indices, each K index is converted into its equivalent three hourly "a index", and then, the daily A index is computed by taking the average of the eight a indices (Reeve, 2010). Table 2.2 illustrates the conversion between K and a (Reeve, 2010).

The following example shows how the A index is calculated. In Figure 2.7, the eight values of K indices on 17-03-2015, were 2, 4, 4, 5, 6, 7, 6 7. Table 2.2 gives the corresponding "a"

Table 2.2: Conversion between K and a.

K	0	1	2	3	4	5	6	7	8	9
a	0	3	7	15	27	48	80	140	240	400

value for each K value: 7, 27, 27, 48, 80, 140, 80, 140. The daily A index is the average of the eight a indices:

A = (7 + 27 + 27 + 48 + 80 + 140 + 80 + 140)/8 = 68.625

2.4.4 F10.7 index

The solar radio flux at a wavelength of 10.7 cm, or simply F10.7, is a daily index commonly used as a proxy for solar activity and to determine the level of radiation from the sun. It is measured day-to-day at the Penticton Radio Observatory in British Columbia, Canada. F10.7 represents the intensity of solar radio emissions at a frequency of 2800 MHz (which corresponds to a wavelength of 10.7 cm), and is measured in solar flux units (1 SFU = 10^{-22} W.m^{-2}.Hz^{-1}) (Poole, 2002; Huang *et al.*, 2009). F10.7 provides information about conditions for long-distance communication since it is highly related to the rate of ionisation and particularly to the electron density in the F2 layer (Poole, 2002). TEC variability is also influenced by F10.7 in such a way that if the solar flux increases TEC magnitude increases (Carrano and Groves, August 7-16, 2008).

2.5 Summary

This chapter provided a brief theoretical background on the ionosphere, including the ionospheric layers and their dynamics as controlled by photoionisation and recombination processes. An overview of TEC variations such that diurnal, seasonal, solar cycle and latitudinal variations was given and illustrated using GPS TEC over Sutherland. A brief mention of causes and phases of geomagnetic storms was provided as well as an overview of solar and geomagnetic indices used in this study.

Chapter 3

Data sources, processing and modelling techniques

This chapter presents a brief introduction to GPS and how TEC is derived from GPS measurements. Details about the data used in this project are provided. Methods used for modelling TEC during storm conditions, i.e. EOF analysis and regression analysis, are mathematically described.

3.1 GPS overview

GPS is a radio-based navigation system developed by the United States Department of Defense, the primary objective being to offer the U.S. military the accurate position, velocity and time (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006). Although GPS was specifically designed for U.S. military purposes and is still funded and controlled by the U.S. Department of Defense, today many civil users benefit from different GPS applications, including aviation, spacecraft guidance, maritime navigation, land transportation, Geographic Information Systems (GIS), mapping, and agriculture (Kaplan and Hegarty, 2005; Misra and Enge, 2006)

3.1.1 GPS structure

GPS is composed of three major segments, namely the space segment, control segment and the user segment (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006).

3.1.1.1 Space segment

The space segment consists of a satellite constellation around the Earth. The baseline GPS constellation consists of 24 satellites, each at an altitude of about 20 200 km above the Earth's surface. The satellite orbits are approximately circular and inclined at an angle of 55^o relative to the equatorial plane, with a semimajor axis of about 26 560 km and eccentricity less than 0.02. Each orbit contains four active satellites and each satellite has an orbital period of

approximately 11.967 hours (Misra and Enge, 2006). Figure 3.1 shows the GPS constellation of 24 satellites distributed in six orbital planes.

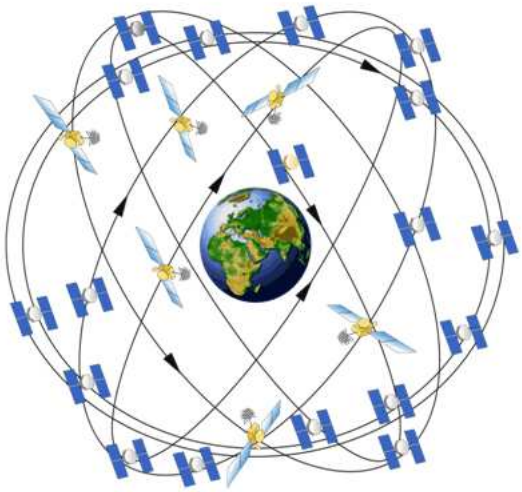


Figure 3.1: GPS constellation of 24 satellites distributed in six orbits (Kaplan and Hegarty, 2005)

Each satellite is identified by a letter that specifies its orbital plane (A, B, C, D, E, F corresponding to six orbital planes) and a number specifying the satellite number in that orbital plane (from 1 to 4 for a baseline constellation) (Misra and Enge, 2006). The pseudorandom noise or pseudorandom number (PRN) is also used to identify a GPS satellite. Since all PRN codes for all GPS satellites in a constellation are known and stored or generated in GPS satellite signal receivers on the ground, when a particular GPS satellite transmits a signal, the latter is processed in a GPS receiver and this makes it possible to determine the position of the ground receiver and to identify the satellite that has transmitted the signal (Hofmann-Wellenhof *et al.*, 1992; Grewal *et al.*, 2007).

3.1.1.2 Control segment

Satellite operations (commands and control functions) are monitored by the control segment also called Operational Control Segment (OPS) or Operational Control System. This comprises three elements: the master control station, the monitor stations and the ground antennas (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006).

- The master control station is where primary commands and control of GPS constellation are performed (Misra and Enge, 2006). The main tasks of the master control station include the generation and the uploading of navigation messages, monitoring of satellite orbits and health, satellite maintenance, satellite repositioning when needed to maintain the optimal GPS constellation, as well as the computation of the precise locations of GPS satellites in space and the uploading of this data to the satellites (GPS.gov, 2015).

- The monitor stations are used for tracking GPS satellites as they pass overhead. The observations from these stations are channelled back to the master control. Furthermore, atmospheric data, range/carrier measurements, and navigation signals are also collected by the monitor stations (GPS.gov, 2015).

- The ground antennas are used to communicate with the GPS satellites. Since the ground antennas support S-band communications links, they transmit information (satellite commands and data) to satellites via S-band radio signals (GPS.gov, 2015; Kaplan and Hegarty, 2005).

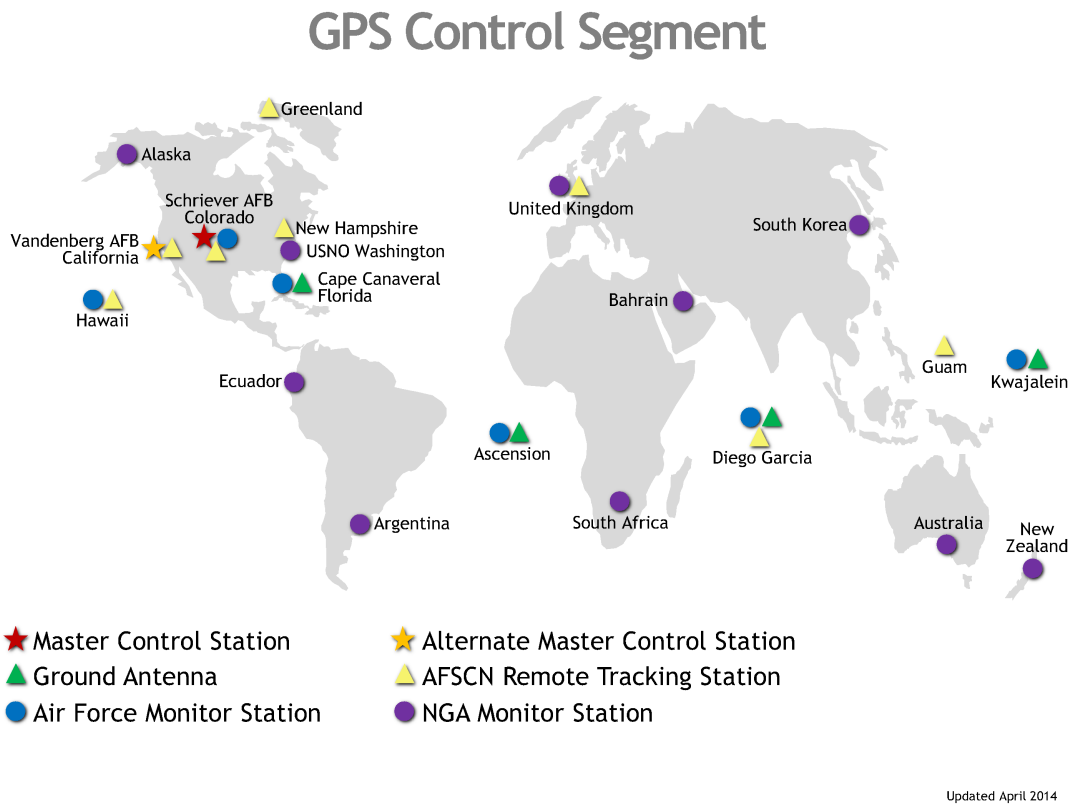


Figure 3.2: The GPS operational control segment consists of a master control station, an alternate master control station, 12 command and control antennas, and 16 monitoring stations (GPS.gov, 2015).

Figure 3.2 shows the GPS control segment which consists of a master control station and an alternate master control station, 12 command and control antennas including four ground antennas co-located with the monitor stations and eight Air Force Satellite Control Network (AFSCN) remote tracking stations, and sixteen monitoring stations including six belonging to the Air Force and ten to the National Geospatial-Intelligence Agency (NGA).

3.1.1.3 User segment

The user segment consists of user receiving equipment, typically referred to as a GPS receiver. As coded signals are transmitted from GPS satellites, they are decoded and processed in a GPS receiver to provide the position, velocity and time to the user. With the baseline constellation, at least four satellites are in view for any GPS receiver since four satellites are required to determine the position in three dimensions (X,Y,Z) and time (Kaplan and Hegarty, 2005).

3.1.2 GPS signals

The L-band which covers a frequency range between 1 GHz to 2 GHz, is a subset of the Ultra-High Frequency (UHF) band, which corresponds to radio frequencies in the range between 0.3 GHz and 3 GHz. Each GPS satellite continuously transmits signals, using two radio frequencies in the L-band, referred to as Link 1 (L1) and Link 2 (L2). The center frequencies of L1 and L2 are $f_1 = 1575.42$ MHz and $f_2 = 1227.60$ MHz, respectively (Misra and Enge, 2006). These two frequencies are derived from the fundamental frequency $f_o = 10.23$ MHz. Then f_1 and f_2 can be defined in term of f_o as follows (Hofmann-Wellenhof *et al.*, 1992; Ya'acob *et al.*, 2010)

$$f_1 = 154 \cdot f_o = 1575.42 \text{ MHz}$$
$$f_2 = 120 \cdot f_o = 1227.60 \text{ MHz}.$$

While two signals are transmitted on L1, one for civilian users and other for users authorised by U.S. Department of Defense, only one signal reserved for authorised users is transmitted on L2. To distinguish civil and authorised users, the Coarse Acquisition (C/A) Code for civilian use (also known as the Standard Positioning Service (SPS)), is modulated only onto the L1 carrier whereas the Precision-Code (P-Code) (also designated as the Precise Positioning Service (PPS)), is available for authorised users and modulated onto both L1 and L2 carriers (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006).

3.1.3 The effect of the ionosphere on GPS signals

The signals transmitted by GPS satellites travel through the ionosphere to GPS receivers. The ionosphere is not uniform in composition and the refractive index changes along the signal path. As a consequence, due to refraction, the signal path is bent and becomes longer than the geometrical straight-line path from the satellite to receiver. As a result, the ionosphere causes the signal to arrive at a GPS receiver later than it would have had if it traveled through the vacuum. The time taken by the signal to travel the distance from satellite to receiver, through the ionosphere, is given by (Misra and Enge, 2006):

$$\tau = \frac{1}{c}\int_S^R n(l)dl \tag{3.1}$$

where the integration is along the signal propagation path, and $n(l)$ is the refractive index profile of the medium and changes along the signal path. $c = 299,792,458$ m.s^{-1} is the speed of light in a vacuum. The time that the signal would take to travel the same distance if the ionosphere behaved as the vacuum is

$$\tau_o = \frac{1}{c}\int_S^R 1 \cdot dl \tag{3.2}$$

The time delay in the signal propagation due to refraction (Misra and Enge, 2006) is

$$\Delta\tau = \frac{1}{c}\int_S^R [n(l) - 1]dl \tag{3.3}$$

The equivalent increase in path length is

$$\begin{aligned} \Delta\rho &= c\Delta\tau \\ &= \int_S^R [n(l) - 1]dl \end{aligned} \tag{3.4}$$

Since the ionosphere is a dispersive medium, both the wave propagation speed and the refractive index depend on the frequency of the signal. The relation of dispersion of the ionosphere is given by (esa navipedia, 2015)

$$\omega^2 = c^2k^2 + \omega_p^2 \tag{3.5}$$

where $\omega = 2\pi f$ and $\omega_p = 2\pi f_p$ are the angular frequencies of electromagnetic signals and ionospheric plasma respectively, and $k = 2\pi/\lambda$ is the wave number. The quantities f and f_p represent the signal and plasma frequencies respectively, and λ is the wavelength of the electromagnetic signals. The frequency ω_p is a critical frequency in the sense that signals with $\omega < \omega_p$ will be reflected and signals with $\omega > \omega_p$ will cross through the plasma (Davies, 1990). The phase and the group velocities are defined by (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006)

$$v_{ph} = \frac{\omega}{k} \tag{3.6}$$

$$v_{gr} = \frac{d\omega}{dk} \tag{3.7}$$

Using Equation 3.5, the phase and the group velocities become

$$v_{ph} = \frac{c}{\sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2}} \tag{3.8}$$

$$v_{gr} = c\sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2} \tag{3.9}$$

By definition, the refractive index (n) of a medium is the ratio of the speed of propagation of the signal in a vacuum to the speed of propagation of the signal in that medium v (Hofmann-Wellenhof *et al.*, 1992):

$$n = \frac{c}{v} \tag{3.10}$$

Using this definition, the phase refractive index of the ionosphere is given by (esa navipedia, 2015)

$$\begin{aligned} n_{ph} &= \frac{c}{v_{ph}} \\ &= \sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2} \end{aligned} \tag{3.11}$$

Similarly, the group refractive index is

$$\begin{aligned} n_{gr} &= \frac{c}{v_{gr}} \\ &= \frac{1}{\sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2}} \end{aligned} \tag{3.12}$$

Taking into account that $\omega_p/\omega = f_p/f$ and using the approximation $(1+x)^\alpha \simeq 1 + \alpha x$, for $|x| \ll 1$, which is the case for the ratio f_p/f (esa navipedia, 2015), one can write the phase and the group refractive indices as follows:

$$n_{ph} = 1 - \frac{1}{2}\left(\frac{f_p}{f}\right)^2 \tag{3.13}$$

$$n_{gr} = 1 + \frac{1}{2}\left(\frac{f_p}{f}\right)^2 \tag{3.14}$$

Knowing that the angular plasma frequency is given by (Chen, 1984)

$$\omega_p = \left(\frac{N_e e^2}{\epsilon_o m_e}\right)^{1/2} \tag{3.15}$$

where $m_e = 9.1094 \times 10^{-31}$ kg is the electron mass, $e = 1.6021 \times 10^{-19}$ C is the electron charge, and $\epsilon_o = 8.854 \times 10^{-12}$ F$\cdot$ m^{-1} is the permittivity of free space, the plasma frequency is then

$$\begin{aligned} f_p &= \frac{\omega_p}{2\pi} \\ &= 8.98\sqrt{N_e} \end{aligned} \tag{3.16}$$

Substituting this expression into Equations 3.13 and 3.14, the expressions of the phase and the group refractive indices become (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006):

$$n_{ph} = 1 - \frac{40.3}{f^2} N_e \tag{3.17}$$

$$n_{gr} = 1 + \frac{40.3}{f^2} N_e \tag{3.18}$$

It is clear that $n_{gr} > n_{ph}$ which implies that $v_{gr} < v_{ph}$ according to Equations 3.11 and 3.12. The fact that the phase and group velocities are different, a group delay and a phase advance are noticed. This simply means that GPS code measurements are delayed and the carrier phases are advanced (Hofmann-Wellenhof *et al.*, 1992).

By replacing the phase and the group refractive indices by their expressions into Equations 3.3 and 3.4, and taking into account Equation 2.1, we get the expression of the excess phase delay (in seconds) experienced by a signal as it propagates through the ionosphere (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006)

$$\begin{aligned} \Delta\tau_{ph} &= \frac{1}{c}\int_S^R [n_p(l) - 1] dl \\ &= -\frac{1}{c}\int_S^R \frac{40.3 N_e(l)}{f^2} dl \\ &= -\frac{40.3}{cf^2} \cdot TEC \end{aligned} \tag{3.19}$$

The excess phase delay (in metres) is

$$\begin{aligned} I_\Phi &= c\Delta\tau_{ph} \\ &= -\frac{40.3}{f^2} \cdot TEC \end{aligned} \tag{3.20}$$

Similarly, the group delay (in seconds) is

$$\Delta\tau_{gr} = \frac{40.3}{cf^2} \cdot TEC \tag{3.21}$$

and when expressed in metres, it becomes

$$\begin{aligned} I_\rho &= c\Delta\tau_{gr} \\ &= \frac{40.3}{f^2} \cdot TEC \end{aligned} \tag{3.22}$$

The phase delay is negative, which means that the phase is advanced, whereas the group delay is positive, which justifies the appellation "group delay" (Misra and Enge, 2006).

3.1.4 Deriving TEC from GPS measurements

Dual-frequency GPS satellites above the ionosphere transmit signals on two frequencies, as mentioned above. Once the signals reach a dual-frequency receiver, the difference in ionospheric delays between L1 (f_1 = 1575.42 MHz) and L2 (f_2 = 1227.60 MHz) carriers of the GPS frequencies can be measured. Assuming that the electromagnetic signals travel along the same path through the ionosphere, and using the Equation 3.22 for the group delay, then TEC can be determined using the equation (Ya'acob *et al.*, 2008, 2009, 2010)

$$P_1 - P_2 = 40.3 \times TEC\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right) \tag{3.23}$$

which becomes, after rearranging the terms

$$TEC = \frac{1}{40.3}\left(\frac{1}{f_2^2} - \frac{1}{f_1^2}\right)^{-1}(P_1 - P_2) \tag{3.24}$$

P_1 and P_2 are the group path lengths corresponding to the high GPS frequency f_1 and the low GPS frequency f_2, respectively.

TEC measured along the signal path from satellite to receiver is called Slant Total Electron Content (STEC). The STEC can be converted into vertical TEC (VTEC), which is measured along the vertical direction, as shown in Figure 3.3. The ionosphere is assumed to be a single thin shell at the average altitude in the range 300 km - 400 km, called the mean ionospheric

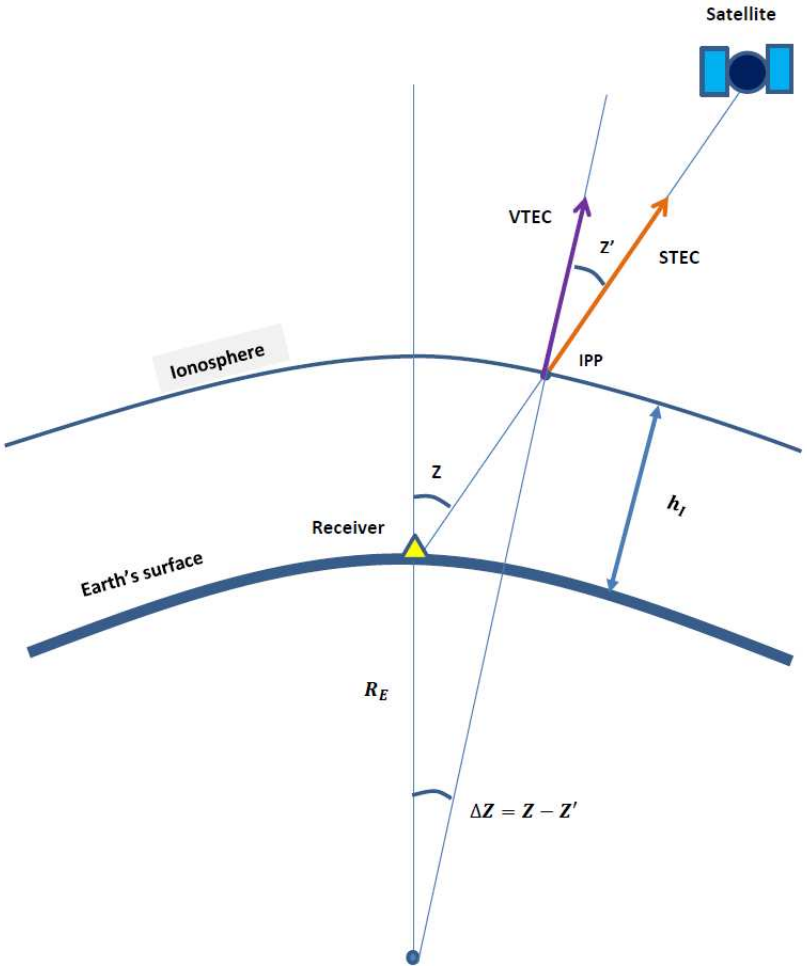


Figure 3.3: STEC to VTEC mapping (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006)

height h_I (Hofmann-Wellenhof *et al.*, 1992; Misra and Enge, 2006). The point of intersection of this spherical shell (at height h_I) and the line of sight between a satellite and a receiver is called the ionospheric pierce point (IPP). The conversion formula from STEC to VTEC is

$$STEC = \frac{1}{cosZ'} VTEC \tag{3.25}$$

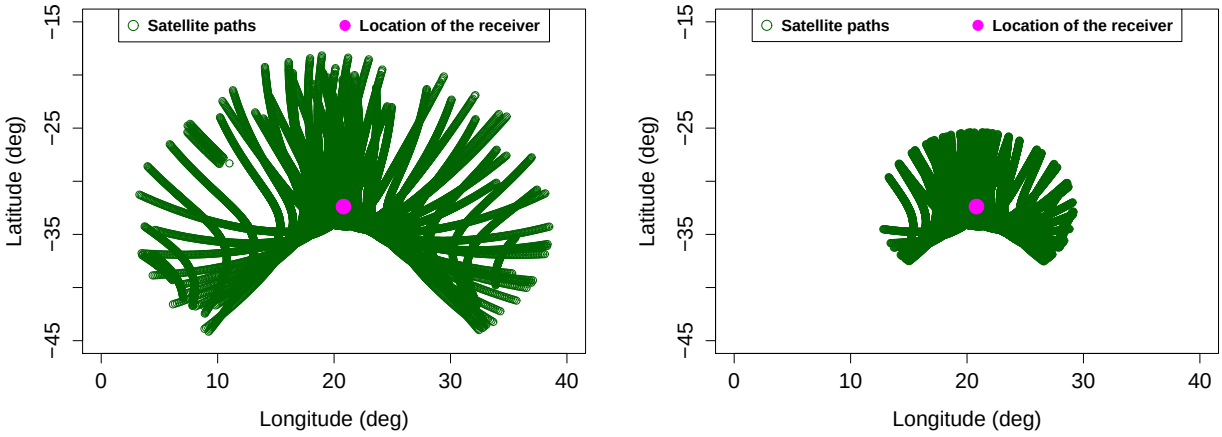
where Z' is the zenith angle of the satellite at the IPP and $(cosZ')^{-1}$ is the obliquity factor. By applying the law of sines, the relationship between the satellite zenith angle Z at the receiver position and Z' can be obtained:

$$\frac{sinZ}{R_E + h_I} = \frac{sinZ'}{R_E} \tag{3.26}$$

where $R_E = 6371$ km is the average radius of the Earth.

3.2 Data processing

The GPS TEC data used in this project was derived from GPS records at Sutherland (32.38^o S, 20.81^o E). Using the TEC analysis software developed at Boston College (Seemala, 2004), the data in RINEX (Receiver INdependent Exchange) format was processed to get both slant and vertical TEC. Since the minimum satellite elevation angle was 4^o, to remove the multipath effects, the satellite elevation angle of 20^o was chosen as a cut-off, and only TEC values corresponding to elevation angles greater than this cut-off was considered in the modelling techniques. Figures 3.4 (a) and (b) show the satellite paths before and after selecting elevation angles greater than 20^o, respectively.



(a) (b)

Figure 3.4: Satellite paths on 17-03-2015 over Sutherland (32.38^o S, 20.81^o E) (a) before selecting elevation angles greater than 20^o, and (b) after selecting elevation angles greater than 20^o.

The values of A index, which was used as one of the inputs to the model, were computed from K indices recorded at SANSA, Hermanus (34.4^o S, 19.2^o E), following the same steps as described in chapter 2. F10.7 and Dst indices were respectively downloaded from the following websites:

1. `http://omniweb.gsfc.nasa.gov/form/dx1.html`

2. `http://wdc.kugi.kyoto-u.ac.jp/dst_final/index.html`

3.3 Modelling techniques

In this section, the two methods that were applied to model TEC during storm conditions are presented. These are the regression and EOF analyses methods. The EOF analysis was used to decompose TEC data in terms of base functions and associated coefficients and to reveal some hidden information in the data. Regression analysis was used to estimate the relationship between the EOF coefficients and indices A and F10.7 used as inputs.

3.3.1 Regression analysis

Regression analysis is a statistical method for investigating relationships between one or more response variables (also called dependent variables, explained variables, predicted variables, or regressands) and several other variables called predictors (or independent variables, explanatory variables, control variables, or regressors) (Chatterjee and Hadi, 2013). Denoting the dependent variable by y and the set of independent variables by x_1, x_2, x_3, ..., x_p, where p denotes the number of independent variables, the main goals of regression analysis can be summarized as follows (Yan, 2009):

- Establish a relationship between the dependent variable y and independent variables x_1, x_2, x_3, ..., x_p.
- Predict the dependent variable y, based on the set of independent variables x_1, x_2, x_3, ..., x_p.
- Identify which variables among x_1, x_2, x_3, ..., x_p contribute more than others to explain the response variable y, so that the relationship between the dependent and independent variables can be determined more efficiently and accurately.

The relationship between the dependent and independent variables can mathematically be approximated by the expression

$$y = f(x_1, x_2, x_3, ..., x_p) + \varepsilon \tag{3.27}$$

where ε is the random error representing the discrepancy in the approximation (Yan, 2009; Chatterjee and Hadi, 2013). There are three types of regression: simple linear regression, multiple linear regression and non-linear regression (Yan, 2009).

- Simple linear regression is used for modelling a linear relationship between two variables: a dependent variable y and the independent variable x. It is stated in the form

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{3.28}$$

where β_0 is the y intercept, β_1 is the slope of the regression line and ε is the random error.

- Multiple linear regression is used for investigating a linear relationship between one dependent variable and more than one independent variables. The general form of multiple linear regression model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_p x_p + \varepsilon \tag{3.29}$$

where β_0, β_1, β_2, β_3, ..., β_p are regression coefficients and ε is the random error.

- Non-linear regression is used for investigating the relationship between dependent variable and independent variables, assuming that this relationship is not linear in regression parameters. An example of a non-linear regression model is

$$y = \frac{\alpha}{1 + e^{\beta t}} + \varepsilon \tag{3.30}$$

where α and β are the regression parameters and ε is the random error (Chatterjee and Hadi, 2013).

The regression parameters in Equations 3.27, 3.28, 3.29 and 3.30, are unknown constants and have to be estimated from the data. This is what is known as parameter estimation or model fitting. The most commonly used method to determine the unknown coefficients is the least squares method (Chatterjee and Hadi, 2013).

3.3.2 Empirical Orthogonal Function (EOF) analysis

The EOF analysis, also known as Principal Component Analysis (PCA) or Natural Orthogonal Component (NOC) algorithm, has been used by ionospheric researchers for modelling ionospheric parameters. Some of these parameters are foF2 (A *et al.*, 2011), TEC during magnetically quiet conditions (Mao *et al.*, 2005, 2008; A *et al.*, 2012), the ionospheric F2 peak height or hmF2 (Zhang *et al.*, 2009; Lin *et al.*, 2014), and the maximum usable frequency factor or M(3000)F2 (Liu *et al.*, 2008). The accuracy of EOF analysis in modelling these ionospheric parameters has been appreciated.

EOF analysis method is a mathematical method that consists of decomposing original dataset, consisting mainly of a number of multiple intercorrelated variables, into a new dataset of a small number of uncorrelated variables, using an orthogonal transformation. Hence, the method reduces the dimensionality of the data, which means that a large set of variables is reduced to a small set of new variables that still contains most of the information in the

original dataset. The EOF analysis is used to decompose the original dataset in terms of ordered base functions and their associated coefficients (A *et al.*, 2012).

3.3.2.1 Goals of EOF analysis

The main goal of EOF analysis is to reveal hidden structures in the data and classify them according to how much of the information stored in data they account for. EOF analysis also

- reduces the dimensionality of the data (Suhr, 2005; Liu *et al.*, 2008; Mankin, 2014)
- reduces redundancy in the data (Mankin, 2014)
- filters some of the noise in the data (Mankin, 2014)
- prepares data for other purposes (Mankin, 2014).

3.3.2.2 Mathematical description of EOF analysis

The original aim of EOF was to decompose a continuous space-time field $X(t, s)$ in terms of basis functions of space, $E_k(s)$ and expansion functions of time $A_k(t)$, in such a way that (Hannachi *et al.*, 2007)

$$X(t, s) = \sum_{k=1}^{p} A_k(t) \times E_k^T(s) \tag{3.31}$$

where t and s denote time and spatial position respectively, and p denotes the number of modes contained in the field. The superscript T over $E_k(s)$ denotes the transpose of the matrix $E_k(s)$. Matrix methods are usually used and the data to be analysed has to be arranged in a matrix format.

Suppose that measurements of a variable, at locations X_1, X_2, X_3, ..., X_p, were taken at times t_1, t_2, t_3, ..., t_n. The measured values can be arranged in a data matrix X of p columns corresponding to X_1, X_2, X_3, ..., X_p and n samples. The original data matrix X is then defined as

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & ... & X_{1p} \\ X_{21} & X_{22} & X_{23} & ... & X_{2p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ X_{n1} & X_{n2} & X_{n3} & ... & X_{np} \end{pmatrix}$$

Each of the p columns of X is interpreted as a time series for a given location and each of the n rows of X, as a map for a specific time. For example, column one represents a time series for location X_1, while the first row is a map for time t_1 (Björnsson and Venegas, 1997).

The matrix X can be decomposed in terms of basis functions E_k and associated coefficients A_k, with $k = 1, 2, ..., p$. The basis functions E_k are the eigenvectors of the covariance matrix Σ, constructed from the original data, and defined by (Smith, 2002; Yu, 2014)

$$\begin{pmatrix} cov(X_1, X_1) & cov(X_1, X_2) & ... & cov(X_1, X_p) \\ cov(X_2, X_1) & cov(X_2, X_2) & ... & cov(X_2, X_p) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ cov(X_p, X_1) & cov(X_p, X_2) & ... & cov(X_p, X_p) \end{pmatrix}$$

where the covariance between two variables X_i, X_j is defined by

$$cov(X_i, X_j) = \frac{1}{n} \sum_{l=1}^{n} (X_{li} - \overline{X}_i)(X_{lj} - \overline{X}_j) \tag{3.32}$$

The mean values $\overline{X}_i$ and $\overline{X}_j$ are defined by

$$\overline{X}_i = \frac{1}{n} \sum_{l=1}^{n} X_{li}$$

and

$$\overline{X}_j = \frac{1}{n} \sum_{l=1}^{n} X_{lj}$$

where X_{li} and X_{lj} are elements of the ith and jth columns, respectively. Note that the definition of the covariance matrix given above is not unique. Further definitions are found in other literature. For example, Weare and Nasstrom (1982); Björnsson and Venegas (1997); Xu and Kamide (2004); De Michelis *et al.* (2010) defined the covariance matrix as

$$\Sigma = X^T X \tag{3.33}$$

while Hannachi *et al.* (2007); Mao *et al.* (2008) defined the covariance matrix as

$$\Sigma = \frac{1}{n} X^T X \tag{3.34}$$

Another definition of the covariance matrix by Shlens (2003); Hannachi (2004); Mankin (2014)

is

$$\Sigma = \frac{1}{n-1} X^T X \tag{3.35}$$

However, the definition is not that critical, since basis functions and their time series will only differ by a constant factor (Björnsson and Venegas, 1997). The covariance matrix is a square symmetric matrix of dimension $p \times p$ and diagonal terms are just the variance of particular measurement types whereas the off-diagonal terms are the covariance between measurement types (Shlens, 2003).

In Algebra any symmetric matrix Σ of dimension $p \times p$ has a set of p orthonormal eigenvectors $(E_1, E_2, E_3, ..., E_p)$ and associated eigenvalues $(\lambda_1, \lambda_2, \lambda_3, ..., \lambda_p)$ (Mankin, 2014). The eigenvalues of Σ are found by solving the characteristic equation

$$|\Sigma - \lambda I| = 0 \tag{3.36}$$

where λ is a parameter and I is a $p \times p$ identity matrix, i.e. a matrix of which the elements are equal to 1 along the main diagonal, and zero elsewhere. 0 is a zero matrix, i.e. with all elements equal to zero. The Equation 3.36 is a polynomial equation of degree p in the parameter λ, and has p roots λ_i, $i = 1, 2, 3, ..., p$ which are the eigenvalues of Σ. For each λ_i there is a corresponding column vector E_i, the ith eigenvector of Σ, given by equation (Riley *et al.*, 2006)

$$\Sigma E_i = \lambda_i E_i \tag{3.37}$$

The eigenvectors are the EOF base functions required. The matrix of the eigenvectors, denoted by E, satisfies the following property:

$$E^T E = E E^T = I \tag{3.38}$$

This means that the EOF base functions are uncorrelated over space, or simply that the eigenvectors are orthogonal to each other; hence, the appellation Empirical Orthogonal Functions (Björnsson and Venegas, 1997). The elements of an eigenvector are the weights e_{ij} and are also known as loadings (Holland, 2008). Eigenvectors are the directions where the data is most spread out. The eigenvector E_1 corresponds to the highest eigenvalue and represents the direction of the greatest variation. The second eigenvector, E_2, corresponds to the second largest eigenvalue and is the direction with the next highest variation, and so on (Björnsson and Venegas, 1997). The eigenvalues of the covariance matrix indicate the fraction of the total variation in the original data explained by each individual EOF component. The quantity

$$r_i = 100 \times \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}\% \tag{3.39}$$

which represents the ratio of ith eigenvalue to the sum of all eigenvalues is interpreted as the proportion of the total variation in the original data explained by the ith EOF component (Hannachi, 2004), while the quantity

$$\rho_k = 100 \times \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{p} \lambda_j}\% \tag{3.40}$$

is interpreted as the proportion of the total variation in the original data explained by the first k EOF components (Zhang *et al.*, 2009).

The next task is to determine the expansion coefficients A_k. Once the eigenvectors of the covariance matrix are known, the original data can be decomposed in a matrix product as follow:

$$X = AE^T \tag{3.41}$$

where E represents the $p \times p$ matrix of EOF base functions E_k (k = 1, 2, 3, ..., p) as columns, and A represents the $n \times p$ matrix of A_k coefficients as columns (k = 1, 2, 3, ..., p). X of dimension $n \times p$ represents the matrix of the original data. By multiplying both sides of Equation 3.41 by $[E^T]^{-1}$, taking into account the definition of the orthogonal matrix as specified by Equation 3.37, the matrix A of A_k coefficients is calculated as

$$A = XE \tag{3.42}$$

In general, for each E_k one can determine the corresponding A_k, by projecting the data onto E_k:

$$A_k = XE_k \tag{3.43}$$

The A_k coefficients are called the principal components (PCs) or expansion coefficients of EOF base functions, and are uncorrelated in time (Björnsson and Venegas, 1997; Zhang *et al.*, 2009).

One of the goals of the EOF analysis/PCA is to reduce the dimensionality of the data and deal with a small number of variables. Therefore, several criteria have been proposed for determining how many PCs should be retained or ignored when using EOF analysis (Holland, 2008; Bremner, 2009).

- Determine the number of PCs to be retained by ignoring the PCs of which the variance

explained is less than 1 when a correlation matrix is used or less than the average variance explained when a covariance matrix is used (Holland, 2008).

- The scree test is based on the scree plot which is a graph of variance as a function of principal component rank (or eigenvalues against the corresponding PC). By looking at a scree plot, the position of an "elbow" (the point at which the curve bends) indicates the number of PCs to be retained.

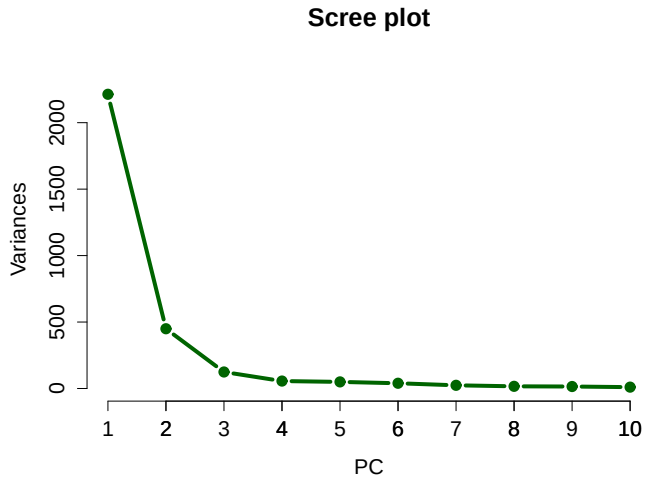


Figure 3.5: Example of a scree plot

Figure 3.5 shows that the elbow in the graph is between 2 and 3. Then, three PCs (the maximum number between 2 and 3) can be retained. It is sometimes difficulty to specify where exactly the position of an elbow is and therefore, the scree test is not considered as a good method for deciding how many PCs should be retained (Bremner, 2009).

- Based on the cumulative proportion of variance, one can retain the PCs which account for a specified proportion of the total variance. For example, if one is satisfied with 95 % of the total variance, then he should use the number of PCs that account for the desired percentage. Although the neglected components account for some percentage of the total variance in the data, they may not be significant enough to consider (Holland, 2008; Bremner, 2009).

- Ignore the last PCs of which the variances explained are all roughly equal (Holland, 2008).

3.4 Summary

In this chapter, some basics of GPS satellites are offered, including GPS architecture and functioning. The effect of the ionosphere on GPS signals is also explained. Due to the ionosphere, the fact that GPS signals delays are proportional to TEC encountered along the signal path from a GPS satellite to a receiver has been utilized to show how TEC can be derived from GPS measurements. The sources and processing of the data were described, as well as the mathematical background on the two modelling methods used in this project. The regression analysis method estimates relationship between a dependent variable and several independent variables and regression parameters are usually determined using the method of least squares. EOF analysis consists of decomposing data in terms of base functions and expansion coefficients. The base functions are the eigenvectors of the covariance matrix obtained from the original data and the expansion coefficients are found by projecting the data onto the base functions. Discovering hidden information in data and the reduction of data dimensionality are primary applications of the EOF analysis. Details about the application of the EOF analysis to TEC decomposition and modelling are presented in the next chapter.

Chapter 4

Results and discussion

This chapter presents results and discussion of TEC modelling during storm conditions. Subsequent to the development of equations for estimating TEC during storm conditions by means of EOF and regression analysis, a comparison of the estimated and the observed TEC, for a number of selected storm periods, was made in order to test the model's performance. Statistical analysis methods were used to evaluate the accuracy of the model. Moreover, the model was tested using TEC over different stations surrounding Sutherland (32.38^o S, 20.81^o E) and its performance was also evaluated.

4.1 Model construction

The TEC modelling technique used in this project is based on two mathematical methods. EOF analysis was used to decompose TEC data into the EOF base functions $E_k(h)$ which represent the diurnal variation of TEC with time, and the associated coefficients $A_k(d)$ which vary with day number of the year, and represent the long-term variation of TEC, i.e., annual, semiannual, seasonal and solar cycle variations. After TEC decomposition, the EOF coefficients $A_k(d)$ were then modelled in terms of A and F10.7 indices, using regression analysis, in order to include geomagnetic and solar activities in the model.

4.1.1 TEC decomposition using EOF analysis

To decompose TEC into EOF base functions and associated coefficients, hourly TEC values over Sutherland (32.38^o S, 20.81^o E), for storm days from 1999 to 2013, were first arranged in a matrix of dimension 865 $\times$ 24 as shown in Figure 4.1. Between 1999 and 2013 there were 257 storm periods with available TEC data and these consisted of 865 storm days. Therefore, 865 rows and 24 columns correspond to 865 storm days and 24 hourly TEC values respectively. Out of 257 storm periods, based on the storm classification by Loewe and Prölss (1997), moderate, strong, severe and great storms were represented by 76.7 %, 19.1 %, 2.3 % and 1.9 % respectively. Note that TEC values that correspond to the storm days of 2000 and 2006 for which the model was tested, were excluded from the matrix of observed TEC. The choice of storm days within a storm period depended on the nature of the storm. Typically,

for a storm which occurred on a particular day, three days of TEC variation were considered, i.e. including one day before and after the storm day.

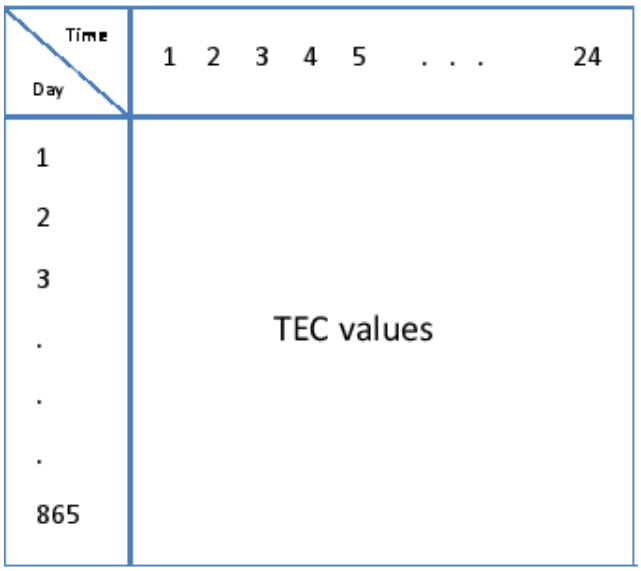


Figure 4.1: The 865 × 24 matrix of hourly TEC values.

The hourly TEC values were then decomposed into EOF base functions E_k and associated coefficients A_k, according to the equation

$$TEC(h,d) = \sum_{k=1}^{24} A_k(d) \times E_k^T(h) \tag{4.1}$$

Here, the $A_k(d)$ coefficient is defined as a 865 × 1 matrix while $E_k(h)$ is a 24 × 1 matrix. $E_k^T(h)$ is the transpose of the matrix $E_k(h)$. Since $E_k^T(h)$ is a 1 × 24 matrix, the product $A_k(d) \times E_k^T(h)$ gives a 865 × 24 matrix, which has the same dimension as the matrix of the observed TEC. The latter is reconstructed by summing 24 terms, as shown in Equation 4.1. Thus, $TEC(h,d)$ represents the hourly TEC values of the observed data expressed as a 865 × 24 matrix, d represents the storm days and h represents UT hour (h = 1, 2, 3, ..., 24). $E_k(h)$ is the k^{th} EOF base function of $TEC(h,d)$ and represents diurnal variation, while $A_k(d)$ is the k^{th} EOF coefficient and reflects long-term variation (solar cycle, annual, semiannual, seasonal) of TEC.

As described in chapter 3, the covariance matrix was computed from the matrix of the observed TEC. This is a 24 × 24 matrix since there are 24 variables which are represented by 24 columns of hourly TEC values. From the covariance matrix, the eigenvalues of the the covariance matrix were obtained. The EOF base functions are then the eigenvectors of the covariance matrix. Theoretically, the original dataset (a 865 × 24 matrix) should be expanded into 24 EOF base functions and associated coefficients to represent the total variation of the

observed TEC. However, due to the quick convergence of the EOF method, only twelve components that account for 99.179 % of the total variation in the original dataset were retained. The modelling of storm-time TEC variability is complicated which is the major reason why twelve components are required for convergence, otherwise previous results have shown that convergence is achieved even after only 3 - 4 components for quiet time foF2, TEC, hmF2 and M(3000)F2 variability (Mao *et al.*, 2005; Liu *et al.*, 2008; Mao *et al.*, 2008; Zhang *et al.*, 2009; A *et al.*, 2011, 2012; Lin *et al.*, 2014). Table 4.1 presents the variances explained by 24 EOF components. The eigenvalue of the covariance matrix is a number which indicates how much variance there is in the data, in the direction of the corresponding eigenvector. Then, what is presented as standard deviations in Table 4.1, are the square roots of the eigenvalues of the covariance matrix.

Table 4.1: Proportion of variance accounted for by each EOF component.

EOF component	Standard deviation	Variances (%)	Cumulative variances (%)
$A_1 \times E_1^T$	47.0498	72.860	72.860
$A_2 \times E_2^T$	21.2023	14.790	87.650
$A_3 \times E_3^T$	11.1475	4.090	91.740
$A_4 \times E_4^T$	7.48319	1.843	93.585
$A_5 \times E_5^T$	7.04793	1.635	95.219
$A_6 \times E_6^T$	6.27843	1.297	96.517
$A_7 \times E_7^T$	4.90091	0.791	97.307
$A_8 \times E_8^T$	4.04910	0.540	97.850
$A_9 \times E_9^T$	3.85219	0.488	98.335
$A_{10} \times E_{10}^T$	3.20430	0.338	98.673
$A_{11} \times E_{11}^T$	2.89474	0.276	98.949
$A_{12} \times E_{12}^T$	2.64644	0.231	99.179
$A_{13} \times E_{13}^T$	2.33150	0.179	99.358
$A_{14} \times E_{14}^T$	1.93128	0.123	99.481
$A_{15} \times E_{15}^T$	1.80190	0.107	99.588
$A_{16} \times E_{16}^T$	1.66369	0.091	99.679
$A_{17} \times E_{17}^T$	1.60300	0.085	99.764
$A_{18} \times E_{18}^T$	1.39172	0.064	99.827
$A_{19} \times E_{19}^T$	1.19576	0.047	99.874
$A_{20} \times E_{20}^T$	1.10360	0.040	99.910
$A_{21} \times E_{21}^T$	1.06291	0.037	99.952
$A_{22} \times E_{22}^T$	0.94750	0.030	99.980
$A_{23} \times E_{23}^T$	0.61897	0.013	99.994
$A_{24} \times E_{24}^T$	0.42997	0.006	100.00

Table 4.1 shows that the first and the second EOF components account for 72.860 % and 14.79% respectively of the total variance in the original TEC data, while combined, they explain 87.650 % of the total variance in the original TEC data. Figure 4.2 shows the diurnal variation of 12 EOF base functions. It is clear that the two first EOF base functions are

characterised by trends which are similar to the general diurnal variation trend of TEC.

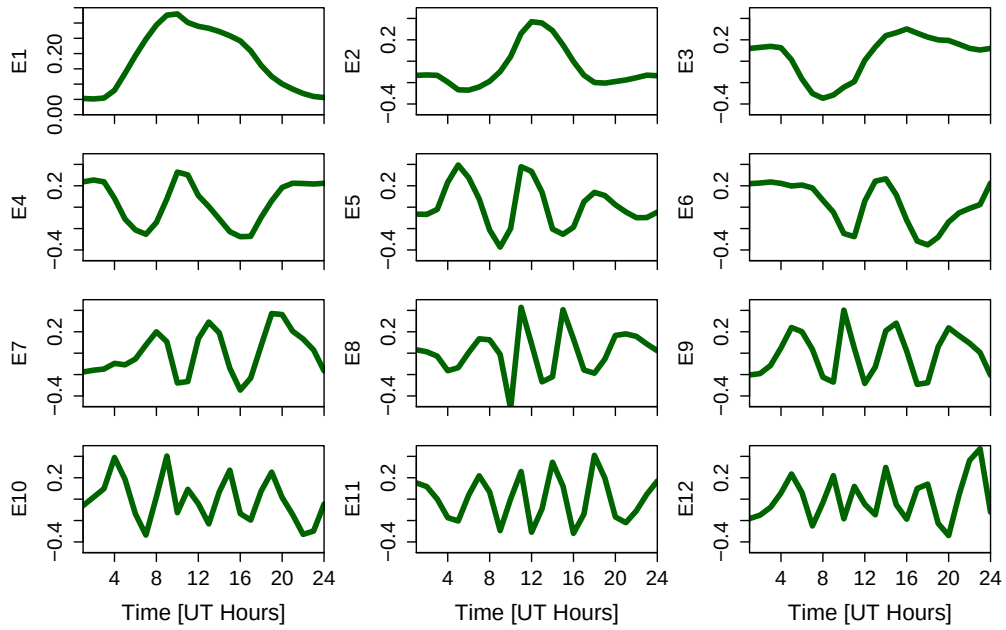


Figure 4.2: Diurnal variation of the first twelve EOF base functions.

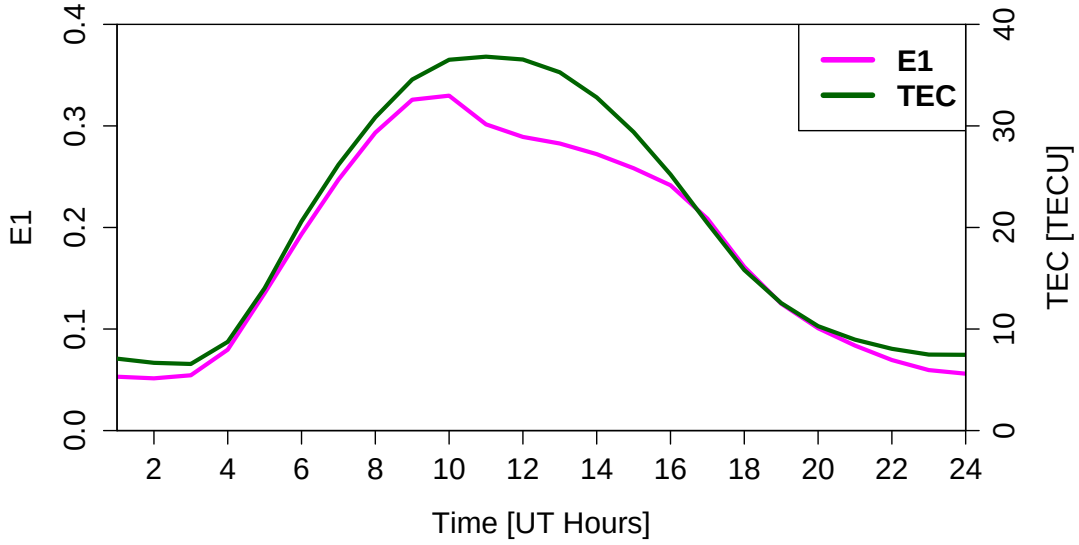


Figure 4.3: Diurnal variation of the first EOF base function and the average TEC.

The first EOF base function is shown in Figure 4.3 together with the average diurnal variation of TEC. The average TEC was calculated over the 865 storm days. For example, to get

the average TEC at time 01:00 UT, we took the sum of all the 865 TEC values at 01:00 UT divided by 865. It is clearly seen that the diurnal variation of the first order base function E_1 and diurnal variation of the average TEC are quite similar, and a high correlation coefficient of 0.9821 was found. Thus, the first order base function E_1 represents the average diurnal variation trend of TEC. A similar observation was also reported in other modelling works of ionospheric parameters by means of EOF analysis (Mao *et al.*, 2005; Liu *et al.*, 2008; A *et al.*, 2011, 2012; Lin *et al.*, 2014). For example, when modelling foF2 with EOF analysis, it was shown that the first order base function E_1 represents the average diurnal variation trend of foF2 (A *et al.*, 2011). In modelling the ionospheric F2 peak height (hmF2) with EOF analysis, Lin *et al.* (2014) found that the first order base function E_1 shows a diurnal variation of hmF2. A similar result was reported by Liu *et al.* (2008) when modelling M(3000)F2, and by A *et al.* (2012); Mao *et al.* (2005) when modelling TEC. The modelling of these two paremeters has demonstrated that the first order base function E_1 represents, on the one hand, the mean diurnal variation of M(3000)F2, and on the other hand, the average diurnal TEC.

Moreover, the diurnal variation of the second order base function E_2 tends towards a trend of diurnal variation slightly similar to the one represented by E_1. The correlation coefficient between E_2 and the average diurnal TEC was found to be 0.6897. However, the high order EOF base functions (from E_3 to E_{12}) represent mainly the short-term variations of TEC, as shown in Figure 4.2, and they are less correlated with the average diurnal TEC. A similar observation was reported by Mao *et al.* (2005). It is important to point out that previous studies employed EOF analysis largely during quiet conditions (Mao *et al.*, 2005; Liu *et al.*, 2008; Mao *et al.*, 2008; Zhang *et al.*, 2009; A *et al.*, 2011, 2012; Lin *et al.*, 2014), while this study attempted to investigate the possibility of using the EOF method on TEC data during disturbed conditions. The remaining EOF base functions may also represent short-term variations of TEC as well as noise effects, and redundancy in data can be decreased by ignoring them.

Once the EOF base functions were determined, Equation 3.43 was used to determine the associated EOF coefficients A_k. Figure 4.4 represents the variation of twelve A_k coefficients for the period 1999 - 2013. Figure 4.4 shows that the long-term variation of the EOF coefficients is characterised by high amplitudes during the solar maximum period (around 2000 - 2003 and 2011 - 2013) and small amplitudes during the solar minimum (around 2005 - 2010). This demonstrates that the EOF coefficients relate to the solar activity. The gap in 2009 is due to missing TEC data for a single storm that occurred on 22-07-2009.

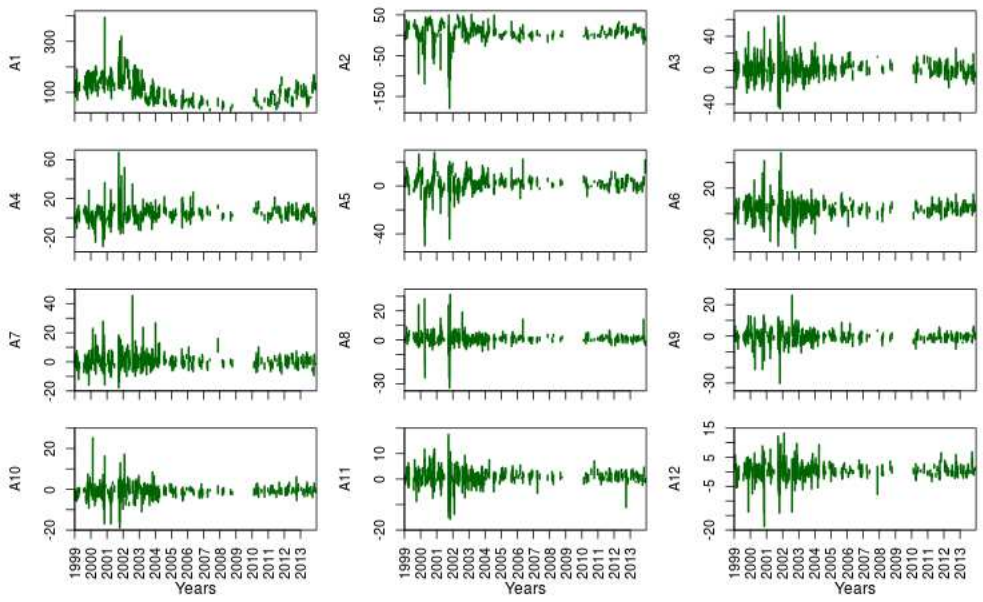


Figure 4.4: Long-term variations of the EOF coefficients A_k.

Figure 4.5 shows the variation of the first EOF coefficient A_1 and F10.7 index, for the period 1999 - 2013. It is clear that the trend of A_1 is similar to that of F10.7, which represents the general level of solar activity. The correlation coefficient between A_1 and F10.7 index was found to be 0.7042. Therefore, the first EOF component $A_1 \times E_1^T$ which explains 72.860 % of the variance of the whole TEC dataset, contains the first order EOF base function which represents the average diurnal variation of TEC, and the first EOF coefficient, which represents long-term variation of TEC. The highest percentage explained by the first EOF component $A_1 \times E_1^T$ justifies the view that the dominant factor that controls TEC variability is solar activity. A similar observation was reported by Mao *et al.* (2008) during TEC modelling over China.

4.1.2 Modelling the EOF coefficients $A_k(d)$ using the regression analysis

After TEC decomposition into base functions $E_k(h)$ and associated coefficients $A_k(d)$, the coefficients $A_k(d)$ need to be modelled in terms of A and F10.7 indices, in order to introduce the influence of solar and geomagnetic activities on TEC into the model. For this purpose, the coefficients $A_k(d)$ were modelled by using the formal Fourier series according to the equation

$$A_k(d) = B_{k1}(d) + B_{k2}(d) + B_{k3}(d) + \epsilon \tag{4.2}$$

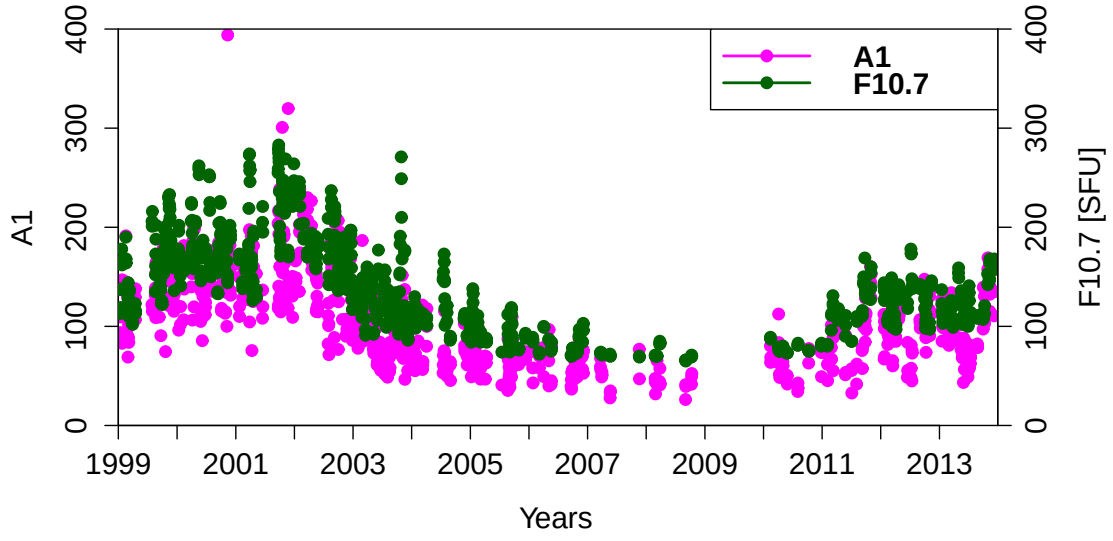


Figure 4.5: Variations of A_1 and F10.7 index for the storm periods of 1999 - 2013.

where ϵ is the model residual error and k = 1, 2, 3, ..., 12, since we need twelve EOF coefficients. Equation 4.2 is a short form of the following equation

$$\begin{aligned} A_k(d) = {} & C_{k1} + D_{k1}.F10.7(d) + E_{k1}.A(d) \\ & + [C_{k2} + D_{k2}.F10.7(d) + E_{k2}.A(d)]cos\left(\frac{2\pi d}{365.25}\right) \\ & + [F_{k2} + G_{k2}.F10.7(d) + H_{k2}.A(d)]sin\left(\frac{2\pi d}{365.25}\right) \\ & + [C_{k3} + D_{k3}.F10.7(d) + E_{k3}.A(d)]cos\left(\frac{4\pi d}{365.25}\right) \\ & + [F_{k2} + G_{k2}.F10.7(d) + H_{k2}.A(d)]sin\left(\frac{4\pi d}{365.25}\right) \end{aligned} \tag{4.3}$$

where the factor 0.25 stands for leap years. When comparing Equations 4.2 and 4.3, the coefficients $B_{k1}(d)$, $B_{k2}(d)$ and $B_{k3}(d)$ are

$$
\begin{aligned}
B_{k1}(d) &= C_{k1} + D_{k1}.F10.7(d) + E_{k1}.A(d) \qquad (4.4)\\
B_{k2}(d) &= [C_{k2} + D_{k2}.F10.7(d) + E_{k2}.A(d)]cos\left(\frac{2\pi d}{365.25}\right)\\
&\quad + [F_{k2} + G_{k2}.F10.7(d) + H_{k2}.A(d)]sin\left(\frac{2\pi d}{365.25}\right)\\
B_{k3}(d) &= [C_{k3} + D_{k3}.F10.7(d) + E_{k3}.A(d)]cos\left(\frac{4\pi d}{365.25}\right)\\
&\quad + [F_{k3} + G_{k3}.F10.7(d) + H_{k3}.A(d)]sin\left(\frac{4\pi d}{365.25}\right)
\end{aligned}
$$

$B_{k1}(d)$ represents a linear function of the F10.7 and A indices, $B_{k2}(d)$ and $B_{k3}(d)$ are harmonic functions with periods of one year and a half year respectively, and amplitudes expressed as linear functions of F10.7 and A indices. Thus, $B_{k1}(d)$, $B_{k2}(d)$ and $B_{k3}(d)$ correspond to the solar cycle, annual, and semiannual variation components in EOF coefficients, respectively. The unknown coefficients C_{k1}, D_{k1}, E_{k1}, ..., F_{k3}, G_{k3} and H_{k3}, in the equations above, were determined by using the linear regression method. However, to reconstruct TEC for all 865 storm days, F10.7 and A can be used as inputs to estimate the EOF coefficients, and then combine the latter with the EOF base functions to get TEC, using Equation 4.1. In particular, for a specific day d, the F10.7 and A indices for day d were used as inputs to estimate the coefficients $A_k(d)$. These were combined with the EOF base functions E_k according to Equation 4.1 to get TEC for that specific day. All the TEC values for storm days for which the model was tested, were obtained by means of this procedure.

4.2 Application of EOF model to selected storms

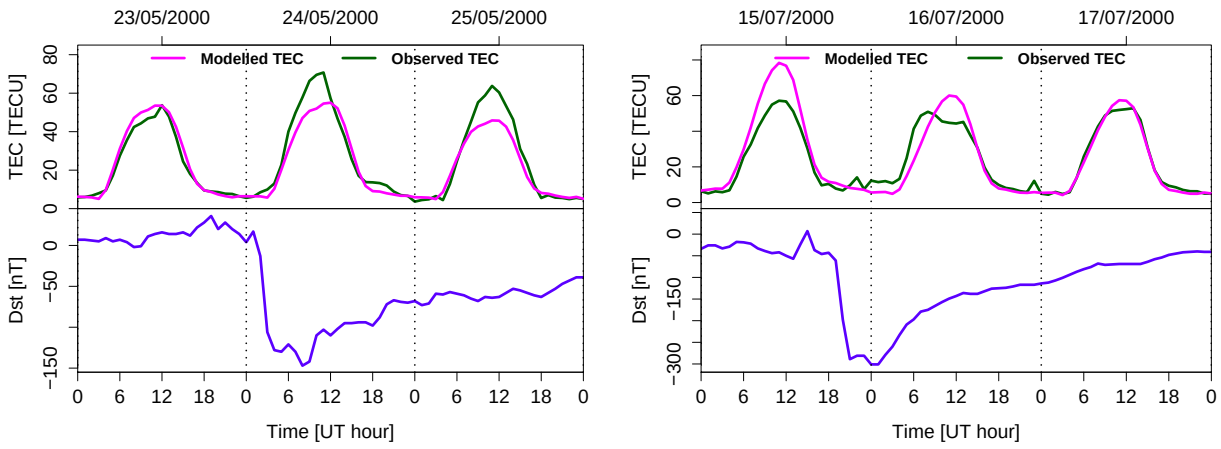
To test the EOF model, specific storms were chosen based on their strength as defined by the minimum value of the Dst index, the storm criterion being Dst $\leq$ -50 nT, and on the solar activity period during which the storm occurred. Three storms were chosen in 2000 and three others in 2006 to represent the solar maximum and solar minimum periods, respectively. TEC values corresponding to the storms selected for verification were excluded from the original dataset of TEC used to build the model, in order to make the model validation independent of data. Since the selected 2000 and 2006 storm periods fell within the database period (1999 - 2013) used for model construction, this validation can also be referred to as interpolation. For extrapolation, the EOF model was used for modelling storms that occurred during 2014 and 2015. Three storms were chosen for 2014 and 2015, based on the available data, and on the provisional Dst index (`http://wdc.kugi.kyoto-u.ac.jp/dst_provisional/index.html`).

4.2.1 TEC modelling during the storms of 2000 and 2006 (Interpolation)

Figure 4.6 shows the observed and modelled TEC variation for selected storms of 2000. For each storm period, the Dst index indicating the strength of the storm is also presented. Figure 4.6 (a) shows the observed and the modelled TEC variation during a strong storm that occurred on 24-05-2000, with a minimum Dst index of about -147 nT. The EOF model is capturing well the TEC magnitude before the storm main phase, during the nighttime and in the early morning, but underestimating TEC during the main and recovery phases. Figure 4.6 (b) shows the observed and modelled TEC variation during a severe storm that occurred between 15 - 17 July 2000, characterised by a minimum Dst of about -301 nT. During the SSC, the EOF model is overestimating TEC while it slightly follows the negative TEC response observed during the main phase. Although the EOF model is not capturing the TEC dynamics of the observed data, it is estimating quite well TEC magnitude during the main and recovery phases.
Figure 4.6 (c) shows the observed and modelled TEC variation during two consecutive storms: strong and severe storms that occurred on 11-08-2000 and 12-08-2000, with minimum Dst values of -106 nT and -235 nT respectively. For both storms, the EOF model is underestimating TEC magnitude but a slight positive response predicted by the model was observed for the storm of 12-08-2000.

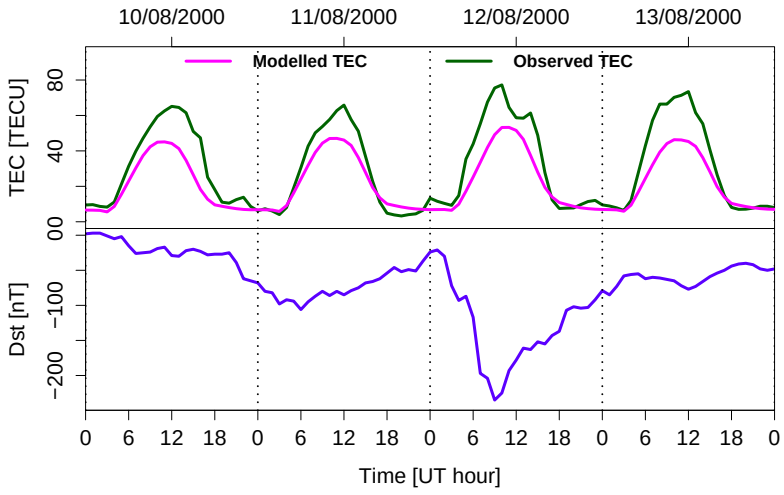
Briefly, for the storms that occurred during the high solar activity period, the EOF model predicts TEC well during the nighttime and early in the morning, while it doesn't perform very well during positive and negative TEC responses.

Figures 4.7 (a) and (b) show the observed and modelled TEC during two moderate storms that occurred on 09-04-2006 and 14-04-2006 with minimum Dst values of -82 nT and -98 nT respectively. For both storms, the EOF model is overestimating TEC during the daytime, except on 09-04-2006. However, the EOF model predicts TEC quite well during the nighttime. It is clear that short-term features in the observed TEC are not captured by the EOF model. Figure 4.7 (c) illustrates the observed and the predicted TEC during a strong storm that occurred on 15-12-2006, with minimum Dst value of -162 nT. Although there is an overestimation of TEC, the EOF model attempts to capture short-term features in TEC variability. The EOF model is empirical in nature and is not expected to perform very well for periods which were not well represented in model development. As Figure 4.5 demonstrates, there was little data for storm periods used in developing the EOF model during low solar activity period.



(a)

(b)



(c)

Figure 4.6: Comparison between observed and modelled TEC for storms that occurred during the high solar activity period on (a) 24-05-2000 (b) 16-07-2000 and (c) 11-12-08-2000

To summarize, the EOF model is overestimating TEC for storms that occurred during 2006, during the low solar activity period. Short-term features are captured only for a strong storm, while the EOF model predicts the TEC magnitude well only for the nighttime.

4.2.2 TEC modelling during the storms of 2014 and 2015 (Extrapolation)

Figure 4.8 shows the observed and modelled TEC during selected storms of 2014. Except for the storm that occurred on 19-02-2014 which is classified as a strong storm with a minimum Dst of -112 nT, all other storms that occurred in 2014 are moderate and most of them do not present any significant TEC response. In addition to the storm of 19-02-2014, three other

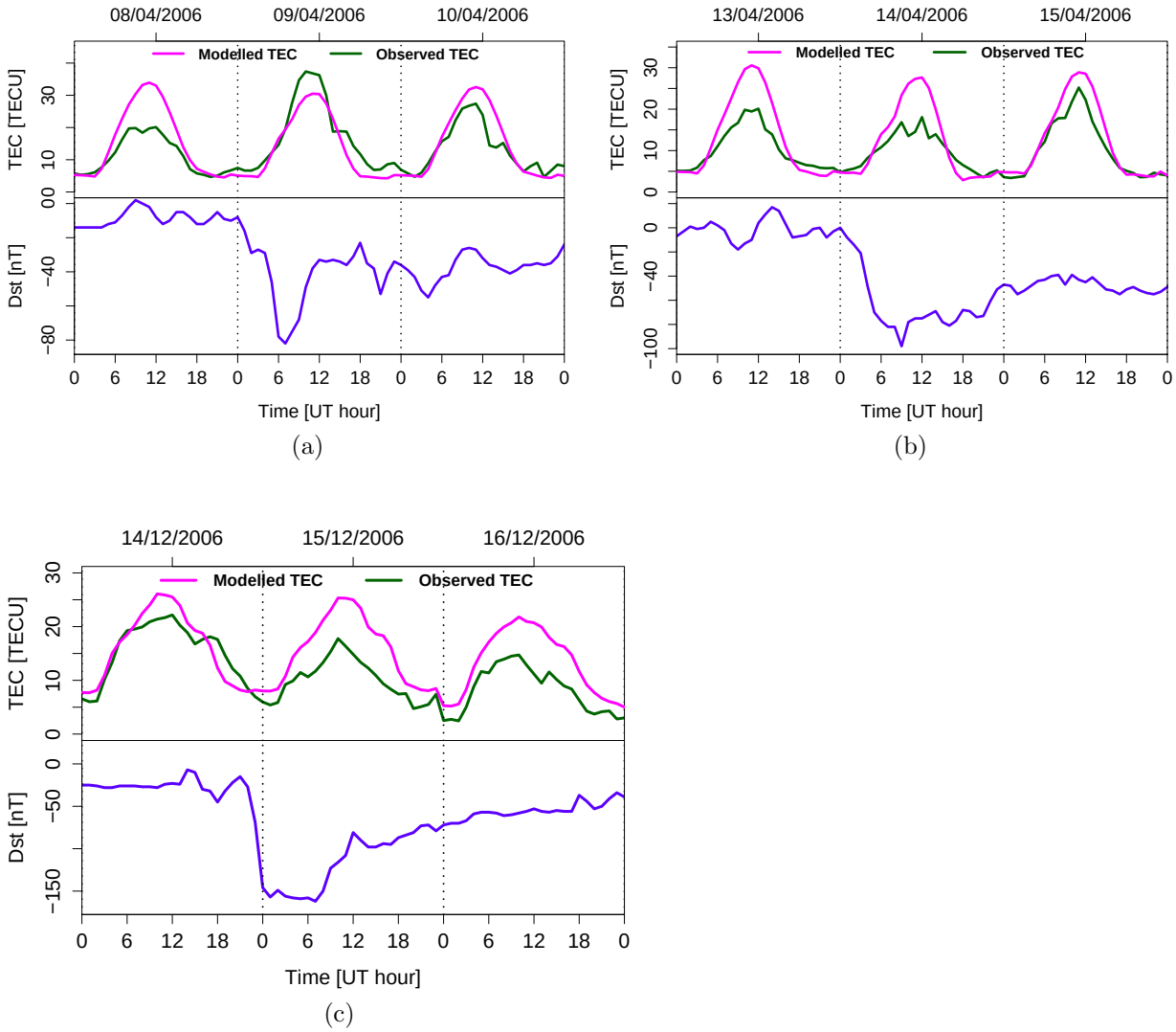
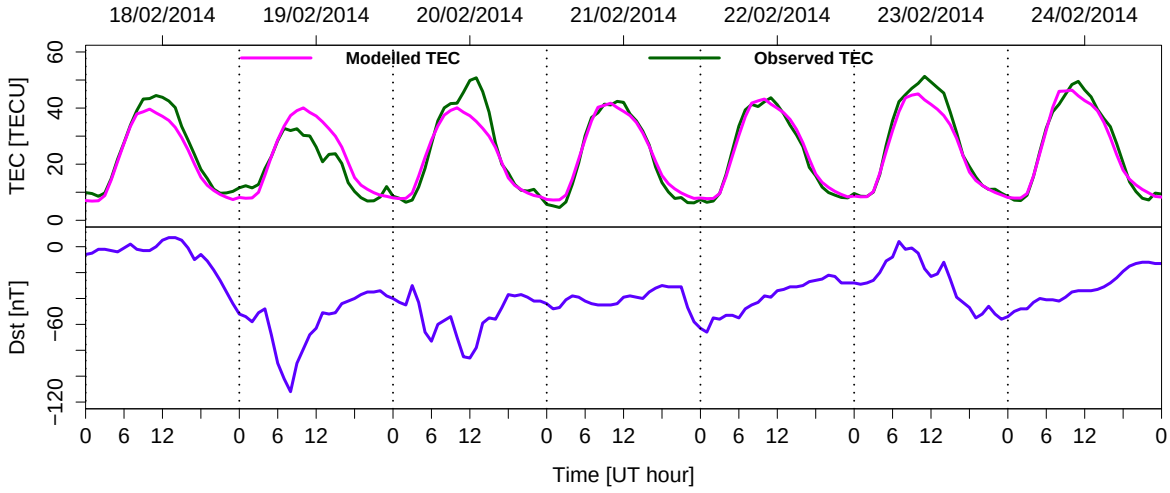


Figure 4.7: Comparison between observed and modelled TEC for storms that occurred during the low solar activity period on (a) 09-04-2006 (b) 14-04-2006 and (c) 15-12-2006

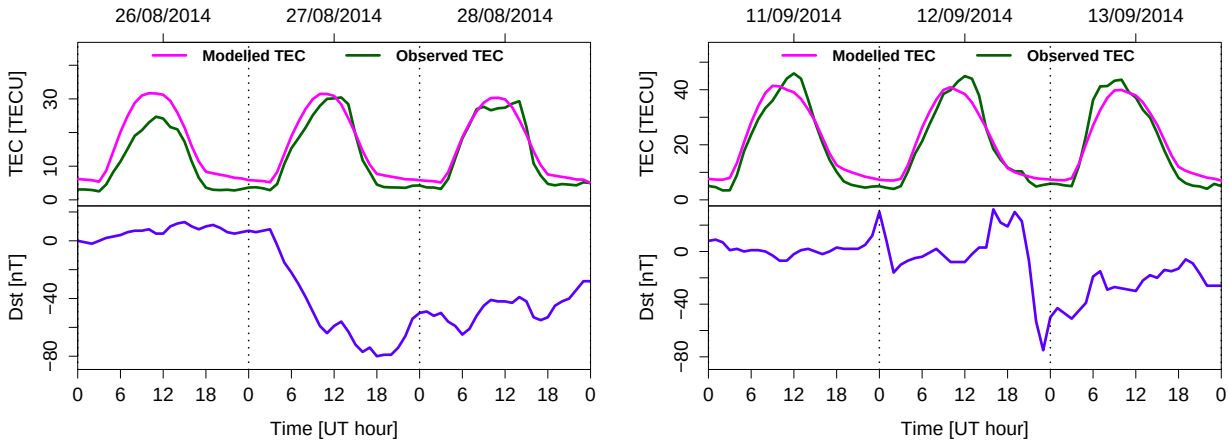
storms selected in February 2014 occurred on 20-02-2014, 22-02-2014 and 23-02-2014, and had provisional minimum Dst values of -86 nT, -66 nT and -56 nT respectively. Furthermore, two storms occurred on 27-08-2014 and 12-09-2014 with provisional minimum Dst values of -80 nT and -75 nT respectively, were also selected. For all these storms, the EOF model predicts the observed TEC well, although there is a small discrepancy between the modelled and observed TEC values.

Figure 4.9 shows observed and modelled TEC during the storms of 2015. The 2015 storms included a moderate (04-01-2015), strong (07-01-2015), and severe storm (17-03-2015) with provisional minimum Dst values of -75 nT, -105 nT and -223 nT respectively. The 17-03-2015 storm caused a positive followed by a negative TEC response and exhibited a long recovery

phase of about four days. Figure 4.9 (a) shows that the EOF model captures quite well short-term features in TEC variation as well as TEC magnitude during the main phase of the 04-01-2015 storm. However the model is overestimating TEC for other days.



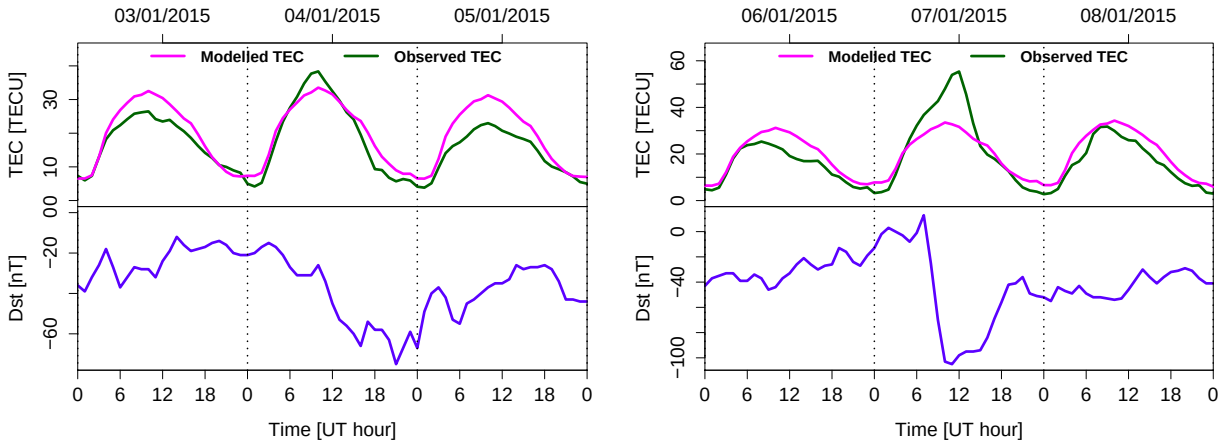
(a)



(b) (c)

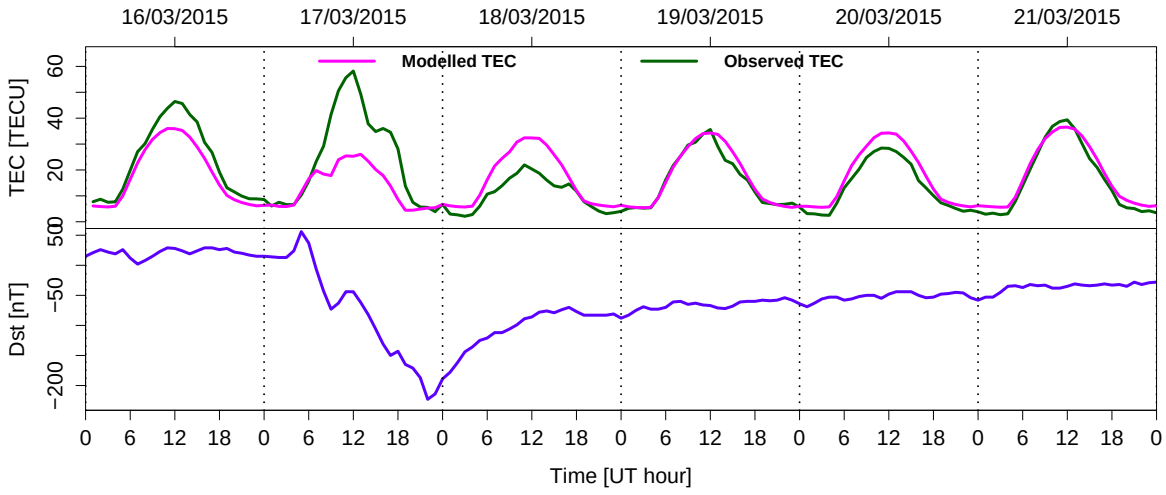
Figure 4.8: Comparison between observed and modelled TEC for storms that occurred (a) February 2014 (b) 27-08-2014 and (c) 12-09-2014

Figure 4.9 (b) shows that, although the model estimates quite well TEC magnitude on the 6^{th} and the 8^{th} of January, 2015, it fails to capture the positive TEC response observed on 07-01-2015. Figure 4.9 (c) shows that the EOF model also fails to capture the TEC enhancement and depression observed on 17-03-2015 and on 18-03-2015 respectively. However, during the recovery phase from 19-03-2015, the EOF is predicting TEC magnitude well. In summary, the EOF doesn't perform well in predicting TEC enhancement and depression for storms in 2015, but it predicts TEC quite well for the remaining storm period.



(a)

(b)



(c)

Figure 4.9: Comparison between observed and modelled TEC for storms that occurred (a) 04-01-2015 (b) 07-01-2015 and (c) 17-03-2015

According to the observations above, the EOF performs well for storms with a non-significant TEC response and during the nighttime. However, the EOF model sometimes fails to capture short-term features observed in TEC. The daily indices A and F10.7 may be the main cause of the failure of the EOF model to follow TEC dynamics since these indices represent the average level of geomagnetic and solar activities over a day, while geomagnetic storm features are short-term varying phenomena.

4.3 Statistical analysis

To evaluate the accuracy of the EOF model for TEC estimation during storm conditions, the root mean square error (RMSE) for each storm was computed, as well as the correla-

tion coefficient between the observed and modelled TEC. For TEC modelling, the RMSE is a measure of the spread between the observed and modelled TEC in such a way that the smaller the RMSE, the better the model. However, when interpreting the RMSE values, other considerations such as solar activity periods and the level of ionospheric disturbance which are correlated to TEC variability, should be taken into account.

For N different observations/predictions, the RMSE for each storm was computed with the formula:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(TEC_{mod} - TEC_{obs})^2} \tag{4.5}$$

where TEC_{mod} and TEC_{obs} represent the modelled and observed TEC respectively. The RMSE values of the twelve selected storms are presented in Figure 4.10.

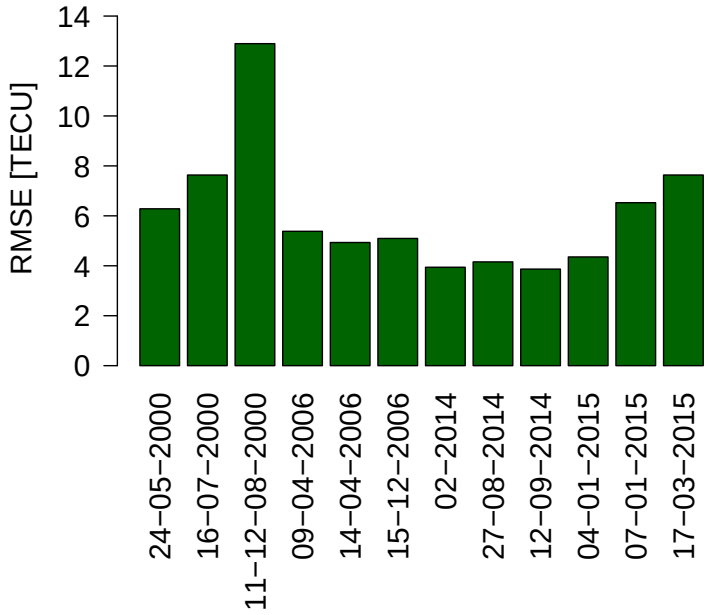


Figure 4.10: RMSE values for selected storm periods. The RMSE of each storm period is shown by a bar that corresponds to the storm date.

Figure 4.10 shows that RMSE values are small for storms that occurred during the solar minimum, i.e. in 2006, and for storms with no significant TEC response. These are specifically the storms which occurred in 2014. However, RMSE values obtained during the solar maximum, i.e. in 2000, and in 2015 when storms with a high significant TEC response were observed, are high. The small RMSE values indicate that on average, the EOF model performs better for storms with no significant TEC response, where deviations between the

observed and modelled TEC are smaller than elsewhere.

Correlation coefficients between the observed and modelled TEC for selected storm periods, were computed to show the relationship between the two variables. Correlation coefficients specify the degree to which the observed and modelled TEC tend to move together. For each storm period, the correlation coefficient was calculated with the formula:

$$R = \frac{cov(TEC_{mod}, TEC_{obs})}{\sigma_{mod}\sigma_{obs}} = \frac{\sum_{i=1}^{n}(TEC_{mod_i} - \overline{TEC}_{mod})(TEC_{obs_i} - \overline{TEC}_{obs})}{\sqrt{\sum_{i=1}^{n}(TEC_{mod_i} - \overline{TEC}_{mod})^2}\sqrt{\sum_{i=1}^{n}(TEC_{obs_i} - \overline{TEC}_{obs})^2}} \tag{4.6}$$

where $cov(TEC_{mod}, TEC_{obs})$ is the covariance between the observed and modelled TEC, and σ_{mod} and σ_{obs} are the standard deviations of the modelled and observed TEC respectively. TEC_{mod_i} and TEC_{obs_i} are the i^{th} modelled and observed TEC while $\overline{TEC}_{mod}$ and $\overline{TEC}_{obs}$ are the average values of the modelled and observed TEC, respectively. The correlations between the observed and modelled TEC for twelve storm periods are shown in Figure 4.11.

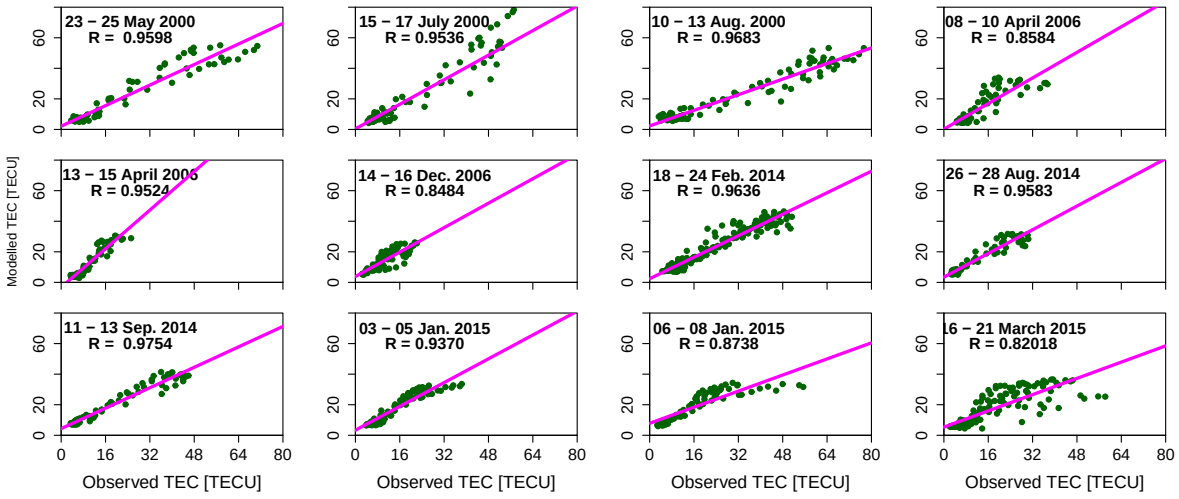


Figure 4.11: Comparison between the modelled and observed values of TEC for twelve selected storms. Also shown, are the correlation coefficients between the observed and modelled TEC. The storm period is specified for each storm.

For all storm periods, the correlation coefficients are high (in the range 0.82 - 0.98), which means that the observed and modelled TEC are highly and positively correlated. The observed and modelled TEC tend to move in the same direction, i.e. the modelled TEC covers most of the information contained in the observed TEC. The highest correlation coefficients

were generally observed for storms in 2000 and 2014, which implies that most of the information was covered for storms that occurred during years of high solar activity and storms with no significant TEC response. Comparing both representations of low and high solar activity periods in the dataset used for the EOF model development, the solar minimum period was not well represented, meaning that the original TEC dataset was dominated by TEC values observed during the solar maximum period. This may be the reason why correlation coefficients are generally higher for storms of 2000 compared with the 2006 storms.

Although the storm periods were not the same, the RMSE values obtained when modelling TEC using the neural network technique (Habarulema *et al.*, 2010) for storms that occurred during both high and low solar activity period were found to be relatively in the same range as the ones given by the EOF model. However, in some cases it was found that the RMSE values for EOF model were smaller, although a further study is needed in order to make a general conclusion.

4.4 Model validation over other stations

Although the EOF model was specifically developed for Sutherland (32.38^o S, 20.81^o E), it was tested with data from other GPS receiver stations and its performance was evaluated. Table 4.2 provides the geographical latitudes and longitudes of the stations where the EOF model was validated.

Table 4.2: Geographic latitudes (GLAT) and longitudes (GLON) of GPS receiver stations used in this project.

Station name	Station code	GLat (^{o}) S	GLon (^{o})E
Sutherland	SUTH	32.38	20.81
Cape Town	CTWN	33.95	18.47
Port Elizabeth	PELB	33.98	25.61
Aliwal North	ANTH	30.68	26.72
Springbok	SBOK	29.67	17.88
Bethlehem	BETH	28.25	28.33
Mafikeng	MFKG	25.81	25.54
Ellisras	ERAS	23.69	27.70
East London	ELDN	33.04	27.83
George	GEOA	34.00	22.38
Ulundi	ULDI	28.29	31.42

Figure 4.12 shows the positions of these stations relative to the ionospheric pierce points (IPPs) over Sutherland (32.38^o S, 20.81^o E), after selecting satellites with elevation angles greater than 20^o to reduce the multipath effects. The rationale behind validating the EOF

model with data from different GPS receiver stations is to determine an appropriate latitudinal separation that will be useful in regional TEC modelling during storm conditions.

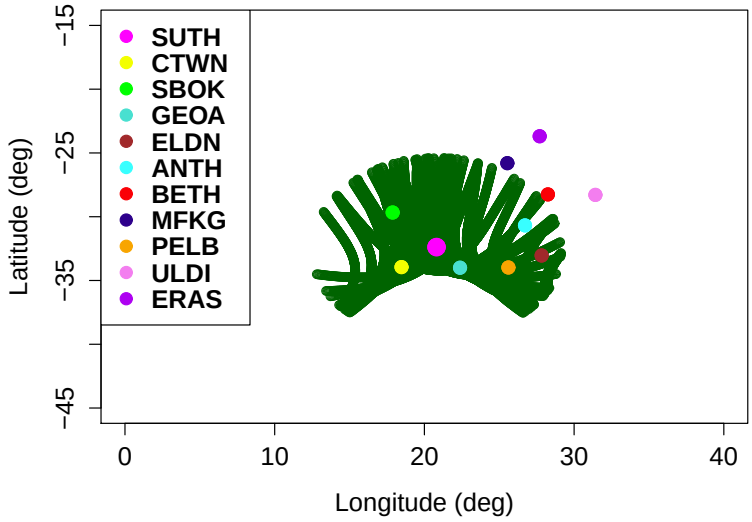


Figure 4.12: Locations of different GPS receiver stations relative to IPPs over Sutherland (32.38^o S, 20.81^o E) after selecting satellites with elevation angles greater than 20^o.

As previously explained, the EOF base functions and coefficients were obtained from TEC over Sutherland (32.38^o S, 20.81^o E) and the modelled TEC was obtained using A and F10.7 indices as inputs to get the modelled EOF coefficients, which in turn, were combined with the base functions. A study was thus done to compare the observed TEC over different stations with the modelled TEC over Sutherland (32.38^o S, 20.81^o E), for selected storm periods. This was done to determine whether the EOF model developed for a specific location could be used to estimate TEC over other stations/locations. For this purpose, different storm periods were selected. The selected storm periods are 18 - 24 February, 2014 (4 storms), 21 - 26 January, 2004 (2 storms) and 06 - 10 March, 2012 (2 storms). For each storm the observed and modelled TEC were compared taking into account sets of stations/locations within and outside the IPP coverage area. Note that it was not possible to validate the EOF model over other stations using the same selected storm periods for Sutherland since for most of the stations, no data was available for 2000 and 2006.

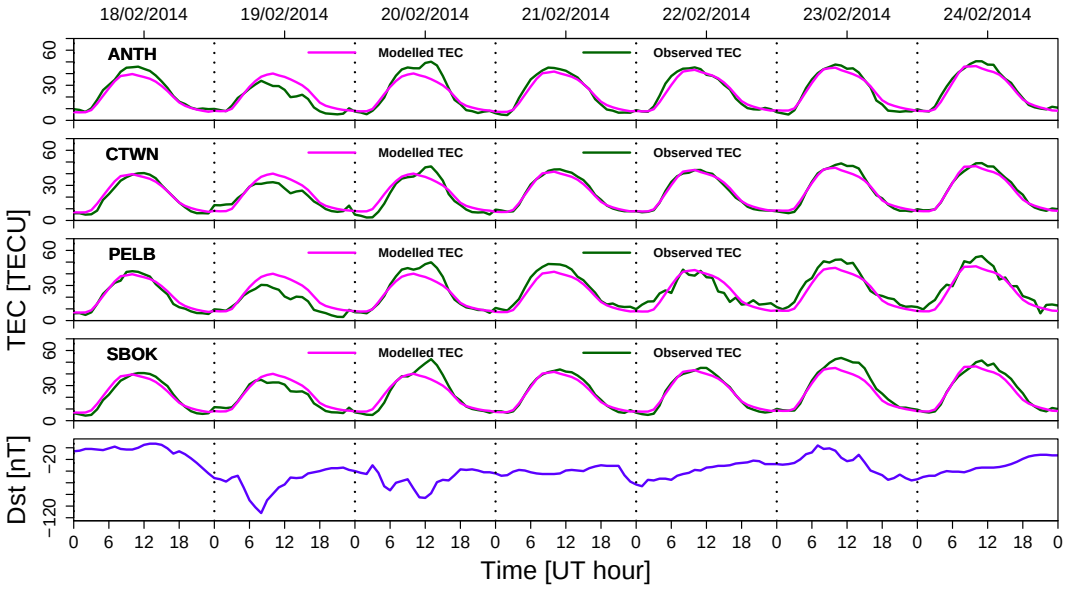
Figure 4.13 shows observed and modelled TEC variations during the storm period 18 - 24 February, 2014. Figures 4.13 (a) and (b) are for stations within and outside the IPP coverage area over Sutherland (32.38^o S, 20.81^o E) respectively. Graphically, it is clear that the EOF model predicts TEC quite well over all eight stations around Sutherland. Figure 4.14 is similar to Figure 4.13, but covers the storm period of 21 - 26 January, 2004. Since there was

no data for CTWN and PELB during this storm period, the two stations were replaced by their respective closest stations GEOA and ELDN. Although most of GPS receiver stations had incomplete datasets (as shown in Figures 4.14 (a) and (b)), the EOF model overestimates TEC, except in a few cases, such as over SBOK and MFKG. Figure 4.15 is similar to Figures 4.13 and 4.14, but covers the storm period 06 - 10 March, 2012. The EOF model predicts TEC quite well over all stations within and outside the IPP coverage area, as shown in Figures 4.15 (a) and (b).

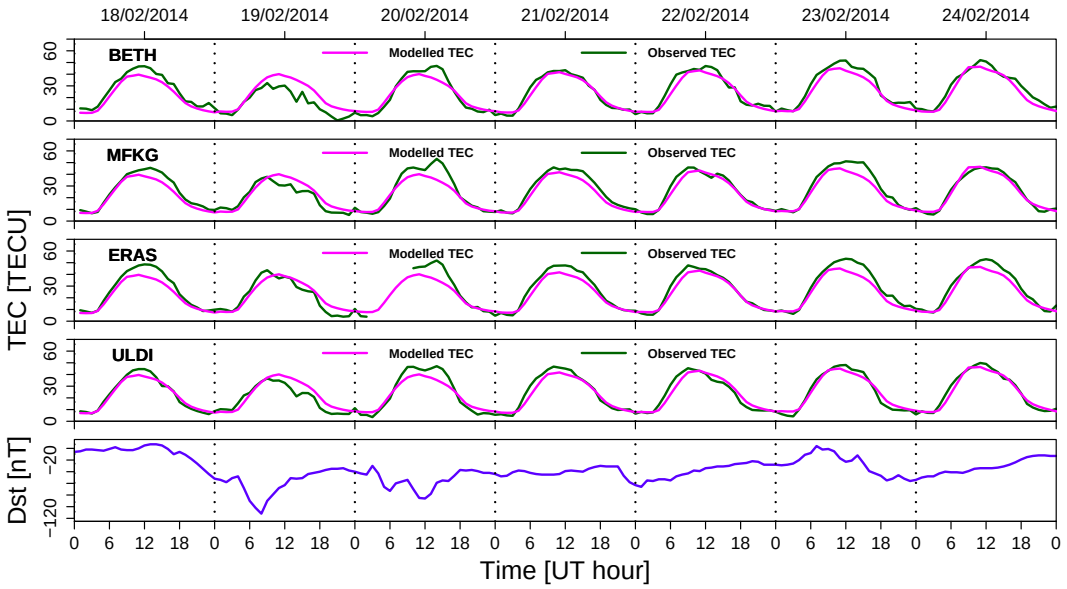
Errors (RMSE) were also computed for each storm period. For the storm period 18 - 24 February, 2014 the RMSE values calculated for different stations are close to the error values obtained when the EOF model was validated over Sutherland (32.38^o S, 20.81^o E). The same result was obtained for the storm period 06 - 10 March 2012. It is observed that the RMSE values generally tend to increase as the distance from Sutherland increases. Thus, for the two storm periods, these results indicate that the EOF model developed for Sutherland can be used to predict TEC over different stations. For example, the RMSE between the modelled and observed TEC over Sutherland for the 18 - 24 February, 2014 storm period is 3.94 TECU compared to the RMSE of 5.23 TECU over ERAS which is separated latitudinally from Sutherland by 8.7^o. Similarly, concerning longitudinal separation, the RMSE for the farthest station (ULDI), at a longitudinal distance of about 10.6^o from Sutherland, was estimated as 4.19 TECU.

However, for the storm period 21 - 26 January, 2004, only RMSE values obtained over SBOK and MFKG are comparable to the RMSE obtained for Sutherland, while for other stations, including GEOA (the closest station to Sutherland) they are higher since the EOF overestimates TEC. This discrepancy may have been influenced by the large amount of the missing data for this storm period and which thus didn't contribute to the computation of RMSE.

Correlation coefficients between the observed and modelled TEC for different stations were computed. Figure 4.17 shows that high correlation coefficients were observed for the storm periods 18 - 24 February, 2014 and 06 - 10 March 2012. This confirms that most of the information contained in the observed TEC is covered in the modelled TEC. For the storm period 21 - 26 January, 2004, the observed and modelled TEC are generally less correlated and this may be due to the missing data in the observed TEC which would have contributed in the computation of the correlation coefficients. The validation of the EOF model for different locations is important, because it contributes to solving the problem of data dimensionality in modelling regional TEC during storm periods, especially in mid-latitudes.



(a)

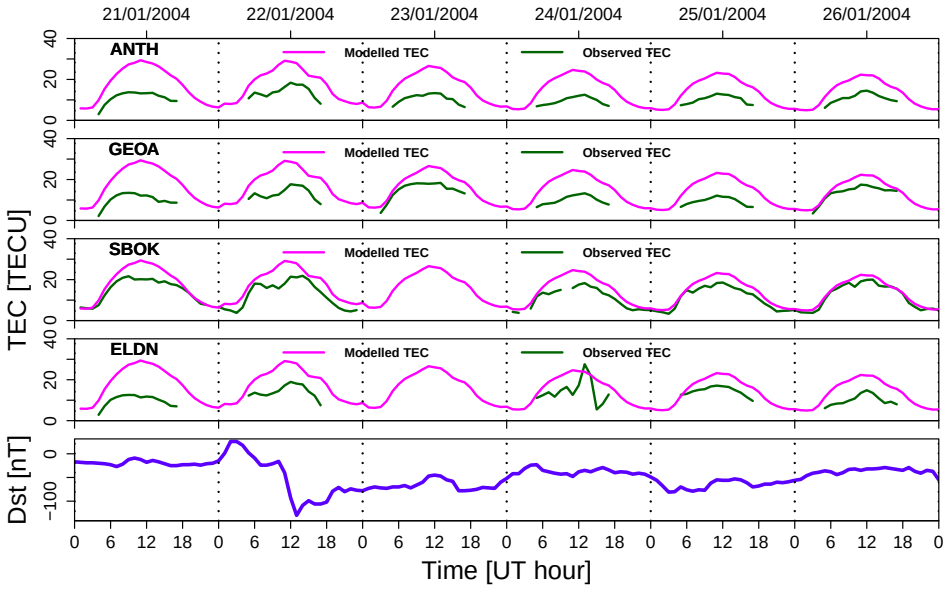


(b)

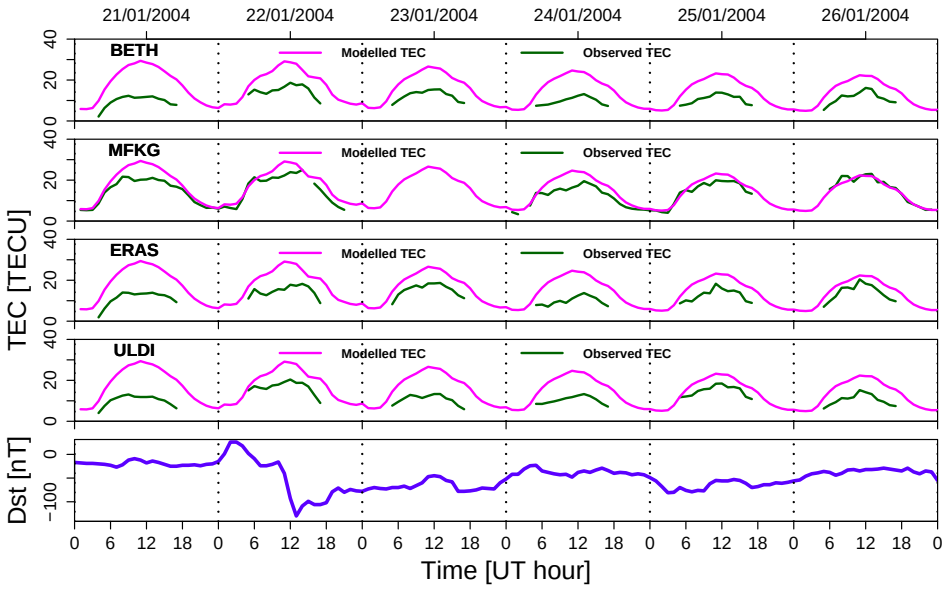
Figure 4.13: Comparison between observed and modelled TEC during the storm period 18 - 24 February, 2014 over (a) ANTH, CTWN, PELB and SBOK (b) BETH, MFKG, ERAS and ULDI.

4.5 Model improvement

Several alternatives were investigated to improve the EOF model developed for Sutherland (32.38^o S, 20.81^o E) by modifying A and F10.7 indices, which had previously been used as inputs. Since most of the storms had a storm period of three days, a 3-day running average



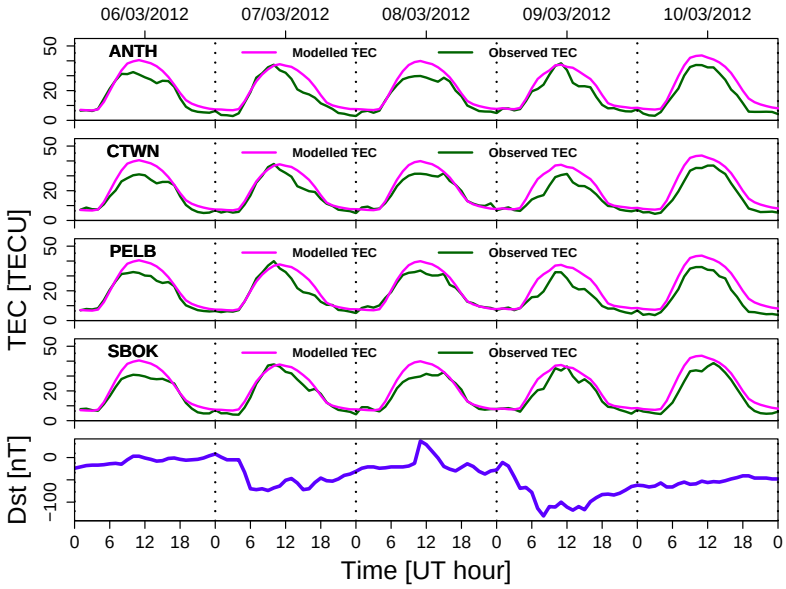
(a)



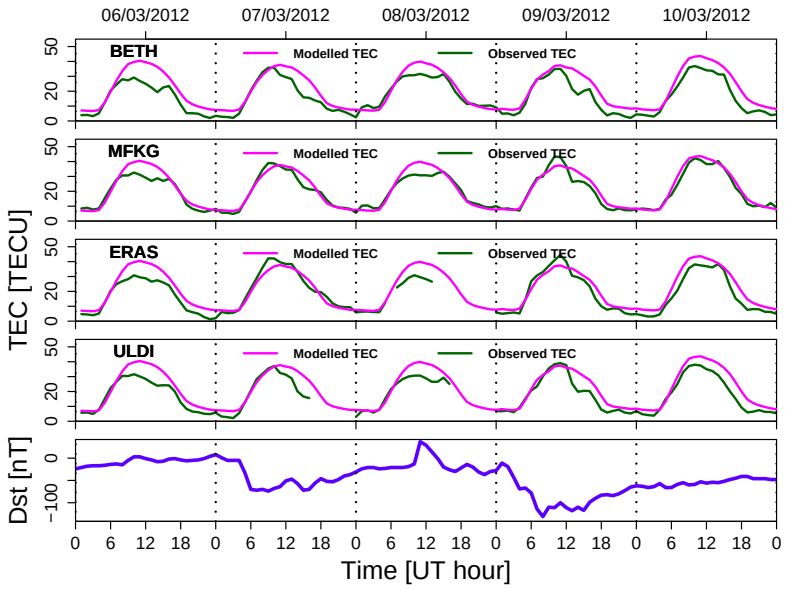
(b)

Figure 4.14: Comparison between observed and modelled TEC during the storm period from 21 - 26 January, 2004, over (a) ANTH, GEOA, SBOK and ELDN (b) BETH, MFKG, ERAS and ULDI.

of A and F10.7 indices were used as inputs. The reason for this, was to try to capture on average of three days, the general information of the level of geomagnetic and solar activities. Different combinations of indices derived from A and F10.7 indices (hereafter referred to as (A, F10.7)), were used as inputs to see if any of them could contribute to improving the EOF model:



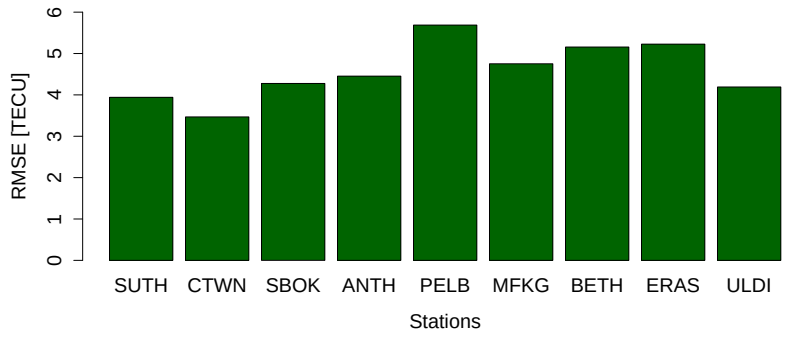
(a)



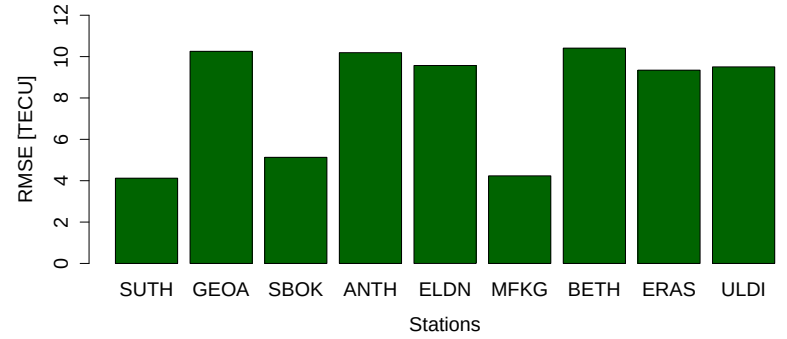
(b)

Figure 4.15: Comparison between observed and modelled TEC during the storm period 06 - 10 March 2012, over (a) ANTH, CTWN, PELB and SBOK (b) BETH, MFKG, ERAS and ULDI.

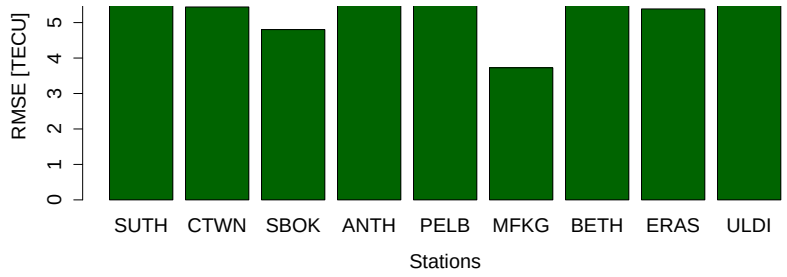
- a 3-day running average of A index was used with F10.7 values as inputs (Amod, F10.7)
- a 3-day running average of F10.7 was used with A values as inputs (A, F10.7mod)
- a 3-day running averages of both A and F10.7 indices were used as inputs (Amod, F10.7mod).



(a)



(b)



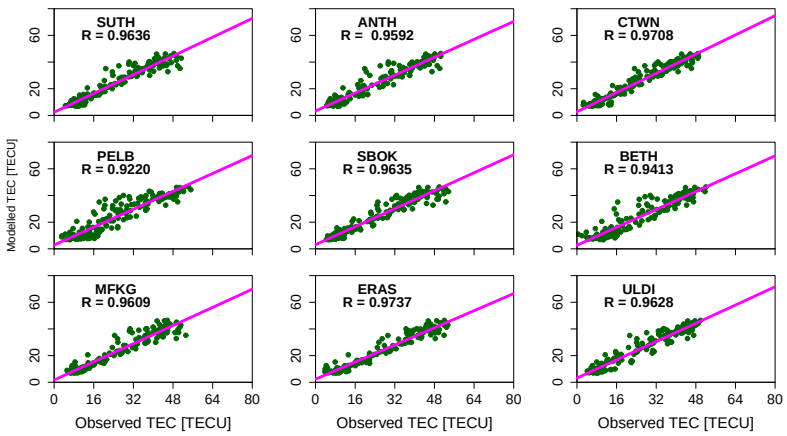
(c)

Figure 4.16: RMSE values for selected storms (a) 18 - 24 February, 2014 (b) 21 - 26 January, 2004 (c) 06 - 10 March 2012.

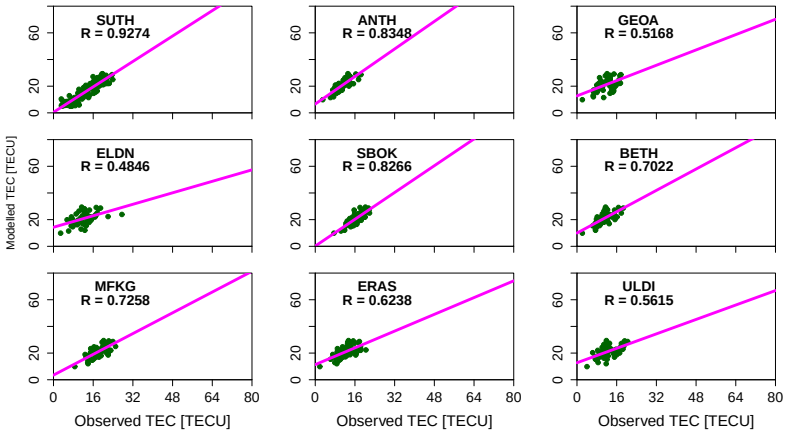
In addition to these, another option was to replace F10.7 by F10.7p, where the latter is computed by using the F10.7 of the day of interest and its average over the previous 81 days (F10.7A), according to the formula

$$\mathrm{F10.7p} = \frac{\mathrm{F10.7} + \mathrm{F10.7A}}{2} \tag{4.7}$$

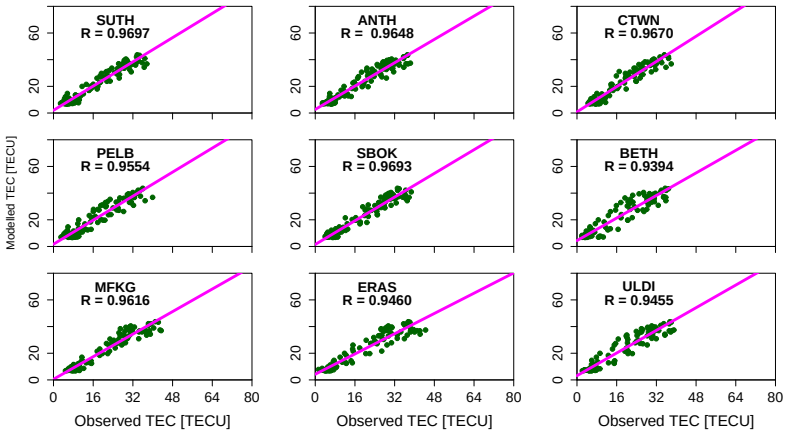
F10.7p represents, in statistical sense, the intensity of the solar EUV flux and it has been shown that it is a better solar proxy for common use than F10.7, since it has the advantage of long-term records (Liu and Chen, 2009; Bergeot *et al.*, 2013). The F10.7p index was recently used as input in modelling foF2 and TEC during quiet conditions using EOF analysis (A *et al.*, 2011, 2012).



(a)



(b)



(c)

Figure 4.17: Comparison between observed and modelled TEC for storms that occurred (a) 18 - 24 February, 2014 (b) 21 - 26 January, 2004 (c) 06 - 10 March 2012. Also shown, are the correlation coefficients between the observed and modelled TEC.

Figure 4.18 shows the errors (RMSE) obtained for each combination of inputs used during the EOF model development for Sutherland (32.38^o S, 20.81^o E). It is clear that small RMSE values were generally observed when A and F10.7p were used as inputs. For a sample of twelve selected storms, 58.33 % of the total number of storms gave the smallest errors (RMSE) when A and F10.7p indices were used as inputs. For other cases, the corresponding percentages are also shown in Figure 4.19. Thus, the use of the A and F10.7p indices may improve the results, compared to other combinations of indices, as specified above.

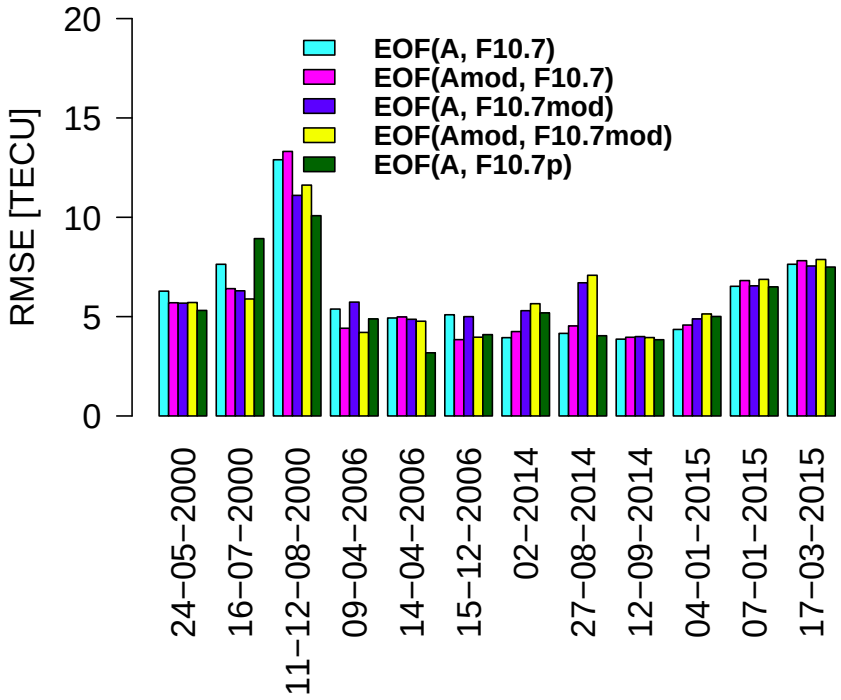


Figure 4.18: RMSE values for selected storms using modified indices. The RMSE values for each storm are shown by bars that correspond to the storm date.

Figure 4.20 compares the observed and modelled TEC over Sutherland (32.38^o S, 20.81^o E), obtained by using different combinations of indices for the storm periods 10 - 13 August, 2000 and 13 - 15 April, 2006. Figure 4.20 (a), shows that the EOF model predicts TEC magnitude better when A and F10.7p indices are used as inputs than in other cases. The predicted TEC values are the highest for almost the whole storm period (except on 10-08-2000 where (Amod, F10.7mod) gave the highest values), and closer to the observed TEC values than for any other combination of inputs. Although the model is still underestimating TEC, the deviation between the observed and the modelled TEC has reduced. In Figure 4.20 (b) the modelled TEC values are close to the observed ones when A and F10.7p are used as inputs. For this case, although the predicted TEC values are the smallest for almost the

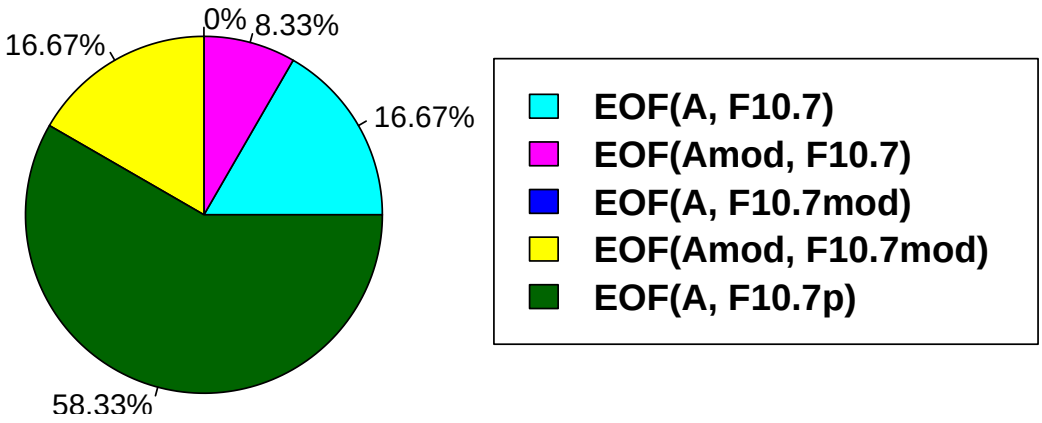
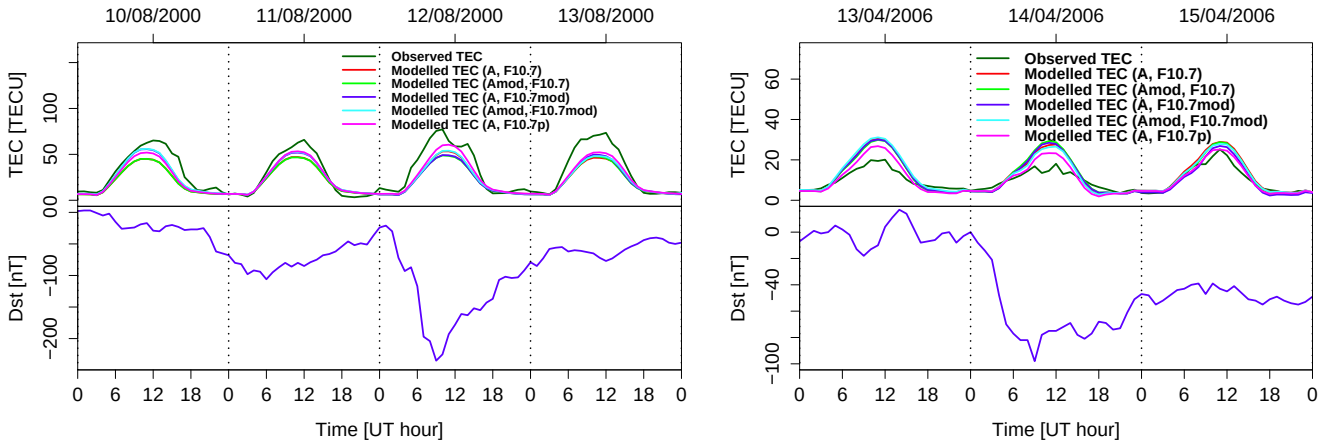


Figure 4.19: Frequency of occurrence each set of indices was found to give the smallest RMSE values.

whole storm period, the model is still overestimating TEC. While further improvement of the EOF model is necessary for better prediction, these observations indicate that A and F10.7p give optimum solutions in modelling storm-time TEC, among different sets of investigated inputs.



(a) (b)

Figure 4.20: Comparison between observed and modelled TEC using modified indices as inputs: (a) storms of 11-12 August, 2000, (b) storm of 14-04-2006.

4.6 Summary

EOF and regression analyses were used to develop analytical equations to estimate TEC over Sutherland (32.38^o S, 20.81^o E) during storm conditions. Specifically, this is the first modelling effort to investigate the usage of EOF analysis in estimating storm-time behaviour. Previous studies used the EOF model to estimate foF2 and TEC during magnetically quiet conditions and the results were very impressive (Mao *et al.*, 2005, 2008; A *et al.*, 2011, 2012). However, having tried to apply the EOF model to storm-time conditions using the global ionospheric maps provided by Jet Propulsion Laboratory, A *et al.* (2012) recommended the use of TEC data of high spatial-temporal resolution for the model to perform well. In this study, EOF analysis was used to decompose the observed TEC dataset into EOF base functions, which mainly represent the diurnal variation of TEC, and associated coefficients, which represent the long-term variation of TEC. Regression analysis was applied to model the EOF coefficients in terms of A and F10.7 indices. Tested on different storms, it was found that the EOF model estimates TEC for storms with no significant TEC response quite well, but it fails to capture some features in the observed TEC. Improvements are therefore still required to capture storm-time behaviour during different phases of geomagnetic storms. The model validation by other stations showed that the EOF model developed for a specific station can be used to estimate TEC over other locations within a latitudinal and longitudinal coverage of 8.7^o and 10.6^o respectively. This is particularly important since it contributes to the reduction of data for storm-time regional TEC modelling. The model improvement based on modified indices has shown that the use of A and F10.7p indices as inputs improves the modelling results.

Chapter 5

Conclusions and future work

The main objective of this study was to develop mathematical equations for estimating TEC over Sutherland (32.38^o S, 20.81^o E) during storm conditions. EOF and regression analyses methods have been used to achieve the objective. This chapter presents conclusions about TEC modelling during storm conditions, challenges related to storm-time TEC modelling, as well as future work.

5.1 Conclusions

The use of EOF analysis to decompose the observed TEC data over Sutherland in terms of EOF base functions and associated coefficients led to the following observations:

- The EOF base functions represent mainly the diurnal variation of TEC as noticed from the first and the second order EOF base functions. The first EOF base function E_1 in particular represents the diurnal variation of the average TEC. The higher order EOF base functions represent short-term variation of TEC as well as noise effects. A similar result was stated in different studies related to the EOF modelling of ionospheric parameters (foF2, TEC, hmF2, M(3000)F2) during magnetically quiet conditions (Mao *et al.*, 2005; Liu *et al.*, 2008; Mao *et al.*, 2008; Zhang *et al.*, 2009; A *et al.*, 2011, 2012; Lin *et al.*, 2014).

- The EOF coefficients represent the long-term variation of TEC. The first order EOF coefficient A_1 in particular represents the solar cycle variation pattern of TEC.

- Twelve EOF components were able to explain 99.179 % of the total variance in the original TEC dataset, leaving only 0.821 % unexplained. Hence, twelve EOF components were used to develop the model. For other works related to modelling of the ionospheric parameters based on EOF method, it was shown that the EOF method converges quickly during quiet conditions and only 3 to 4 EOF components were retained to build the model (Mao *et al.*, 2005; Liu *et al.*, 2008; Mao *et al.*, 2008; Zhang *et al.*, 2009; A *et al.*, 2011, 2012; Lin *et al.*, 2014). However, since modelling storm-time

TEC is complicated due to high variability in the data, it has been necessary to retain 12 EOF components. The first component $A_1 \times E_1^T$, which accounts for 72.86% of the total variance in the original TEC dataset, justifies the conclusion that the main cause of TEC variation is solar activity.

The use of regression analysis after TEC decomposition, allowed for:

- the introduction of solar and geomagnetic activities in the model by modelling the EOF coefficients in terms of the A and F10.7 indices.
- the identification of the solar cycle, annual, and semiannual variation components in the EOF coefficients.

For a specific storm day, the A and F10.7 indices were used as inputs to get the modelled EOF coefficients, which were then combined with the base functions to get the modelled TEC for that specific day. A comparative study of the modelled and the observed TEC for selected storm periods, led to the following conclusions:

- The EOF model predicts the TEC magnitude well for storms with no significant ionospheric TEC response.
- For storms with significant TEC response, the EOF predicts the TEC magnitude well during the nighttime and early morning and for some storms, during the recovery phase. However, TEC enhancement and depression and some short-term features in the observed TEC are not captured sufficiently by the EOF model.
- The failure of the EOF model to capture positive and negative TEC response and, to an extent short-term features in the observed TEC well, may be due to the use of the daily indices A and F10.7 as inputs. These indices represent on average, the daily level of geomagnetic and solar activity while geomagnetic storm features are short-term varying phenomena.

A statistical analysis of the observed and modelled TEC showed that:

- Small RMSE values for storms with no significant TEC response and storms that occurred during periods of low solar activity, confirm that the EOF model performs better for storms with no significant TEC response and for storms that occurred during periods of low solar activity.
- The high correlation between the observed and modelled TEC indicates that the modelled TEC covers most of the information contained in the observed TEC.

The model validation with data from other stations showed that the EOF model developed for a specific station may be used to estimate TEC over other locations within a latitudinal and longitudinal coverage of 8.7^o and 10.6^o respectively. This result indicates that during regional storm-time TEC modelling, it may not be necessary to compute TEC data for all the closest receiver stations, since most of the needed information can be extracted from measurements at one location. This is an important step towards reducing data dimensionality problem especially for high temporal resolution modelling.

The improvement of the model by using the modified A and F10.7 indices, showed that the use of the A and F10.7p indices as inputs to model the EOF coefficients, improves the accuracy of the model. However, this is not enough to achieve a better performance and further improvements to the EOF model are still needed.

5.2 Challenges in modelling storm-time TEC using the EOF analysis

The EOF model was based on TEC data derived from GPS records at Sutherland (32.38^o S, 20.81^o E), for storm periods only. However, during the solar minimum a small number of storms were observed, compared to the solar maximum period. This means that the TEC dataset used to develop the EOF model was dominated by TEC values for storms during solar maximum.

The A and F10.7 daily indices are recorded for each day and this is a long-time compared to the timescale of geomagnetic storms features. This may be the cause of the failure of the EOF model to capture short-term features in the observed TEC.

5.3 Future work

The EOF model needs to be improved for better performance. To improve modelling of TEC over Sutherland (32.38^o S, 20.81^o E), an alternative would be to replace the A index by the Dst index and then use the combination (Dst, F10.7p) as inputs to model the EOF coefficients. Since the Dst index is an hourly index, while the F10.7p is a daily index, the minimum or the average daily for a specific day would be used with the F10.7p for that day, as inputs. The output would be compared with the modelled TEC, using the combination (A, F10.7p) to determine which inputs would be the best.

Future work will involve comparisons of the EOF technique with other modelling approaches

such as neural networks, auto-covariance prediction methods, and other empirical models, e.g. IRI.

The EOF model was used to perform single station TEC modelling in mid-latitude region. In future this method will be extended to other African latitudes for region storm -time TEC modelling. It is hoped that the success of this will form a basis of applying this method to global storm-time ionospheric TEC modelling.

References

A, E., Zhang, D.-H., Xiao, Z., Hao, Y.-Q., Ridley, A. J., and Moldwin, M.: Modeling ionospheric foF2 by using empirical orthogonal function analysis, Annales Geophysicae, 29, 1501–1515, 2011.

A, E., Zhang, D., Ridley, A. J., Xiao, Z., and Hao, Y.: A global model: Empirical orthogonal function analysis of total electron content 1999 – 2009 data, Journal of Geophysical Research: Space Physics, 117, A03 328, 2012.

Bagiya, M. S., Joshi, H., Iyer, K., Aggarwal, M., Ravindran, S., and Pathan, B.: TEC variations during low solar activity period (2005–2007) near the equatorial ionospheric anomaly crest region in India, 27, 1047–1057, 2009.

Balan, N., Bailey, G., Abdu, M., Oyama, K., Richards, P., MacDougall, J., and Batista, I.: Equatorial plasma fountain and its effects over three locations: Evidence for an additional layer, the F3 layer, Journal of Geophysical Research: Space Physics (1978–2012), 102, 2047–2056, 1997.

Baumjohann, W., Treumann, R. A., and Treumann, R. A.: Basic Space Plasma Physics, vol. 57, Imperial College Press, London, 1997.

Bergeot, N., Tsagouri, I., Bruyninx, C., Legrand, J., Chevalier, J.-M., Defraigne, P., Baire, Q., and Pottiaux, E.: The influence of space weather on ionospheric total electron content during the 23rd solar cycle, Journal of Space Weather and Space Climate, 3, A25, 2013.

Björnsson, H. and Venegas, S. A.: A Manual for EOF and SVD Analysis of Climatic Data, MCGill University, Department of Atmospheric and Oceanic Sciences and Center of Climate and Global change Research, 1997.

Borries, C., Berdermann, J., Jakowski, N., and Wilken, V.: Ionospheric storms A challenge for empirical forecast of the total electron content, Journal of Geophysical Research: Space Physics, 120, 3175–3186, 2015.

Bremner, S.: Analysing Transient Effects in the Ionosphere Using Narrowband VLF Data, Master's thesis, University of KwaZulu-Natal, Westville, 2009.

Campbell, W. H.: Introduction to Geomagnetic Fields, Press syndicate of the University of Cambridge, Cambridge, United Kingdom, 1997.

Campbell, W. H.: Earth Magnetism: A Guided Tour Through Magnetic Fields, Harcourt/Academic Press, San Diego, U.S.A, 2001.

Carrano, C. S. and Groves, K. M.: TEC Variability and Scintillation at Low Latitudes Measured Using High-Rate GPS Receivers, 2008 URSI General Assembly, August 7-16, 2008.

Chatterjee, S. and Hadi, A. S.: Regression Analysis by Example, Fourth Edition, John Wiley & Sons, U.S.A, 2013.

Chauhan, V., Singh, O., and Singh, B.: Diurnal and seasonal variation of GPS-TEC during a low solar activity period as observed at a low latitude station Agra, Indian Journal of Radio & Space Physics, 40, 26–36, 2011.

Chen, F. F.: Introduction to Plasma Physics and Controlled Fusion, volume 1: Plasma physics, Plenum Press, New York, 1984.

Davies, K.: Ionospheric radio, Peter Peregrinus, London, United Kingdom, 1990.

De Michelis, P., Tozzi, R., and Consolini, G.: Principal components' features of mid-latitude geomagnetic daily variation, Annales Geophysicae, 28, 2213–2226, 2010.

esa navipedia: European Space Agency, `http://www.navipedia.net/index.php/Ionospheric_Delay`, (Accessed 22-04-2015), 2015.

Fuller-Rowell, T., Codrescu, M., and Wilkinson, P.: Quantitative modeling of the ionospheric response to geomagnetic activity, Annales Geophysicae, 18, 766–781, 2000.

GPS.gov: Official U.S. Government information about the Global Positioning System (GPS) and related topics, `http:www.gps.gov/systems/gps/control/\#personnel`, (Accessed 20-04-2015), 2015.

Grewal, M. S., Weill, L. R., and Andrews, A. P.: Global Positioning Systems, Inertial Navigation, and Integration, John Wiley & Sons, New York, 2007.

Habarulema, J. B.: A contribution to TEC modelling over Southern Africa using GPS data, Ph.D. thesis, Rhodes University, Grahamstown, South Africa, 2010.

Habarulema, J. B., McKinnell, L.-A., and Cilliers, P. J.: Prediction of global positioning system total electron content using neural networks over South Africa, Journal of Atmospheric and Solar-Terrestrial Physics, 69, 1842–1850, 2007.

Habarulema, J. B., McKinnell, L.-A., Cilliers, P. J., and Opperman, B. D.: Application of neural networks to South African GPS TEC modelling, Advances in Space Research, 43, 1711–1720, 2009.

Habarulema, J. B., McKinnell, L.-A., and Opperman, B. D.: TEC measurements and modelling over Southern Africa during magnetic storms; a comparative analysis, Journal of Atmospheric and Solar-Terrestrial Physics, 72, 509–520, 2010.

Habarulema, J. B., McKinnell, L.-A., Burešová, D., Zhang, Y., Seemala, G., Ngwira, C., Chum, J., and Opperman, B.: A comparative study of TEC response for the African equatorial and mid-latitudes during storm conditions, Journal of Atmospheric and Solar-Terrestrial Physics, 102, 105–114, 2013.

Hannachi, A.: A Primer for EOF Analysis of Climate Data, Departement of Meteorology, University of Reading, Reading, United Kingdom, 2004.

Hannachi, A., Jolliffe, I., and Stephenson, D.: Empirical orthogonal functions and related techniques in atmospheric science: a review, International Journal of Climatology, 27, 1119–1152, 2007.

Herman, J. and Goldberg, R.: Sun, Weather And Climate., Library of Congress Cataloging in Publication Data, Washington, D.C., 1978.

Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J.: GPS. Theory and Practice, Springer-Verlag Wien, Vienna, 1992.

Holland, S. M.: Principal Components Analysis (PCA), University of Georgia, Department of Geology, Athens, 2008.

Huang, C., Liu, D.-D., and Wang, J.-S.: Forecast daily indices of solar activity, F10. 7, using support vector regression method, Research in Astronomy and Astrophysics, 9, 694–702, 2009.

Huang, Z. and Roussel-Dupré, R.: Total electron content (TEC) variability at Los Alamos, New Mexico: A comparative study: FORTE-derived TEC analysis, Radio Science, 40, 2005.

Hunsucker, R. D. and Hargreaves, J. K.: The High-latitude Ionosphere and its Effects on Radio Propagation, Cambridge University Press, Cambridge, 2002.

Jakowski, N., Béniguel, Y., De Franceschi, G., Pajares, M. H., Jacobsen, K. S., Stanislawska, I., Tomasik, L., Warnant, R., and Wautelet, G.: Monitoring, tracking and forecasting ionospheric perturbations using GNSS techniques, Journal of Space Weather and Space Climate., 2, A22, 2012.

Kaplan, E. and Hegarty, C.: Understanding GPS: Principles and Applications, Second Edition, Artech House, Norwood, 2005.

Kumar, S., Tan, E., Razul, S., See, C., and Siingh, D.: Validation of the IRI-2012 model with GPS-based ground observation over a low-latitude Singapore station, Earth, Planets and Space, 66, 17, 2014.

Lin, J., Yue, X., Zeng, Z., Lou, Y., Shen, X., Wu, Y., Schreiner, W. S., and Kuo, Y.-H.: Empirical orthogonal function analysis and modeling of the ionospheric peak height during the years 2002–2011, Journal of Geophysical Research: Space Physics, 119, 3915–3929, 2014.

Liu, C., Zhang, M.-L., Wan, W., Liu, L., and Ning, B.: Modeling M (3000) F2 based on empirical orthogonal function analysis method, Radio Science, 43, 1003, 2008.

Liu, L. and Chen, Y.: Statistical analysis of solar activity variations of total electron content derived at Jet Propulsion Laboratory from GPS observations, Journal of Geophysical Research: Space Physics (1978–2012), 114, A10 311, 2009.

Loewe, C. A. and Prölss, G. W.: Classification and mean behavior of magnetic storms, Journal of Geophysical Research: Space Physics, 102, 14 209–14 213, 1997.

Love, J. J. and Gannon, J. L.: Movie-maps of low-latitude magnetic storm disturbance, Space Weather, 8, S06 001, 2010.

Mankin, E.: Principal Components Analysis: A How-To Manual for R, University of Colorado Boulder, 2014.

Mao, T., Wan, W.-X., and Liu, L.-B.: An EOF based empirical model of TEC over Wuhan, Chinese Journal of Geophysics, 48, 827–834, 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 Science, 43, RS2009, 2008.

McNamara, L. F.: The Ionosphere: Communications, Surveillance, and Direction Finding, Krieger Publishing Company, Malabar, Florida, 1991.

Misra, P. and Enge, P.: Global Positioning System: Signals, Measurements and Performance, Second Edition, Lincoln MA: Ganga-Jamuna Press, 2006.

Moeketsi, D. M.: Solar cycle effects on GNSS-derived ionospheric total electron content observed over southern Africa, Ph.D. thesis, Rhodes University, Grahamstown, South Africa, 2008.

Moldwin, M.: An Introduction to Space Weather, Cambridge University Press, Cambridge, 2008.

Mosert, M., McKinnell, L., Gende, M., Brunini, C., Araujo, J., Ezquer, R., and Cabrera, M.: Variations of foF2 and GPS total electron content over the Antarctic sector, Earth, planets and space, 63, 327–333, 2011.

Ondoh, T. and Marubashi, K.: Science of Space Environment, Ohmsha, Ltd., Chiyoda-ku, Tokyo, Japan, 2001.

Oyeyemi, E. O., McKinnell, L., and Poole, A.: Near-real time foF2 predictions using neural networks, Journal of Atmospheric and Solar-Terrestrial Physics, 68, 1807–1818, 2006.

Poole, I.: Understanding Solar Indices, vol. 86, pp. 38–40, American Radio Relay League, 2002.

Rao, S., Galav, P., Sharma, S., and Pandey, R.: Low-latitude TEC variability studied from magnetically conjugate locations along 73E longitude, Journal of Atmospheric and Solar-Terrestrial Physics, 104, 1–6, 2013.

Rathore, B. S., Gupta, D. C., and Kaushik, S. C.: Effect of solar wind plasma parameters on space weather, Research in Astronomy and Astrophysics, 15, 85–106, 2014.

Reeve, W. D.: Geomagnetism Tutorial, Reeve Observatory Anchorage, Alaska, USA, 2010.

Riley, K. F., Hobson, M. P., and Bence, S. J.: Mathematical methods for physics and engineering, Third Edition, Cambridge University Press, Cambridge, 2006.

Rishbeth, H. and Garriott, O. K.: Introduction to Ionospheric Physics, Academic press, New York, 1969.

Seemala, G. K.: GPS-TEC analysis application, Institute for Scientific Research, Boston College, Boston, 2004.

Shlens, J.: A tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition, Princeton University, New Jersey, 2003.

Smith, L. I.: A tutorial on principal components analysis, University of Otago, Dunedin, New Zealand, 2002.

Srivastava, N. and Venkatakrishnan, P.: Solar and interplanetary sources of major geomagnetic storms during 1996–2002, Journal of Geophysical Research: Space Physics, 109, A10 103, 2004.

Stankov, S., Stegen, K., and Warnant, R.: Seasonal variations of storm-time TEC at European middle latitudes, Advances in Space Research, 46, 1318–1325, 2010.

Suhr, D. D.: Principal component analysis vs. exploratory factor analysis, SUGI 30, pp. 203–230, 2005.

Toffoletto, F.: Geomagnetic Storms and Substorms, Rice University, Houston, Texas, 2004.

Tsurutani, B. T.: The Interplanetary Causes of Magnetic Storms, Substorms and Geomagnetic Quiet, p. 103, Space Storms and Space Weather Hazards, Jet Propulsion Laboratory, California Institute of Technology, 2001.

Tsurutani, B. T. and Gonzalez, W. D.: The interplanetary causes of magnetic storms: A review, vol. 98, pp. 77–89, American Geophysical Union, 1997.

Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L., Guarnieri, F. L., Gopalswamy, N., Grande, M., Kamide, Y., Kasahara, Y., Lu, G., Mann, I., *et al.*: Corotating solar wind streams and recurrent geomagnetic activity: A review, Journal of Geophysical Research: Space Physics, 111, A07S01, 2006.

Veenadhari, B., Kumar, S., Ram, S. T., Singh, R., and Alex, S.: Corotating interaction region (CIR) induced magnetic storms during solar minimum and their effects on low-latitude geomagnetic field and ionosphere, Indian Journal of Radio and Space Physics, 41, 306–315, 2012.

Watthanasangmechai, K., Supnithi, P., Lerkvaranyu, S., Tsugawa, T., Nagatsuma, T., and Maruyama, T.: TEC prediction with neural network for equatorial latitude station in Thailand, Earth, Planets and Space, 64, 473–483, 2012.

Weare, B. C. and Nasstrom, J. S.: Examples of extended empirical orthogonal function analyses, Monthly Weather Review, 110, 481–485, 1982.

Xu, W.-Y. and Kamide, Y.: Decomposition of daily geomagnetic variations by using method of natural orthogonal component, Journal of Geophysical Research: Space Physics, 109, A05 218, 2004.

Ya'acob, N., Abdullah, M., Ismail, M., and Zaharim, A.: Total electron content (TEC) and model validation at an equatorial region, in: Proceedings of the 13th WSEAS International Conference on Applied Mathematics, pp. 204–208, World Scientific and Engineering Academy and Society (WSEAS), 2008.

Ya'acob, N., Abdullah, M., Ismail, M., and Zaharim, A.: Model validation for GPS total electron content (TEC) using 10th polynomial function technique at an equatorial region, WSEAS Transactions on Computers, 8, 1533–1542, 2009.

Ya'acob, N., Ismail, M., and Abdullah, M.: GPS total electron content (TEC) prediction at ionosphere layer over the Equatorial region, INTECH Open Access Publisher, Universiti Kebangsaan Malaysia, 2010.

Yan, X.: Linear Regression Analysis: Theory and Computing, World Scientific Publishing Co. Pte. Ltd., New Jersey, 2009.

Yizengaw, E. and Essex, E.: Storm Time Seasonal Variation of TEC on the Southern Hemisphere Mid-Latitude Regions Using Signals from GPS Satellites, 4'th Oersted International Science Team (OIST-4) Proceedings, Copenhagen, Denmark, 2002.

Yu, J.-Y.: Part 4: Time series II, Lecture notes, Department of Earth System Science, University of California, Irvine, 2014.

Zhang, M.-L., Liu, C., Wan, W., Liu, L., and Ning, B.: A global model of the ionospheric F2 peak height based on EOF analysis, Annales Geophysicae, 27, 3203–3212, 2009.