FORECASTING OF GLOBAL VERTICAL TOTAL ELECTRON CONTENT BASED ON TRIGONOMETRIC B-SPLINE WITH LONG SHORT TERM MEMORY

TRİGONOMETRİK B-SPLİNE TABANLI KÜRESEL DİKEY TOPLAM ELEKTRON İÇERİĞİNİN UZUN KISA SÜRELİ BELLEK (LSTM) İLE TAHMİNİ

İREM YILDIZ

Assist. Prof. Dr. MURAT DURMAZ

Supervisor

Submitted to Graduate School of Science and Engineering of Hacettepe University

as a Partial Fulfilment to the Requirements

for the Award of the Degree of Master of Science

in Geomatics Engineering

2021

To My Family...

ABSTRACT

FORECASTING OF GLOBAL VERTICAL TOTAL ELECTRON CONTENT BASED ON TRIGONOMETRIC B-SPLINE WITH LONG SHORT TERM MEMORY

YILDIZ, İrem

M.S., Department of Geomatics Engineering

Supervisor : Assist. Prof. Dr. Murat Durmaz

June 2021, 80 pages

Short term forecasting of Ionosphere is not only an important topic for both near real-time applications such as single frequency point positioning and navigation, but also monitoring the ionosphere by data assimilation methods. In this study, short term forecasting of global ionosphere on the basis of Trigonometric B-splines is studied with both Deep Learning methods such as LSTM and also conventional methods such as SARIMA. In addition, dimension reduction with Principal Component Analysis is also investigated. The Trigonometric B-spline coefficient time series of Global Ionosphere is obtained by generating coefficients using approximately 20 years of IGS global ionosphere maps in IONEX format. After examining the data, two different methods are proposed on the basis of trends. One is assuming the trend as constant, and the other is a combination of linear and annual trend by Facebook Prophet library. Performance of LSTM and SARIMA models are investigated in the forecasting of individual B-spline coefficient, and also in terms of forecasting Spatial Mean and Principal Components. In addition, a block based LSTM model is also proposed. Best model for each method is established by means of hyper parameter search. Then these best models are compared on days of both quiet and storm ionospheric conditions.

According to the results, the combination of dimension reduction with SARIMA model performs better in both quiet and storm days, with 56.17% and 32.59% improvement with respect to persistent ionosphere model, respectively. The proposed block based LSTM model and PCA LSTM provide close results to the SARIMA model with 15% and 21% improvement especially around 00:00 UT. In addition, up to 56% improvement

is achieved in the PCA SARIMA model in 2010 on selected days. Although LSTM provides a blackbox model building, feature engineering based on SARIMA model parameters in LSTM models may provide better results.

Keywords: Vertical Total Electron Content, Ionosphere Forecasting, LSTM, SARIMA, PCA

ÖZET

TRİGONOMETRİK B-SPLİNE TABANLI KÜRESEL DİKEY TOPLAM ELEKTRON İÇERİĞİNİN UZUN KISA SÜRELİ BELLEK (LSTM) İLE TAHMİNİ

YILDIZ, İrem
Yüksek Lisans, Geomatik Mühendisliği Bölümü
Tez Yöneticisi : Yrd. Doç. Dr. Murat Durmaz

Haziran 2021 , 80 sayfa

İyonosferin kısa dönemli tahmini, hem tek frekanslı konumlandırma ve navigasyon yarı gerçek zamanlıya uygulamalar için hem de veri asimilasyon yöntemleri ile iyonosferin izlenmesi için önemli bir konudur. Bu çalışmada, küresel iyonosferin Trigonometrik B-spline temelinde kısa dönemli tahmini, hem LSTM gibi Derin Öğrenme yöntemleri hem de SARIMA gibi geleneksel yöntemlerle incelenmiştir. Ayrıca Temel Bileşen Analizi ile boyut küçültme de incelenmiştir. Küresel İyonosfer'in Trigonometrik B-spline katsayı zaman serisi, IONEX formatında yaklaşık 20 yıllık IGS küresel iyonosfer haritaları kullanılarak katsayılar üretilerek elde edilmiştir. Veriler incelendikten sonra trend bazında iki farklı yöntem önerilmiştir. Biri trendin sabit olduğunu varsayarken, diğeri Facebook Prophet tarafından lineer trend ve yıllık trendin bir kombinasyonu kullanır. LSTM ve SARIMA modellerinin performansı, bireysel B-spline katsayısının tahmininde ve ayrıca Mekansal Ortalama ve Temel Bileşenlerin tahmin edilmesi açısından çalışılmıştır. Ayrıca blok tabanlı bir LSTM modeli önerilmiştir. Her yöntem için en iyi model, hiper parametre araması yoluyla belirlenir. Daha sonra her yöntemin en iyi modeli hem normal hem de fırtınalı iyonosferik koşullardaki günlerde karşılaştılmıştır.

Sonuçlara göre SARIMA modeli ile boyut küçültme kombinasyonu, hem normal hem de fırtınalı günlerde, persistence iyonosfer modeline göre sırasıyla %56.17 ve %32.59 iyileştirme ile daha iyi performans göstermektedir. Önerilen blok tabanlı LSTM modeli, özellikle 00:00 UT civarında normal günlerde %15, PCA LSTM model ile 21% iyileştirme ile SARIMA modeline yakın sonuçlar vermektedir. Ayrıca 2010 yili için seçilmiş günlerde PCA SARIMA modelde 56%'lara varan iyileşme sağlanmıştır. LSTM bir kara kutu model yapısı sağlamasına rağmen, LSTM modellerinin eğitiminde SARIMA mo-

delinin parametrelerine dayali öznitelik mühendisliği yapılarak daha iyi sonuçlar elde edilebilir.

Anahtar Kelimeler: Dikey Toplam Elektron İçeriği, İyonosfer Tahmini, LSTM, SARIMA, PCA

ACKNOWLEDGMENTS

I would like to thank my supervisor, Asst. Prof. Dr. Murat Durmaz, who has provided me with his knowledge and support throughout my work.

I would like to thank my jury members Prof. Dr. Çetin Mekik, Assoc.Prof.Dr. Sultan Kocaman Gökçeoğlu, Assoc.Prof.Dr. Ali Özgün Ok, Asst. Prof. Dr. Kamil Teke, for their suggestions and comments.

I would like to thank my family, my mother Serap Yıldız, my father Hüseyin Yıldız and my brother Yiğit Muzaffer Yıldız, for always motivating me and making me feel their love.

I would like to thank Didem Ateş, Yağmur Filizkaya, Mervecan Gökçe who have always been by my side, supporting and loving.

I would like to thank Sıla Baş who did not spare her experiences.

I would like to thank Recep Can, who is always by my side and supports me endlessly, whose love I always felt, and who I know will always be by my side.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

This study is about developing a model for short term forecasting of global Vertical Total Electron Content (VTEC) based on Trigonometric B-spline coefficients. State of the art methods, both with deep learning and statistical background are investigated for forecast model building. This chapter provides the problem definition with an overview of the recent studies about the ionosphere modeling and forecasting. The motivation of the study and its goals are also listed with an outline of the manuscript.

1.1 Motivation

Ionosphere is the upper layer of atmosphere with ionized gases and free electrons. It is ranging approximately from 50 km to 1000 km. There is a long lasting and active research on monitoring and prediction of Ionosphere, especially the electron density of the electron content of the Ionosphere, since the free electrons affect the electromagnetic wave propagation through Ionosphere. Thus, systems relying on radio communications such as Global Navigation Satellite Systems (GNSS), radio links, space weather and satellite telecommunication systems are all affected by Ionosphere [4]. The conditions in the Ionosphere may change with seasonal and non-seasonal patterns, mostly caused by the solar activity. Ionospheric variability may also have a significant coupling with thermosphere and some geophysical phenomena, for example climate change [1].

Thanks to the dual frequency code pseudorange and carrier phase measurements from Global Navigation Satellite Systems (GNSS), the Total Electron Content (TEC) in the Ionosphere can be obtained with high precision. The column integral of electron

density N_e along the path from ground based receiver to a GNSS satellite is called Slant Total Electron Content (STEC).

$$\int_s N_e(s)\, ds \tag{1.1}$$

where s is the slant path between satellite and receiver. STEC may be converted into the Vertical TEC (VTEC) by an elevation dependent mapping function [2]. Using the geometry-free linear combination of dual frequency GNSS measurements global coverage of VTEC can be obtained with high spatial and temporal resolution utilizing a globally established reference stations. These measurements then be used to estimate global VTEC by different methods.

There are many methods for estimation of Ionospheric VTEC representation in the literature. Most of them are using an expansion in some form of global basis functions such as Spherical Harmonics (SH) [2] or Trigonometric B-splines [5]. Global VTEC maps estimated with different methods by different analysis centers are combined within a weighting procedure, resulting in the International GNSS Service Ionosphere products in IONEX format [6] since as early as 1999. These maps published with a delay extending a week due to post-processing and integration. With the advancement in the internet and data speed, nowadays rapid and short term predictions of Global Ionosphere Maps (GIMs) are available [4]. However, the latency in data collection, processing and combination results in delays for estimated state of the Ionosphere reaching up to a few hours [5]. This requires efficient short term forecasting of VTEC representation parameters for real-time users and prediction of GIMs for a couple of days for planning purposes.

There are many studies that propose models for forecasting of global TEC maps [2; 7; 8; 9], with some others proposing regional TEC forecasting [10; 11]. Besides, there are studies working on local method for TEC estimation by [12; 13; 14; 15]. Several works have provided successfull results by statistical methods such as autoregressive (AR) [8] and autoregressive moving average (ARMA) [16]. [8] proposed that global VTEC maps may be forecasted by adaptive AR modeling of SH coefficients. [16] put forward the combination method, which contains seasonal effect model and ARMA to estimate short term forecasting of TEC and compared the combination method with

classical ARMA model.

With the increase in the data, artificial neural networks may be an alternative to estimation of TEC. Regional maps of TEC over Europe using a technique of bezier surface-fitting were estimated using neural networks [10]. Another model involving neural network and principal component analysis (PCA) is proposed by [14]. A genetic algorithm based neural network (GA-NN) is also developed [11] where GA was applied to optimize initial weights of the neural network. [15] proposed adaptive neuro-fuzzy inference system (ANFIS) and Principle Component Analysis (PCA) for reducing the dimension of data in order to forecast TEC. A deep neural network to forecast global TEC maps is also proposed recently [7]. There have been some studies based upon Long Short Term Memory (LSTM) which is a kind of recurrent neural network (RNN). [12; 13] proposed LSTM based method with the aim of forecasting ionospheric TEC for a single station . [9], on the other hand, proposed an LSTM model to forecast the 256 SH coefficients where they also utilized solar extreme ultraviolet (EUV) flux, disturbance storm time (DST) index, and hour of the day to predict next one, hour TEC in their estimation. There have been a study where global TEC model was created with machine learning-based XGBoost or XGBDT [17].

Highly accurate and precise short term forecasting of GIMs is still an important topic for real-time monitoring, positioning and navigation. Generally successful short-term forecasting requires accurate knowledge of the current state of the Ionosphere. Thus, forecasting is in general goes hand in hand with some sort of state estimation with new measurements [5]. In this context, short-term forecasting global Ionosphere model coefficients such as SH or B-splines may provide better results if supported by a data assimilation procedure. Successful near real-time global ionosphere estimation with Trigonometric B-splines [5] may be combined with a short-term forecasting of coefficients to improve the usability of the estimation with real-time users as well as assimilation quality inside the Kalman filter.

The main purpose of this study is to investigate short term-forecasting of global VTEC maps based on Trigonometric B-splines targeting mostly for real-time users and data assimilation of Ionospheric state. The B-spline based VTEC modeling is compared to SH under different data distributions and benefits are shown [18]. Similar advan-

tages may also be obtained in geomagnetic storm conditions with proper short-term forecasting and data assimilation.

Trigonometric B-spline coefficient time series used in this study is obtained by using 20 years of final IGS GIMs starting from year 2000. For every hour in each day, the grid VTEC in the IGS GIMs are used as measurements to estimate Trigonometric B-spline coefficients for that hour in solar geomagnetic reference frame. The time series spanning 20 years are then used for studing different forecasting methods developed throught the study. On of the methods handled in this study is based on prediction of individual coefficients using LSTM, which is a deep learning method [19]. The forecasting results of LSTM model for individual coefficients are compared with the Seasonal Autoregressive Integrated Moving Average (SARIMA) [20]. A Block-based LSTM model is also proposed to improve forecasting quality of coefficients representing highly varying equatorial region. In this method, for each coefficient, neighboring B-spline coefficients are also given as input to the training process. Additionally, dimension reduction methods based on PCA have also been studied with both LSTM and SARIMA based models. Finally, developed models are compared with each other in quiet and highly active storm conditions of ionosphere.

1.2 Objectives

The main goal of the thesis is to investigate short-term global Ionosphere forecasting by examining state-of-the-art methods including LSTM, SARIMA and PCA. In order to achieve these goal, the following objectives are determined:

- Time series of Trigonometric-Bspline coefficients for global VTEC from 20 years of IGS VTEC maps are required.
- Forecasting method should be investigated using state-of-the-art methods namely LSTM, SARIMA and PCA.
- Models for forecasting of individual Bspline coefficients should be developed.
- Models for forecasting all B-spline coefficients with PCA based dimension reduction should be investigated.

- Testing and validating should be done both in terms of forecasting quality of individual coeffients and also the resulting VTEC maps.

- Comparison of the success of the proposed methods with a discussion of future work should be listed.

1.3 Methodology

The methodology part contains mainly three stage:

- Downloading gobal IGS VTEC maps from 2000 to 2020. Using the iontrace [3] library, time series of Trigonometric B-spline coefficients will be generated.

- Single coefficient forecasting based on LSTM and SARIMA models will be developed. Due to the computing resources available (Laptop, powered by Intel i7 8th generation processor with 16 GB RAM), the initial model selection is based on one year of data (2013).

- Apply PCA based dimension reduction to the VTEC coefficients and apply LSTM and SARIMA models to Principal Components. Due to the computing resources available, the initial model selection is based on one year of data (2013).

- Select best model for single coefficient forecasting and PCA based forecasting methods using validation errors of individual models.

- Re-train the best models with a longer dataset covering 2013-2015 including high storm conditions during 2015.

- Compare the re-trained best models in 2018 for both quiet and storm days based on coefficients and resulting maps.

- List conclusions and futurework directions.

1.4 Outline

In Chapter 1 Introduction the motivation, objectives, methodology and outline are given. In Chapter 2 Forecasting of Ionosphere contains Structure and Variations of Ionosphere; observation methods are briefly listed and how to extract TEC from GNSS, the Ionosphere models, how to create IONEX file, B-splines basis function and lastly forecasting methods which are statistical methods and neural networks are described. Chapter 3 , Methodology, describes preparing B-Spline Coefficient time series, forecasting of B-spline Coefficients, LSTM based and conventional forecasting methods, and lastly validation. Chapter 4 Application presents results of the study. The manuscript ends with the last Chapter 5 Conclusion. This chapter contains summary of the study, results and future work.

CHAPTER 2

FORECASTING OF IONOSPHERE

This chapter provides a brief summary of background information both for Ionosphere and its forecasting. The first section contains information on the structure and variations of the Ionosphere; the second section presents observing of the Ionosphere; the third section consists of the Ionosphere models, and its forecasting methods are described in the last section.

2.1 The Ionosphere

The Ionosphere is a layer of top atmosphere, located with ranging altitude between 50 km and 1000 km from the surface of the Earth. It mainly consists of ionized gases and free electrons mostly due to sunlight in Ultra Violet (UV) band. The Ionosphere is generally divided into regions with respect to the quantity of free electrons changing with height.

2.1.1 The Structure of the Ionosphere

Variations of electron density within the ionosphere represented by different regions depending on altitude. These layers are named D, E and F region. The D region, with the least ionization, is at the bottom of the Ionosphere which is within the altitude interval of 50 km to 90 km from the surface of the Earth. As the height increases, amount of the free electrons increase in the Ionosphere. Ionization in this layer takes place with high energy X-rays. While the ionization occurs in D region right after sunrise, the layer disappears at night by the recombination of free electrons and ions.

Region E is located between altitudes of 90 km and 140 km where ionization is greatly reduced but does not disappear fully at nights. The highest layer of the Ionosphere which is the F region is starts at an altitude of 150 km. This region is the most important in terms of radio communications and navigation systems. While this region is a single layer during the night, it is separated into F1 and F2 during the daylight.

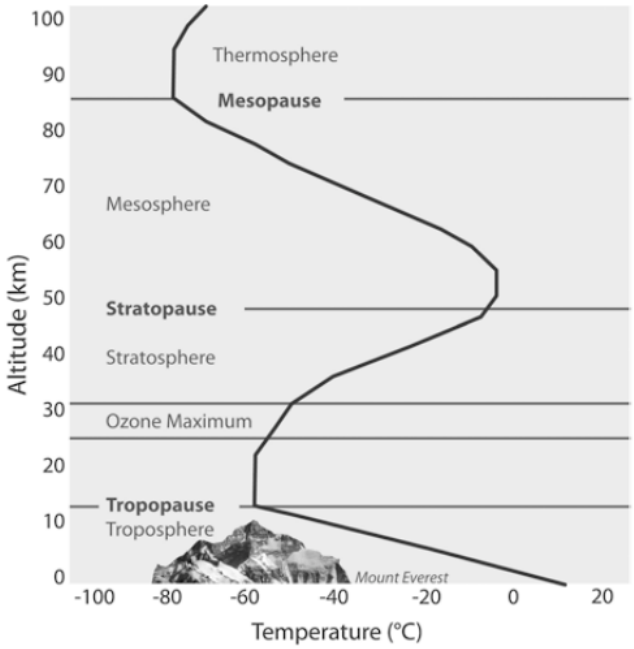

Figure 2.1: Regions of Atmosphere with changing temperature profile [1]

The F1 region can be found at altitude approximately 140-210 km which changes with solar activity, season as well as geomagnetic activity. The F2 layer is a highly variable layer with the highest ionization. Its depth and intensity varies with time, season, and sunspot activity. The highest electron density through the Ionosphere is in the F2 layer. This layer does not disappear day or night. The Figure 2.2 shows the these regions within day and night. The $foF2$ is known as the critical frequency corresponding to the peak of the F2 layer [1].

2.1.2 Variations of the Ionosphere

The most important temporal variations within the Ionosphere is due to the daily motion of the Sun. The number of free electrons in the ionosphere are also affected

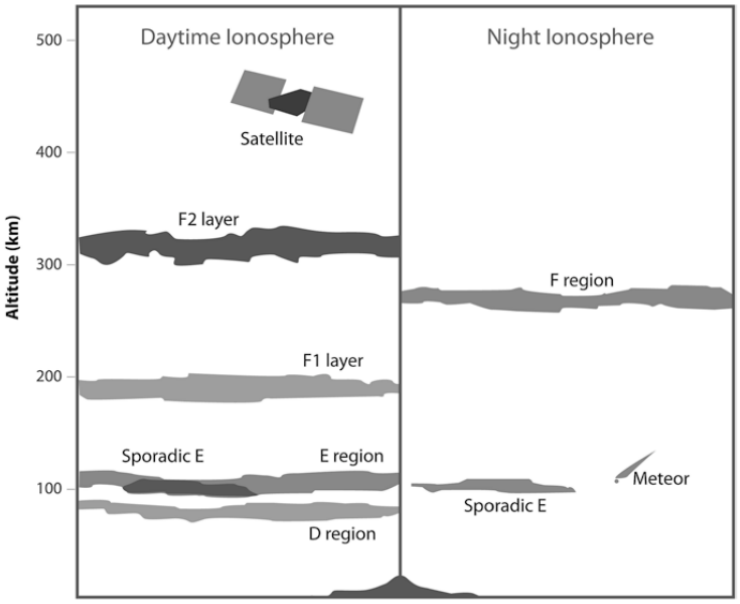

Figure 2.2: Regions within the Ionosphere [1]

by solar flares, sunspots and the changes in the geomagnetic field. Spatial variations within the Ionosphere can be divided by high latitude, mid-latitude and regions near the Equator. In The high and mid-latitudes variations are generally smaller than in other regions. Periodic changes occur in the amount of electrons within the Ionosphere. The ionization amount differs day and night mainly because of the sunlight. The ionization in day is high, while it is low at nights. Also there are seasonal changes depending on the incidence of the the sunlight. The electron content is low in summer and winter and while it is high in spring and autumn. The Ionospheric changes may also be caused by some factors such as geomagnetic storms, sunspots and solar activity. Solar Flux $\Phi 10.7$ can be used as an indicator of solar activity. Sunspots are highly concentrated magnetic fields on the sun's surface. Periodic changes in solar energy is on the average accounts for a 11-year cycle which is named solar cycle. Geomagnetic storms happen on Earth due to Coronal Mass Ejections Shots and affect the global magnetic field and the Ionosphere [1]. Geomagnetic activity can be monitored by and index combined from Ds, the auroral electrojet activity and DST, the ring current in the magnetosphere. There is also Kp and Ap indexes which may be used to indicate the geomagnetic activity. These indices may be used to evaluate geomagnetically disturbed days.

2.2 Observing Ionosphere

Methods based on the refraction of electromagnetic waves by the Ionosphere can be used to sound electron density or content. Some of those are Ground-based GNSS, Occultation, Satellite altimetry and Ionosonde observations.

high spatial and temporal resolution TEC measurements of the ionosphere can be obtained with GNSS. GNSS receivers get coded information regularly sent by satellites, providing the receiver's position on Earth by measuring the distance between the receiver and the satellite. However, significant delays are caused by free electrons in the Ionosphere. For this reason, the GNSS is designed to provide ranging signals in more than one frequency. The receiver can then use differential techniques to either mitigate Ionospheric delays or get a measurement of delays due to Ionosphere by geometry-free linear combination [4].

Although ground based GNSS provides valuable information about Ionosphere, it can only measure integrated electron density along a slant path. Recently, occultation measurements between satellite pairs can be used to probe ionosphere from different angles by dedicated satellite missions [21]. Horizontal scanning of the ionosphere can be done with the COSMIC-I and COSMIC-II satellites. Satellite altimetry missions such as of Jason may detect changes in sea level as well as directly measure vertical TEC [22]. Ionosonde stations emit varying frequency of radio waves for vertical sounding of the Ionosphere where critical parameters of the F region may be measured [1].

2.2.1 Extracting TEC from GNSS Measurements

The Ionosphere causes delays in the transmission of radio signals. It can be shown that the delay of electromagnetic waves ın the Ionosphere is related with the slant TEC along the path of propagation [23]. In this case, the STEC can be defined as:

$$STEC = \int_r^s N_e(s)\, ds \tag{2.1}$$

Where N_e is the electron density measured in electrons per m^2 along the path ranging from the satellite s to the GNSS receiver r. The Vertical TEC (VTEC) can be obtained

from STEC by [2]:

$$VTEC = \frac{1}{m(z)} STEC \tag{2.2}$$

In Eq. 2.2, $m(z)$ is mapping function depending on zenith z angle as shown Fig. 2.4 utilizing namely the single-layer model [2]. The Single-layer model, assumes that the Ionosphere is a thin layer at a certain height. Zenith z and Ionospheric pierce point (IPP) are found from the geometry between receiver and the satellite. Then, the mapping function can be calculated as:

$$m(z) = \frac{1}{\cos z'} = \frac{1}{\sqrt{(1 - \sin^2 z')}} \tag{2.3}$$

Dual frequency measurements obtained through GNSS (such as GPS and GLONASS) may provide TEC measurements based on pseudo-ranges or carrier phases. The geometry-free linear combination of measurements from two frequencies removes geometry induced errors. Measurements of pseudo-ranges are much more noisy but they are absolute. On the other hand, the carrier-phase measurements are very precise but suffers from cycle-slips and integer ambiguities. The Ionospheric observable P and Φ derived from combinations of the pseudo-ranges and carrier-phases respectively can be written as [3]:

$$P^s_{r,f_1} - P^s_{r,f_2} = \alpha STEC + DCB_r + DCB_s + \epsilon_P \tag{2.4}$$

$$\Phi^s_{r,f_2} - \Phi^s_{r,f_1} = \alpha STEC + B_r + B_s + C^s_{arch,r} + \epsilon_\Phi \tag{2.5}$$

where

- $P^s_{r,f_1} - P^s_{r,f_2}$ and $\Phi^s_{r,f_2} - \Phi^s_{r,f_1}$ are geometry-free linear combination of pseudo-range and carrier-phase measurements, respectively.
- P and Φ are pseudo-range and the carrier-phase measurements observed from satellite s to receiver r.
- f_1, f_2 are frequencies.

- B_r and B_s are inter frequency biases of receiver and satellite, DCB_r and DCB_s are differential code biases of receiver and satellite.
- α is constant depending of frequencies.
- $C^s_{arch,r}$ is the combined ambiguity bias of the carrier-phases.
- ϵ_P and ϵ_Φ are measurement errors of code based and phase based combinations.

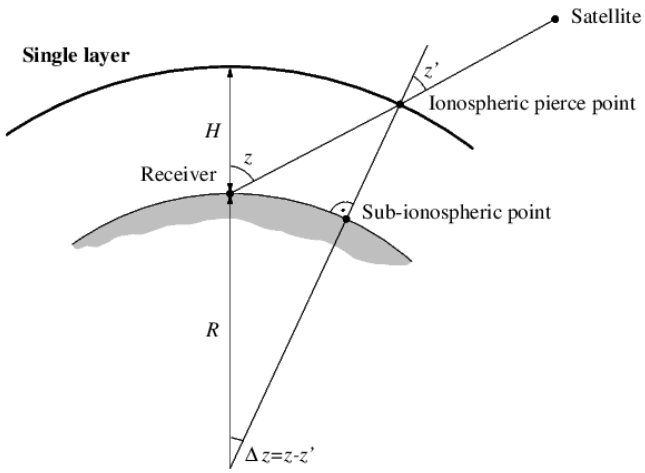

Figure 2.3: Single Layer Model [2]

Generally, leveling of carrier phase based geometry free linear combination is preferred as measurements of TEC since they have higher precision. The leveling procedure estimates the Common phase bias defined as:

$$CPB^r_s \approx \frac{1}{N}\sum_{j=1}^{N}(\Phi^s_{r,f_2} - \Phi^s_{r,f_1} - P^s_{r,f_2} - P^s_{r,f_1})_j \tag{2.6}$$

where N is the number of observations measured along a continuous arc without cycle-clips. Then, the leveled-geometry free phase observation can be computed as [5]

$$\tilde{L}^s_{r,4} = \Phi^s_{r,f_2} - \Phi^s_{r,f_1} - CPB^r_s = \alpha STEC + b_r + b^r + \epsilon_{L_4} \tag{2.7}$$

2.3 Ionosphere Models

Although the Ionosphere effects can be eliminated using dual frequency ionosphere-free combination, many receivers still operates only with single frequency. Ionosphere models are need for the dissemination of correction information for single frequency users as well as monitoring and prediction of the ionosphere. Many models have been developed for the representation of Ionospheric state ranging from physics-based models such as data assimilation to various empirical models based on mathematical models coefficients of which are estimated from measurements.

Maybe the mostly used global Ionosphere models are International Reference Ionosphere (IRI) and Klobuchar Model. Klobuchar model is used for ionosphere correction on single frequency GPS users and distributed in the navigation message. The IRI, and also the NeQuick model are analytical functions developed from various observations [1]. Global Ionosphere Maps (GIM) are VTEC maps are generated by various IGS analysis centers such as Center for Orbit Determination in Europe and Jet Propulsion Laboratories having different models for the spatio temporal variations of the VTEC. There are also regional and local Ionosphere models.

2.3.1 Coordinate Systems

In the sun fixed coordinate system, mean coordinates are shifted to the left according to the position of the sun.

$$\lambda_{SF} = \lambda - \lambda_0 \tag{2.8}$$

where λ_{SF} is the longitude in the sun fixed coordinate system. λ_0 is the longitude of the mean sun [3].

In the sun fixed geomagnetic coordinate system, not only the longitude is shifted, but also the north pole aligned the geomagnetic dipole axis (λ_P, φ_P), which may be implemented as a simple rotation.

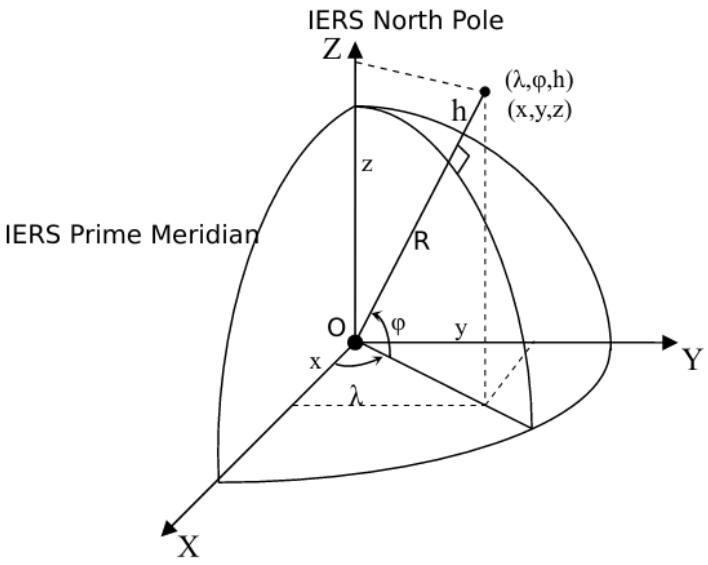

Figure 2.4: Spherical Coordinates [3]

$$\lambda_m = \lambda_{SF} \tag{2.9}$$

$$\varphi_m = \arcsin \sin \varphi \sin \varphi_P + \cos \varphi \cos \varphi_P \cos (\lambda \lambda_P) \tag{2.10}$$

where are φ_m, λ_m are latitude and longitude in sun fixed geomagnetic coordinate system, respectively.

2.3.2 International Reference Ionosphere

IRI is a physical and experimental ionosphere model which is also accepted to be an ISO standard (ISO:16457)[24]. IRI is developed by a international working group supported by Committee on Space Resarch (COSPAR) and the International Union of Radio Science (URSI). It not only provides TEC information, but also the electron density density predictions. In this regard, IRI can provide 3D Ionosphere model. IRI estimates are good in areas with high measurement density. An important point for IRI is that studies and testing are ongoing. The IRI model is being continuously updated with recent data and improved mathematical models of global and temporal variations within Ionosphere. Numerous independent studies have validated the IRI by comparing with independent measurements [25].

IRI can be used to obtain electron density, TEC, temperature and other key physical parameters of Ionosphere for a specific datetime and location. Data sources included in the modeling of IRI are mostly solar parameters, ionosonde and incoherent scatter radars [1].

The web-based software for IRI-2016 model can be found on the `https://ccmc.gsfc.nasa.gov/modelweb/models/iri2016_vitmo.php`. Fortran code can be accessed on this website: `http://irimodel.org/IRI-2016/`.

2.3.3 The NeQuick Model

The NeQuick is a time dependent electron density model which is a fast-running model specifically designed for ionospheric applications, allowing the electron concentration to be calculated at any location in the Ionosphere [26]. NeQuick 2 is the latest version of the NeQuick ionosphere electron density model which is continuously tested and evaluated [26].

NeQuick 2 web model can be accessed on the web: `https://t-ict4d.ictp.it/nequick2/nequick-2-web-model`.

2.3.4 Spherical Harmonic Model

The most frequently applied basis expansion for global functions is the SH expansion. The global distribution of VTEC is modeled by a series of coefficients for SH expansion [2].

$$VTEC(\varphi,\lambda) = \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} \tilde{P}_{nm}(sin\varphi)(a_{nm}cos(m\lambda) + b_{nm}sin(m\lambda)) \tag{2.11}$$

where φ and λ are geomagnetic latitude and sun-fixed longitude. The degree and order of the model is given as n and m. The coeffients a_{nm} and b_{nm} are the coefficients of the model and $\tilde{P}_{nm}$ are the normalized associated Legendre Polygnomials. The unknown parameters of the model are estimated within a least square adjustment procedure utilizing ground based GNSS measurements [2].

2.3.5 Trigonometric B-Spline Model

Global VTEC may also be expressed as a series expansion in terms of Trigonometric B-splines. The Series expansion of B-spline function is obtained by a tensor product of polynomial and Trigonometric B-splines [5].

$$VTEC(\varphi,\lambda) = \sum_{k_1=0}^{K_{J_1}-1} \sum_{k_2=0}^{K_{J_2}-1} d_{k_1,k_2}^{J_1,J_2} N_{J_1,k_1}^2(\varphi) T_{J_2,k_2}^2(\lambda) \tag{2.12}$$

where $d_{k_1,k_2}^{J_1,J_2}$ are unknown series coefficients, $N_{J_1,k_1}^2(\varphi)T_{J_2,k_2}^2(\lambda)$ are interpolating polynomial B-spline functions on the solar-geomagnetic latitude φ and trigonometric B-splines on the solar-geomagnetic longitude λ, respectively. J_1 and J_2 are levels, k_1 and k_2 are geometrical positions. K_{J1} is the number of polynomial B-spline functions, which is equal to $2^{J_1}+2$. K_{J2} is the number of trigonometric B-spline functions, which is equal to $3\text{x}2^{J_2}+2$.

The parameters of the model are estimated within a Kalman Filter or a least squares estimation procedure combining ground based GNSS measurements.

2.3.6 IGS Global Ionosphere Maps

IGS has been working to produce ionospheric products on a daily basis since 1999. The products are disseminated in Ionosphere Exchange Format (IONEX) format as a global grid of VTEC values. IGS VTEC maps are obtained by weighting the maps produced by Ionospheric Associate Analysis Centers (IAAC). Each of the analysis center uses different methods to produce an estimated global VTEC in IONEX format.

The daily data collected by the globally distributed reference stations goes to the RINEX directories of IGS. Global ionosphere maps are produced using RINEX files which have carrier-phase and pseudo-code data of GNSS stations spread around the world. Computational centers of CODE, ESA, JPL etc. download RINEX and use their own ionosphere modeling software. STEC values obtained from the RINEX file are used to estimate grid VTEC where both satellite and receiver DCBs are also estimated.

IGS final ionosphere maps are created by combining data from individual IAAC according to [6]:

$$I_i = \tilde{f}_i(t, \lambda, \phi) + \epsilon \tag{2.13}$$

where I_i is the Ionosphere model and i is the index of computational centers, $\tilde{f}_i$ is each center's own model. For example $\tilde{f}_i$ can be seen as SH expansion for CODE.

$$f_{igs} = \sum_{i=1}^{N} \frac{w_i \tilde{f}_i(t, \lambda, \phi) + \epsilon}{\sum w_i} + \epsilon_{igs} \tag{2.14}$$

IGS Ionosphere maps f_{igs} are produced with the weighted averages of those coming from each center. In this way, IGS maps are improved in terms of accuracy. After the centers produce the functions $\tilde{f}_i$. The IONEX file consists of raster Ionosphere map created in the form of a grid per hour. Thus, IONEX files are created for each center. In addition, IGS uses JASON altimetry data not available in the estimation procedure of the analysis centers. The IONEX file use a spherical Earth model.

One IONEX file contains grid global VTEC with a resolution 2.5°x5°. The file also lists the number of the satellites and the receivers used in the estimation procedure with estimated differential code biases for the satellites and the receivers. Temporal resolution is one hour. The latitude and longitude ranges of the IONEX files are given as 87.5° / -87.5° in geographic latitude and -180° / 180° geographic longitude respectively.

Each IONEX file consists of a header and a data sections. The header contains common information record for the whole file [27]. An example of the IGS IONEX file is shown in Fig. 2.5 and 2.6. The Fig. 2.5 presents header of the file and Fig. 2.6 shows the end of the file. The IONEX is combined IGS TEC maps as shown Fig. 2.5.

Interpolation being temporally and spatially can be applied to find a map between two times. Temporal and spatial VTEC values are calculated by interpolation [27]. Suppose VTEC maps are given:

```
     1.0            IONOSPHERE MAPS     MIX                 IONEX VERSION / TYPE
cmpcmb v1.2          GRL/UWM             19-may-18 04:48     PGM / RUN BY / DATE
ionex file containing IGS COMBINED Ionosphere maps           COMMENT
global ionosphere maps for day 120, 2018                     DESCRIPTION
IONEX file containing the COMBINED IGS TEC MAPS and DCBs     DESCRIPTION
      IONEX files of the following IAACs were combined: cod DESCRIPTION
                                                         jpl DESCRIPTION
Contact address: Andrzej Krankowski                          DESCRIPTION
                 Geodynamics Research Laboratory             DESCRIPTION
                 University of Warmia and Mazury (GRL/UWM)   DESCRIPTION
                 Oczapowski St. 1                            DESCRIPTION
                 10-957-Olsztyn, POLAND                      DESCRIPTION
                 e-mail:  kand@uwm.edu.pl                    DESCRIPTION
                                                             DESCRIPTION
  2018     4    30     0     0     0                        EPOCH OF FIRST MAP
  2018     5     1     0     0     0                        EPOCH OF LAST MAP
  7200                                                       INTERVAL
    13                                                       # OF MAPS IN FILE
  COSZ                                                       MAPPING FUNCTION
     0.0                                                     ELEVATION CUTOFF
combined TEC calculated as weighted mean of input TEC valuesOBSERVABLES USED
   345                                                       # OF STATIONS
    32                                                       # OF SATELLITES
  6371.0                                                     BASE RADIUS
     2                                                       MAP DIMENSION
   450.0 450.0   0.0                                        HGT1 / HGT2 / DHGT
    87.5 -87.5  -2.5                                        LAT1 / LAT2 / DLAT
  -180.0 180.0   5.0                                        LON1 / LON2 / DLON
    -1                                                       EXPONENT
TEC values in 0.1 tec units;  9999, if no value available   COMMENT
DCB values in nanoseconds, reference is Sum_of_SatDCBs = 0  COMMENT
DIFFERENTIAL CODE BIASES                                     START OF AUX DATA
   G01    -7.812     0.019                                  PRN / BIAS / RMS
   G02     9.181     0.043                                  PRN / BIAS / RMS
   G03    -5.669     0.019                                  PRN / BIAS / RMS
   G04     3.933     0.169                                  PRN / BIAS / RMS
   G05     3.073     0.060                                  PRN / BIAS / RMS
   G06    -7.019     0.032                                  PRN / BIAS / RMS
   G07     3.197     0.008                                  PRN / BIAS / RMS
   G08    -7.387     0.014                                  PRN / BIAS / RMS
   G09    -4.955     0.031                                  PRN / BIAS / RMS
   G10    -5.502     0.025                                  PRN / BIAS / RMS
```

Figure 2.5: Top Overview of IGS IONEX MAP in 19 May 2018

```
   G  yar2 50107M004          -3.705     0.000             STATION / BIAS / RMS
   G  yar3                     4.312     0.000             STATION / BIAS / RMS
   G  yarr 50107M004           1.462     0.016             STATION / BIAS / RMS
   G  yell 40127M003           6.819     0.008             STATION / BIAS / RMS
   G  yssk 12329M003         -13.970     0.009             STATION / BIAS / RMS
   G  zamb 34601M001          -2.417     0.000             STATION / BIAS / RMS
   G  zeck 12351M001           8.577     0.115             STATION / BIAS / RMS
   G  zim2                   -12.865     0.000             STATION / BIAS / RMS
   G  zim3                   -13.974     0.000             STATION / BIAS / RMS
   G  zimj 14001M006           8.052     0.000             STATION / BIAS / RMS
   G  zimm 14001M004         -11.661     0.000             STATION / BIAS / RMS
DIFFERENTIAL CODE BIASES                                   END OF AUX DATA
                                                           END OF HEADER
    1                                                      START OF TEC MAP
 2018     4    30     0     0     0                       EPOCH OF CURRENT MAP
   87.5-180.0 180.0   5.0 450.0                           LAT/LON1/LON2/DLON/H
  68   69   69   69   70   70   70   70   70   70   71   71   70   70   70   70
  70   70   70   70   69   69   69   68   68   68   67   67   67   66   66   65
  65   65   65   65   65   64   64   64   64   64   64   64   64   64   64   64
  64   64   64   64   65   65   65   65   65   65   65   65   65   66   66   66
  67   67   67   67   68   68   68   68   68
   85.0-180.0 180.0   5.0 450.0                             LAT/LON1/LON2/DLON/H
  69   70   70   71   71   72   73   73   74   74   74   74   75   74   74
  74   74   73   72   71   71   70   69   68   67   66   65   65   64   63   63
  62   62   62   61   61   61   61   61   61   61   61   62   62   62   62   62
  63   63   63   63   63   63   63   63   63   64   64   64   64   65   65   65
  65   66   66   67   67   67   68   69   69
   82.5-180.0 180.0   5.0 450.0                             LAT/LON1/LON2/DLON/H
  69   70   71   72   72   73   74   75   76   77   77   78   78   78   78   78
  78   77   77   75   74   73   71   70   68   66   64   63   62   60   60   58
  58   57   57   57   57   57   57   58   58   59   59   60   61   61   61   62
  61   62   62   62   62   62   62   62   62   63   62   63   63   63   63   64
  64   65   65   65   66   66   67   68   69
   80.0-180.0 180.0   5.0 450.0                             LAT/LON1/LON2/DLON/H
  69   70   71   72   73   74   75   76   77   79   79   80   81   81   82   82
  81   81   79   78   76   74   72   70   67   65   62   60   58   56   55   53
  53   52   51   52   52   53   53   54   55   56   57   58   59   60   60   61
  61   61   61   61   61   61   61   61   62   61   62   61   62   62   62   62
  63   63   64   64   65   66   67   68   69
```

Figure 2.6: Mid Overview of IGS IONEX MAP in 19 May 2018

$$VTEC_k(\lambda, \varphi) = VTEC(\lambda, \varphi, t_k) \tag{2.15}$$

where λ is the geographic latitude, φ geographic latitude and t_k is the time in UT.

Temporal interpolation is given as:

$$VTEC(\lambda, \varphi, t) = \frac{t_{k+1} - t}{t_{k+1} - t_k} VTEC_k(\lambda, \varphi) + \frac{t - t_k}{t_{k+1} - t_k} VTEC_{k+1}(\lambda, \varphi) \qquad (2.16)$$

where $t_k <= t < t_{k+1}$. The spatial interpolation scheme uses bi-linear interpolation of the four nearest grid points to obtain a VTEC for a specific point.

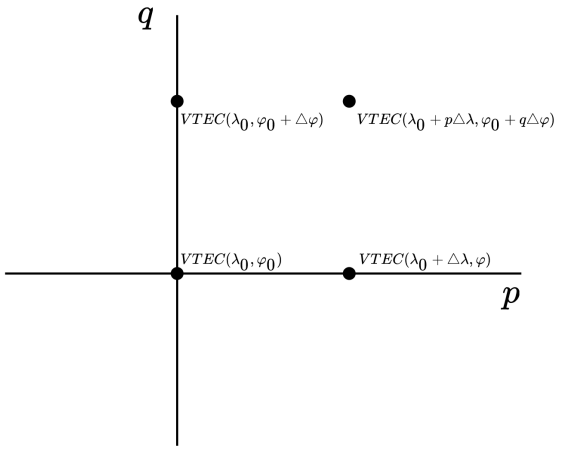

Figure 2.7: IONEX Spatial Interpolation [2]

2.3.7 Assimilation Models

Global Assimilation of Ionospheric Measurements (GAIM) is a physics based data assimilation model developed by Utah State University Global GAIM and uses a Kalman filter basically assimilating a variety of real-time (or near-real-time) measurements [28]. GAIM consists of Ionospheric Forecast Model (IFM), Ionosphere–Plasmasphere Model for simulation, Kalman filters for data assimilation providing 3D electron density distribution between 90 and 25,000 km for global and regional forecasts [1].

2.3.8 Local and Regional Ionosphere Models

There are numerous studies working on local TEC estimation for example by [12; 13; 14; 15]. Regional maps of TEC over Europe are produced by technique of bezier surface-fitting using neural networks [10]. A study was done using a tensor product

spline expansion regionally [29]. Another study focuses on local ionosphere prediction using single station VTEC values [12].

2.4 Forecasting Methods for Ionosphere

Forecasting of the Ionosphere is a topic of high research interest [1]. Ionospheric forecasting consist of forecasting for solar and geomagnetic activity, ionospheric storm and other parameters of Ionosphere. Near real-time ionospheric forecasting is required for radar and surveillance applications or mobile communications. In addition, short term forecasting is important in aviation, positioning and navigation. In this study, the main focus of forecasting is one hour ahead forecasting using the historical time series of VTEC parameters. In this regard, two main methods of many are listed here namely the statistical methods and black box methods based on neural networks.

2.4.1 Statistical Methods

A time series can be considered as a list of data points lined up over a specified period of time, where time is an independent variable. There are the following components in a time series:

- a trend : time dependent increasing or decreasing of values
- a seasonal component : periodic variations of known frequency
- noise : various random effects mostly described by stochastic models
- level : an average value or constant offset

The trend in a time series can be expressed as a polynomial [20]. It can also be modeled with a series of sine and cosine of constant amplitude and specific frequency. A recently developed method (and tool named prophet) based on generalized linear models can effectively model linear and periodic trend in time series [30].

A weak-sense stationary time series is generally described by a constant mean, finite variance with covariance independent of time. For example, white noise is a stationary

time series while a time series which has a linear trend is not stationary. If X_t is stationary, $E[X_t] = \mu$ is independent of t and autocovariances $\text{Cov}(X_{t+k}, X_t)$ depend only on time lag k for all t. This is called weak-stationarity [20].

Autocorrelation defines the linear relationship between lagged values in a time series. Autocorrelation Function (ACF) may be important in determining the orders of lagged interactions in a time series. It can show how the values with specific lags correlate with each other. An autocorrelation plot can reveal a trend in the time series. If a time series has no autocorrelation, it can be specified as white noise which is an independent and identically distributed random process.

Autocovariance between X_t and X_{t+k} can be calculated by [20]:

$$Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] \tag{2.17}$$

where k is the lag. Depending only on the time lag, autocovariance function can be defined as:

$$\gamma(k) = E[(X_t - \mu)(X_{t+k} - \mu)] \tag{2.18}$$

Then, autocorrelation function for a stationary time series can be written as:

$$\rho(k) = \frac{\gamma(k)}{\sqrt{\gamma(0)\gamma(k)}} = \frac{\gamma(k)}{\gamma(0)} \tag{2.19}$$

Different methods such as multiple linear regression, the autocorrelation method, moving average (MA), autoregressive (AR) models, autoregressive moving average (ARMA) are available for stationary time series analysis and modeling.

Autoregressive (AR) model applies linear combination of previous data for the prediction of the next sample. If current observation can be written as a weighted combination of previous observations with some random error then it is called AR. In other

words, there is correlation between current and previous observations. The model is called AR(p) where p is the order of the model [31]:

$$\tilde{X}_t = X_t - \mu \tag{2.20}$$

where $\tilde{X}_t$ is current observation is regressed. The mean is assumed to be constant where assumption of constant variance and mean comes from stationarity [20].

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + ... + \phi_p \tilde{X}_{t-p} + Z_t \tag{2.21}$$

where Z_t is noise, ϕ are the weights. The Partial Autocorrelation Function Function (PACF) can be use to select order p. PACF finds correlation between current value and residual of previous lags.

Moving Average utilizes the previous errors to predict the next value. It is called MA(q), q is order [31]:

$$\tilde{X}_t = Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - ... - \theta_q Z_{t-q} \tag{2.22}$$

where Z_t is white noise, and θ is weight. The ACF can be utilized to select the order q. Autoregressive Moving Average consists of AR(p) and MA(q) which can be written as ARMA((p,q).

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + ... + \phi_p \tilde{X}_{t-p} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - ... - \theta_q Z_{t-q} \tag{2.23}$$

Autoregressive Integrated Moving Average Model (ARIMA) is generally applied for non-stationary time series. The stationarity of non-stationary series can be achieved by performing a series of difference operations suitable for operation of the ARIMA model. It is called AR(p)I(d)MA(q) or ARIMA(p, d, q) [32]:

$$\tilde{X}_t = (\tilde{X}_t - \tilde{X}_{t-1}) - (\tilde{X}_{t-1} - \tilde{X}_{t-2}) - .. - (\tilde{X}_{t-d+1} - \tilde{X}_{t-d}) \tag{2.24}$$

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + ... + \phi_p \tilde{X}_{t-p} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - ... - \theta_q Z_{t-q} \tag{2.25}$$

Eq. 2.24 provides differentiating from the non-stationary data. Eq. 2.25 is the same ARMA's formula provided before.

The method shall be chosen depending on the data at hand. It is important to detect the stationarity as well as seasonality in the dataset.

The auto regressive models can be used in stationary data or with pre elimination of trend from the dataset. It has been used to 24 hour ahead prediction of $foF2$ [1].

$$\Phi = \frac{foF2 - foF2_{med}}{foF2_{med}} \tag{2.26}$$

where Φ is random process over time. where the non-stationary part $foF2_{med}$ is eliminated from the dataset. So autoregressive model can be of the form:

$$\Phi(t_0) = \sum_{k=1}^{n} \beta_k \Phi(t_0 - \tau_k) \tag{2.27}$$

where n is order of AR, β is the weight.

2.4.1.1 Seasonal autoregressive integrated moving average

Seasonal autoregressive integrated moving average (SARIMA), which is Box-Jenkins methodology, is a statistical method and extension of ARIMA. SARIMA is capable of non-seasonal data. SARIMA contains two parts which are non-seasonal part, seasonal part.

$$SARIMA(p,d,q)(P,D,Q)_s \tag{2.28}$$

As seen from the equation, SARIMA has seven parameters. p is the order of the non-seasonal autoregressive model, d is the number of non-seasonal differences, q is the order of non-seasonal moving average model, P is the order of seasonal autoregressive model, D is the number of seasonal differences, Q is the order of seasonal moving average model and s is the periodic term.

$$\Phi_P(B^s)\phi_p(B) \bigtriangledown_s^D \bigtriangledown^d X_t = \Theta_Q(B^s)\theta_q(B)_t \tag{2.29}$$

where Φ_P, Θ_Q and $\bigtriangledown_s^D$ are seasonal operators autoregressive, moving average and differencing respectively, ϕ_p, θ_q and $\bigtriangledown^d$ are non-seasonal operators autoregressive, moving average and differencing respectively, Z_t is white noise.

Suppose the model orders are chosen as $SARIMA(0,1,1)(0,1,1)_{12}$, the fitted time series can be shown as [20]:

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + Z_t - \theta_1 Z_{t-1} - \Theta_1 Z_{t-12} + \Theta_1\theta_1 Z_{t-13} \tag{2.30}$$

$\mathrm{X}_{t+l} = X_{t-1+l} + X_{t-12+l} - X_{t-13+l} + Z_{t+l} - \theta_1 Z_{t-1+l}$
$- \Theta_1 Z_{t-12+l} + \Theta_1\theta_1 Z_{t-13+l}$

where l is future ahead time.

$\hat{X}_t(l) = [\mathrm{X}_{t-1+l}] + [X_{t-12+l}] - [X_{t-13+l}] + [Z_{t+l}] - \theta_1[Z_{t-1+l}]$
$- \Theta_1[Z_{t-12+l}] + \Theta_1\theta_1[Z_{t-13+l}]$

where $\hat{X}_t(l)$ is forecasted value ahead with respect to l.

$[X_{t+l}]$ and $[Z_{t+l}]$ describes conditional expectations (mean) to X_{t+l}, Z_{t+l} any t time. $[Z_{t+j}]$ is assumed zero because Z_{t+j} has not occurred yet for $j > 0$. But if $j <= 0$ is, Z_{t+j} has occurred and is applied. Since Z_{t+j} cannot be observed by itself, this can be

calculated with equation $[X_{t+j}] = X_{t+j} - \hat{X}_{t+j-1}$ for $j <= 0$. If the time of observation has already passed, $[X_{t+l}] = X_{t+l}$, for $j <= 0$. If time t is in future, $[X_{t+j}] = \hat{X}_t(l)$, for $j > 0$.

Augmented Dickey-Fuller (ADF) statistical test can be applied to test the stationarity of a time series. If the dataset is not stationary the Null hypothesis is accepted. Thus, we can also tell if it is stationary with the selected p-value threshold [31].

To determine all of the orders of a SARIMA model, partial autocorrelation function (PACF) and autocorrelation function (ACF) can be utilized. In addition, Akaike's information criterion (AIC) can be used for model selection in a hyper parameter search procedure. The model with the smallest AIC value is usually chosen as the best model [20].

$$AIC = -2 \ln \hat{\sigma}_a^2 + 2r \tag{2.31}$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of the residual variance, r is the number of parameters estimated in the model including a possible constant term.

There are some studies in the literature that can estimate VTEC using a statistical models. For example [8] proposed that global VTEC maps were forecast by adaptive AR with spherical harmonic (SH) coefficients produced by CODE. Another model is proposed by [33] where the predicted TEC is compared with the TEC provided by IGS. Local and region models can be estimated using ARIMA model [32]. This study compared with conventional model ARIMA and deep learning method. Another study used least-squares collocation technique to determine stochastic and deterministic part of the prediction model in a least squares adjustment procedure where with a correlated noise assumption, the noise characteristic is estimated by detrending and fed back [2].

2.4.2 Neural Network Methods

Neural Networks (NN) which may be used for linear and non-linear model estimation can be good at modeling complex relationships in the data. VTEC data or $foF2$

parameter of Ionosphere are non-linear and also have periodic components. A Neural network is a set of connected neurons as shown in 2.9.

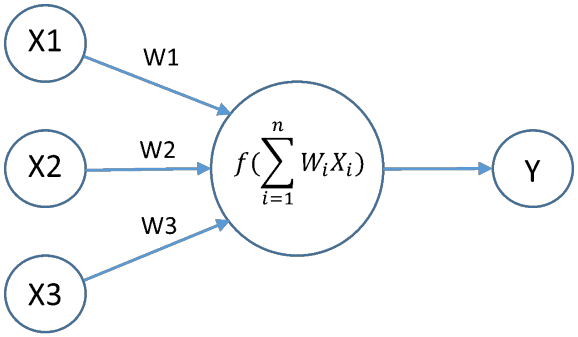

Figure 2.8: A model of a Neuron (`https://i.stack.imgur.com/gzrsx.png`)

A neural network is constitute of layers of neurons, an input layer, one or more hidden layers and an output layer. Each layer of neurons receives its input from the previous layer and output of each neuron feeds the following connected layer. The neuron provides the output according to an activation function value of the weighted sum of its inputs. Thus, finding the optimal weights $W_{1,2,..,n}$, in a network of neurons, complex relationships between the inputs and outputs may be modeled. Different NN has complex special structures to solve more complex problems such as convolutional neural network (CNN), recurrent neural networks (RNN) and even more specialized structures such as Long Short Term Memory (LSTM).

Figure 2.9: The modular Neural Network [1]

Optimization algorithms are generally used to estimate the weights in the so called training of the NN. These optimization algorithms are named stochastic gradient descent, adagrad, adadelta, adam, adamax.

Activation functions apply non-linear transformations to coming inputs. Example of activation functions are sigmoid, hyperbolic tangent(tanh), and Rectified Linear Unit (ReLU) functions. Loss function measure error between ground-truth value and predicted value. The examples of loss functions can be Mean-Squared Error (MSE),

cross entropy, binary cross entropy.

2.4.2.1 Long Short-Term Memory

LSTM is a kind of special RNN, which provides a solution to the vanishing gradients problems of RNN. LSTM has memory cell component for storing important historic information using three gate structures.

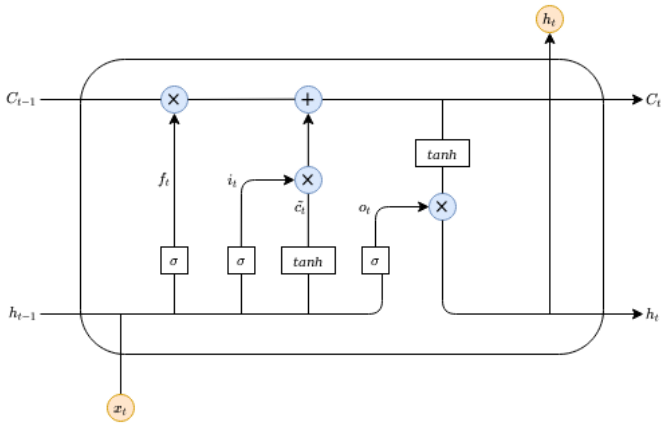

Figure 2.10: Block of LSTM

These gates determine which information needs to be added or cleared. Cell State C_t can be thought of as the memory of a network. It ensures that previous information is kept. The information to be transported is determined by the gates as shown Fig. 2.10.

In Eq. 2.32, f_t is that information from the previous cell h_t and the current information X_t are inserted into the sigmoid activation function. The forget gate f_t decides how much memory kept from previous memory state C_{t-1}. Information with 0 is forgotten and information with 1 continues to be carried by Cell State.

Another gate is the input gate i_t in Eq. 2.33, providing what information to write into current memory state C_t. i_t does Cell State C_t update. It is decided whether to update the previous and current information according to the result of the sigmoid operation. Information with 0 is considered unimportant and information with 1 is considered important. In addition, the tanh activation function, which compresses the data between -1 and 1, is used to regulate the network. Then the sigmoid and tanh

function outputs are multiplied and it is decided which information will be updated.

In Eq. 2.34, The exit gate determines the input of the next cell h_{t+1}. It is also used for guesswork. First pass the previous information and the information of the current input through the sigmoid function. Then pass the existing information on the Cell State through the tanh function. Finally, it decides what information will be the input for the next cell h_{t+1} by multiplying the two results. When the gate operations for the current cell are completed, the Cell State that will go to the next cell and the Hidden State (ht) information defined as the input information of the cell are decided. Depending on the current cell state C_t, the output of LSTM $h_t$is determined by the output gate o_t. [19]

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2.32}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2.33}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{2.34}$$

$$\tilde{c_t} = \tanh(W_{\tilde{c}}.[h_{t-1}, x_t] + b_{\tilde{c}}) \tag{2.35}$$

$$C_t = (C_{t-1} * f_t + i_t * \tilde{c_t}) \tag{2.36}$$

$$h_t = o_t * \tanh(C_t) \tag{2.37}$$

There are some studies in the literature that estimate VTEC using a neural-network-based model. One study is used LSTM based model using TEC values and related parameters to estimate local TEC values [12]. The study compares multilayer perceptron network (MLP) and the results finds the LSTM-based model more reliable to other model. LSTM-based method to estimate Global Ionospheric TEC based on SH coefficients are proposed [9].

2.4.3 Dimension Reduction with Principal Component Analysis

Principal component analysis (PCA) is a method that can be used in multivariate data sets. The main goal here is to create new uncorrelated orthogonal variables while preserving variation as much as possible. These are called principal components [34]. It can provide dimension reduction if strong correlations among the variates. The

first principal component generally provides the highest variance in the dataset with decreasing order with other PCs. PCA can be calculated as :

1. The average of each variable is subtracted such that the data has zero mean. Most of the time, the data is also normalized to have a unit variance.

$$X_{std} = X - \mu \tag{2.38}$$

   where X_{std} represents the zero mean data.

2. Compute Covariance Matrix of the data X_{std}

3. Compute Eigen Values λ and Eigen Vectors V of the data Covariance Matrix. Each Eigen Vector corresponds to a principal component. Eigen values are scalar and dimension of the vector is [number of components(PCs),number of features]. Eigen Values can give variance of PCs and the dimensions are [number of features] .

$$\frac{1}{N-1} X_{std}^T X_{std} V = \lambda V \tag{2.39}$$

4. Eigen Vectors are used to obtain Principal Components (PCs) which may also be seen as a transformation. In order to obtain PCs X_{pca}:

$$X_{pca} = X_{std} \dot{.} V^T \tag{2.40}$$

5. Inverse transform to obtain the original data X_{org} can be computed:

$$X_{org} = X_{pca} \dot{.} V + \mu \tag{2.41}$$

PCA is actually a method that can be used to reduce input dimensions. The above-mentioned Sections 2.4.1 and 2.4.2 can both be applied to the forecasting of PCs. It can be used as a dimension reduction in ionosphere estimation of the VTEC coefficients. In a study, the historical VTEC values and features obtained by PCA are used to predict VTEC values based on NN [14].

CHAPTER 3

METHODOLOGY

This chapter explains the overall methodology applied to achieve the objectives of the study. First of all, downloading and forming time series of Trigonometric B-spline coefficients using historic IGS IONEX maps is described. Afterwards, different methods to forecast both individual coefficients or Principal Components of the highest variance are listed. The application of different methods to the generated B-spline coefficients are given in detail. The proposed methods in this study are Single-Pixel , PCA-based forecasting and Block-based forecasting of Trigonometric B-spline coefficients. Finally, methods for the validation of he results are listed.

3.1 Preparing B-Spline Coefficient Time Series

The basis functions used in this study are tensor products of Trigonometric B-splines in the solar geomagnetic longitude and natural B-spline functions in solar geomagnetic latitude direction as given in Eq. 2.12. The level parameters J_1 and J_2 are selected as 4 and 3, respectively since they provide a good trade off between goodness of fit in reconstruction and number of parameters according to [5]. The procedure is straightforward as given in Fig. 3.1. The daily final IGS IONEX maps are downloaded from `ftp://cddis.gsfc.nasa.gov/gps/products/ionex/YEAR/DOY` between years 2000 and 2020. For each hour in each day, global grid VTEC with associated geographic latitude and longitude are obtained with the help of GPSTK library and Iontrace software [35]. The geographic latitude and longitude are converted to solar geomagnetic latitude and longitude using the Spacepy Python library [36]. Then, the Trigonometric B-spline coefficients for each hour are estimated using Ordinary Least

Squares adjustment using grid VTEC as measurements in solar geomagnetic coordinates. As can be seen from the Fig. 3.1, the resulting VTEC coefficients provide a general overview of the spatial variation of the ionosphere with localized basis function. Thus, resulting coefficient array can be seen as an image representation of the hourly snapshot coefficients, where each pixel corresponds to a Trigonometric B-spline coefficient.

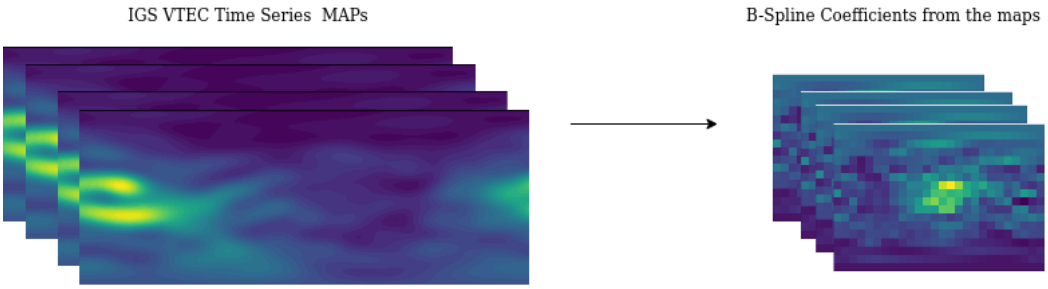

Figure 3.1: From IGS Maps To B-Spline Coefficients

3.2 Forecasting of B-spline Coefficients

This section provides proposed models for ionosphere prediction using the estimated coefficients. Firstly, a single pixel (single coefficient) based model is proposed for ionosphere prediction. For single pixel forecasting, both the LSTM-based deep learning model and the statistical SARIMA-based model are developed and tested. Two additional methods have also been proposed to account for the high temporal variance and also the spatial correlation of pixels with each other. One of them utilizes PCA-based models where the principal components are forecast using again LSTM-based and SARIMA-based models. The other method is to use a block of pixels (the neighbouring pixels of the target pixel) for forecasting. Since multivariate models are not discussed in the study, forecasting based on bock-pixel approach are produced only for the LSTM. The overall workflow of the model building, selection and validation activities are provided in Fig. 3.2.

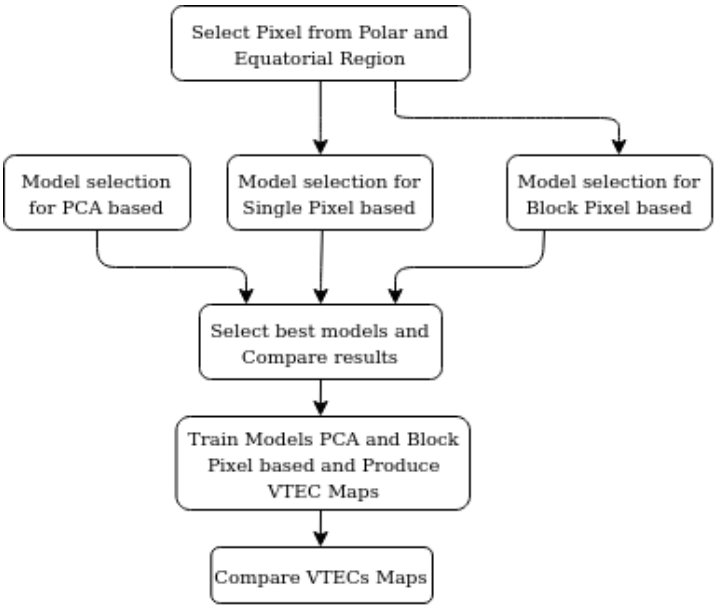

Figure 3.2: General workflow of the study

As can be observed from Fig. 3.3, due to the solar geomagnetic coordinate system selection, center pixels of the resulting coefficients are generally higher due to high ionization depending on mean Sun position. Thus, coefficients for the geomagnetic pole regions generally have lower values than the equatorial region. In order to save computing resources, two pixels are selected for model building and selection stage. These pixels are selected from the region with high electron content and the region with low electron content. These pixels are used for model selection for the two of the three specified methods namely Single-Pixel and Block-Pixel base. For the PCA-based method, PCs are generated by applying eigen decomposition of the temporal covariance of pixels for a period of days. Only the first PC due to high variance explanation ratio is used for model selection. The best models are selected based on hyper parameter search and validation error metrics. A general approach to data assimilation with Ionosphere is to assume that the ionospheric conditions do not change with short periods of time (generally one hour for GIMs). Thus, estimation procedures either apply a no-change constraint to the coefficients or assume a random-walk with the assumption that the coefficients of the next hour is equal to the current state. This assumption will be called Persistence Model in the rest of the manuscript. Each method is compared with the persistence model to search for an improvement of one-hour forecasting. The models are trained by using one year of data from 2013. The best models of each method are

then used to produce final models in another training process where longer time-span (2013-2015) is used to estimate model coefficients. According to the final results, the coefficients are compared with persistence model both in terms of forecast coefficients and reconstructed global VTEC maps using quiet and active storm conditions.

3.2.1 Pixel Based Forecasting

Since each B-spline coefficient is in solar geomagnetic coordinates, it may be considered as an individual time series. In this sense, the VTEC representation has 468 (18 for latitude and 26 for longitude) coefficients where a model for each coefficients shall be obtained. By modeling each coefficient individually, we can obtain the forecast B-Spline coefficients, which then can be used to reconstruct forecast maps. Fig. 3.3 shows that an example single pixel taken from each hour in the the data set. Fig. 3.4 shows the time series of a single B-Spline coefficient which corresponds to VTEC[9,13]. Matplotlib library is used to generate Fig. 3.4 as well as other figures in the scope of thesis [37]. As shown in Fig. 3.5, there can be a daily and also some seasonal repeating patterns in the time series of B-Spline coefficients. Over a 3-year period, a pattern can be seen in the coefficient values on the same days within a 24-hour period.

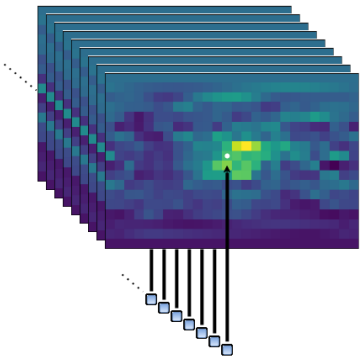

Figure 3.3: Same pixel for a time series of VTEC coefficients

3.2.1.1 LSTM Based

Here the LSTM based model is proposed for forecasting of a single coefficient. Firstly, data preparation is carried out for each of the B-spline coefficients. The same procedures are applied to train each model for 468 coefficients in total. The inputs of the model are determined according to the number of previous time steps used for pre-

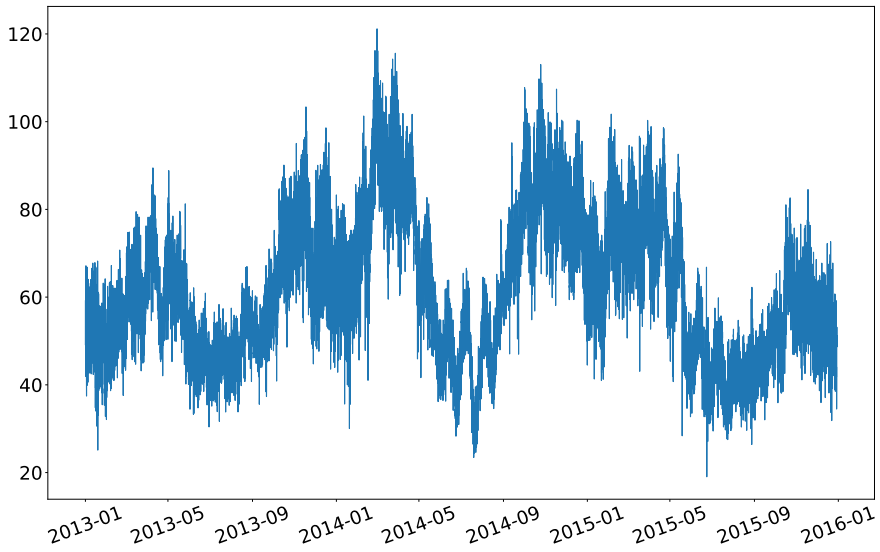

Figure 3.4: A time series of B-Spline coefficient VTEC[9,13] between 2013 and 2015

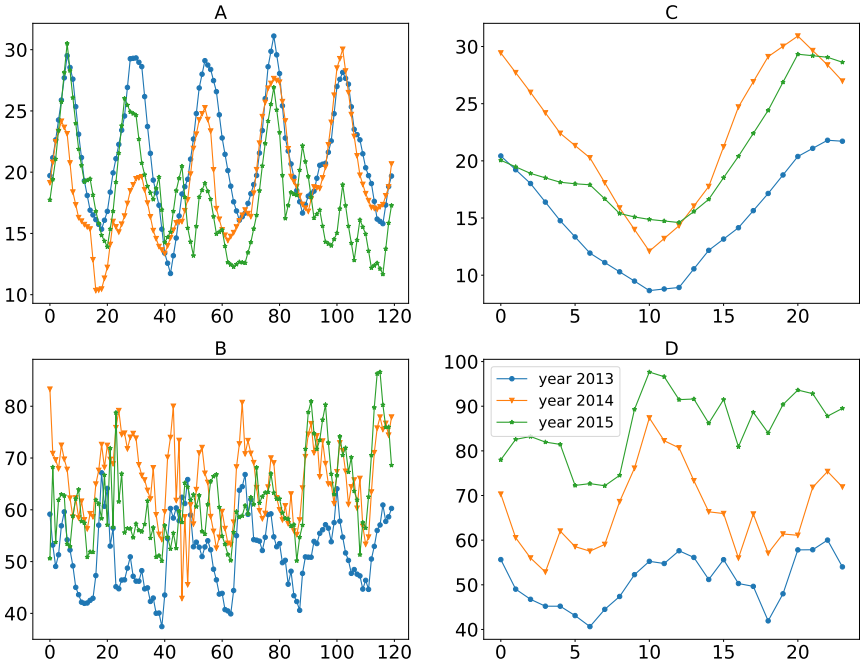

Figure 3.5: Daily Repeatable Pattern of B-Spline Coefficients. A) is First Five Day of B-Spline Coefficient [0,13], B) is First Five Day of B-Spline Coefficient [9,13], C) is One Day of B-Spline Coefficient [0,13], D) is One Day of B-Spline Coefficient [9,13]

diction, the number of forecast future steps (which is one for single hour forecasting). In this study, Tensorflow Keras [38] is used for deep learning models. LSTM layers receive three-dimensional data. They are as follows: [N, time steps, feature]. N is sample size or batch size. Feature is the number of input features which is one in this section. That is, only the previous time steps of the coefficients are used as input. For example, let the data be a time series from time step 1 to 15, previous time steps 6 and future time steps as 1, the input data set that will enter the model is created as follows:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 5 & 6 & 7 & 8 & 9 \\ 5 & 6 & 7 & 8 & 9 & 10 \\ 6 & 7 & 8 & 9 & 10 & 11 \\ 7 & 8 & 9 & 10 & 11 & 12 \\ 8 & 9 & 10 & 11 & 12 & 13 \\ 9 & 10 & 11 & 12 & 13 & 14 \end{bmatrix}_{[N,6]} \tag{3.1}$$

The two-dimensional [N, 6] data set is reshaped into a 3-dimensional $[N, 6, 1]$ matrix. The output data set that will enter the training phase is created as follows:

$$\boldsymbol{y} = \begin{bmatrix} 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \end{bmatrix}^T \tag{3.2}$$

After the training data set $(\boldsymbol{y}, \boldsymbol{X})$ is prepared, a layered architecture consisting of one or more LSTM layers followed by a Dense layer is created. Here, the number of neurons of the last Dense layer, which can be called the fully connected layer, is determined with the number of future steps. In the Fig. 3.6, an example network architecture consisting of two LSTM layers with additional hidden layers is presented. Due to computational power available, the models are produced by varying the time step and future step values only. The number of previous time steps are selected as six and twelve, while future steps are chosen as one for short term forecasting and twenty-four for the next day prediction performance. Activation function of Dense layer is selected

as linear. The loss function is chosen as the mean squared error (MSE) and optimizer is selected "adam" for all model training phases. After estimating a forecast model for all coefficients, results from each model are combined to reconstruct the global VTEC coefficients which can be used to generate forecast VTEC maps.

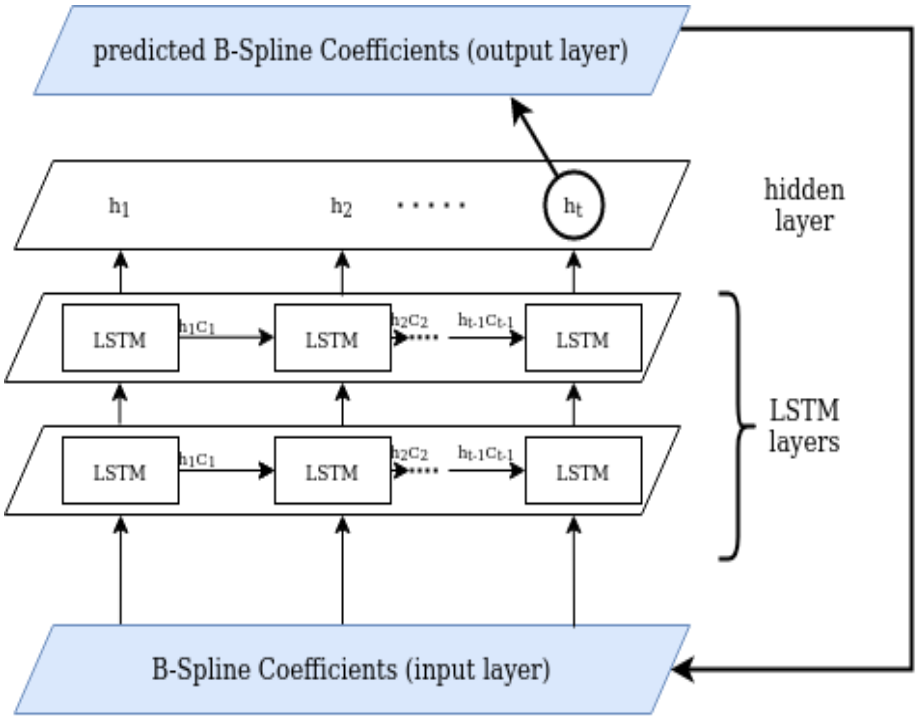

Figure 3.6: Network Architecture

3.2.1.2 SARIMA Based

The overall flow of steps for choosing a statistical prediction model can be listed as in in the Fig. 3.7. After the data has been analyzed, a transformation can be applied if necessary. For example, a logarithmic function can be used to scale the time series data. Then data is controlled whether it is stationary or non-stationary. If the data is determined to be non-stationary, ACF PACF can be looked at by taking the difference of the first order. If the data has seasonal components, the ACF and PACF can be checked again by taking the difference by the period. Parameters for the model can be selected with the help of figures. However, for the most accurate result can be obtained by an hyper parameter search utilizing the AIC values. The lowest value AIC determines the appropriate model for forecasting. The data is then fitted with

these parameters.

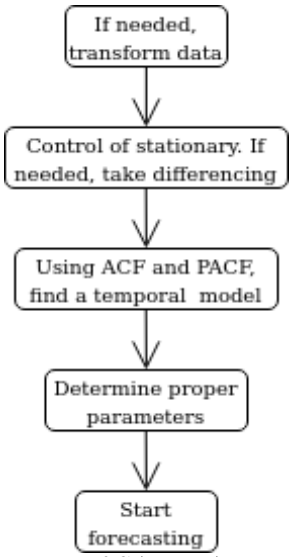

Figure 3.7: Flow of SARIMA model building

Due to seasonal patterns in the time series, especially with daily periods, SARIMA model with seasonality 24 has been selected in the study. When we interpret the auto-correlation plot of B-spline coefficients given in Fig. 3.8, a clear daily seasonality with a strong indicator of trend visible with the slow decrease in the autocorrelation function.

Some statistical information of the training dataset for the coefficient VTEC[9,13] between the years 2013-2015 is as follows. ADF Statistic: -3.124145 with associated p-value: 0.024817 and Critical Values: 1%:-3.43 and 5% : -2.862. We can see that statistics -3.124145 is larger then the threshold -3.43 at 1%. Thus, the null hypothesis may not be rejected clearly. Thus it may be a good idea to treat the dataset as non-stationary. But we can see that the two values are very close to each other. Indeed, the ADF value remains below critical value for 5%. Therefore, both stationary and non-stationary conditions should be considered for the time series.

According to autocorrelation function plot for the raw data provided in Fig. 3.8, auto-correlation is high for every 24 hours and there is a slow decrease as lag increases. This situation can be considered as an indicator of non-stationarity. If the first differences $\bigtriangledown X_t = X_t - X_{t-1}$ and a seasonal difference $\bigtriangledown_{24} X_t = x_t - X_{t-24}$ are taken as combined difference $\bigtriangledown \bigtriangledown_{24} X_t$, ACF plots can be obtained as in the Figs. 3.9, 3.10 , respectively.

After taking these regular and seasonal differences, the data can be considered as stationary. These indicators will be used when building SARIMA models not only for single pixel but also for PCA based models.

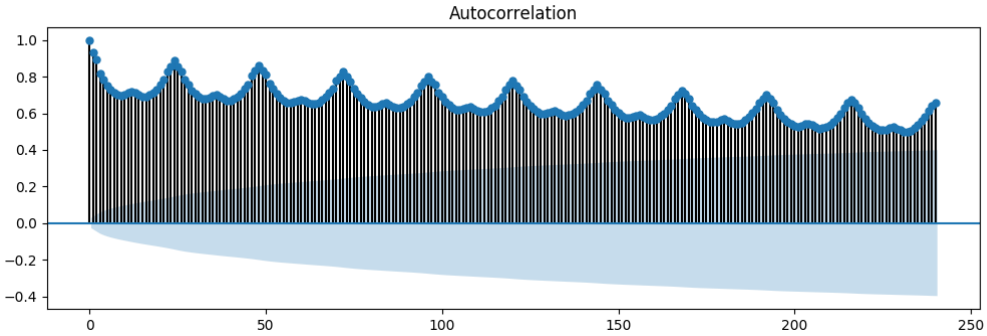

Figure 3.8: ACF of Raw dataset 240 hours in 2013

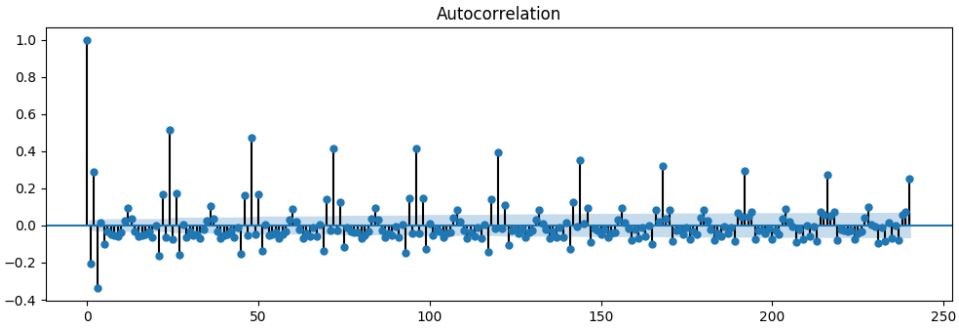

Figure 3.9: ACF of the first differences

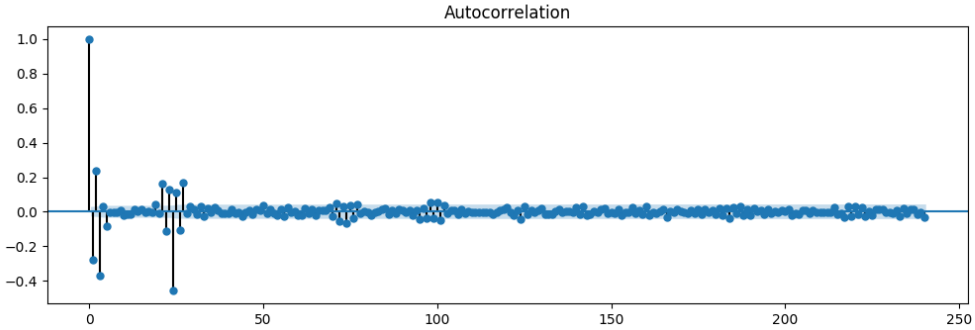

Figure 3.10: ACF of the first differences and daily seasonality

SARIMA contains two parts which are non-seasonal part and a seasonal part. AIC is utilized to obtain the best parameters in an hyper parameter search procedure. After model selection and training, forecasts of each B-spline coefficient can be obtained and forecast VTEC maps may be reconstructed.

3.2.2 PCA Based Forecasting

In the proposed PCA-based forecasting method dimension reduction is obtained by considering temporal and spatial variations of 468 pixels. The time series data set is collected in a large data matrix with columns representing B-spline coefficients and rows representing time evolution. Although generally, the data covariance matrix is calculated with the removal of temporal trend in each column, here we first calculate spatial mean of each hour as an additional time series. The spatial means will also be considered as an additional time series to be forecast alongside with the principal components. The spatial means are removed from all B-spline coefficients with corresponding hour. Then temporal means of the residuals which is the average of each pixel time series is subtracted from the data matrix value. Second, data covariance matrix is calculated. The dimension of the covariance matrix is [468x468]. Third, eigen values and eigen vectors are computed from the covariance matrix. Variance explanation ratio of each eigen value is calculated and eigen values are sorted in decreasing order of variance explanation. This way, it is determined how many Principal Components are required to represent the training dataset. Since the first 135 PCs can account for 99% of the data variance, it was decided to use only the first 135 PCs. In this manner, the number of time series has been reduced from 468 to 135 + 1 with the new variables, namely the PCs, uncorrelated with each other. Both LSTM-Based and Sarima-Based models are trained for each variable and also for spatial means . Finally, the forecast B-spline coefficients can be reconstructed according to the reconstruction formula given in Chapter 2, with the addition of forecast spatial means.

3.2.3 Block-Based Pixel Forecasting

This method is proposed due to handle the spatial variation of the B-spline coefficients. The pixels are related with each other locally because of spatial autocorrelation as can be seen from the coefficients maps provided before. Thus, neighboring pixels may contain valuable information for the forecasting of target pixels. The general flow is shown in Fig. 3.11. First of all, we decompose the pixels as in the individual pixel estimation method. However unlike the single pixel methods, neighboring pixels are taken according to the windowing as seen in Fig. 3.12. Neighboring pixels are used as

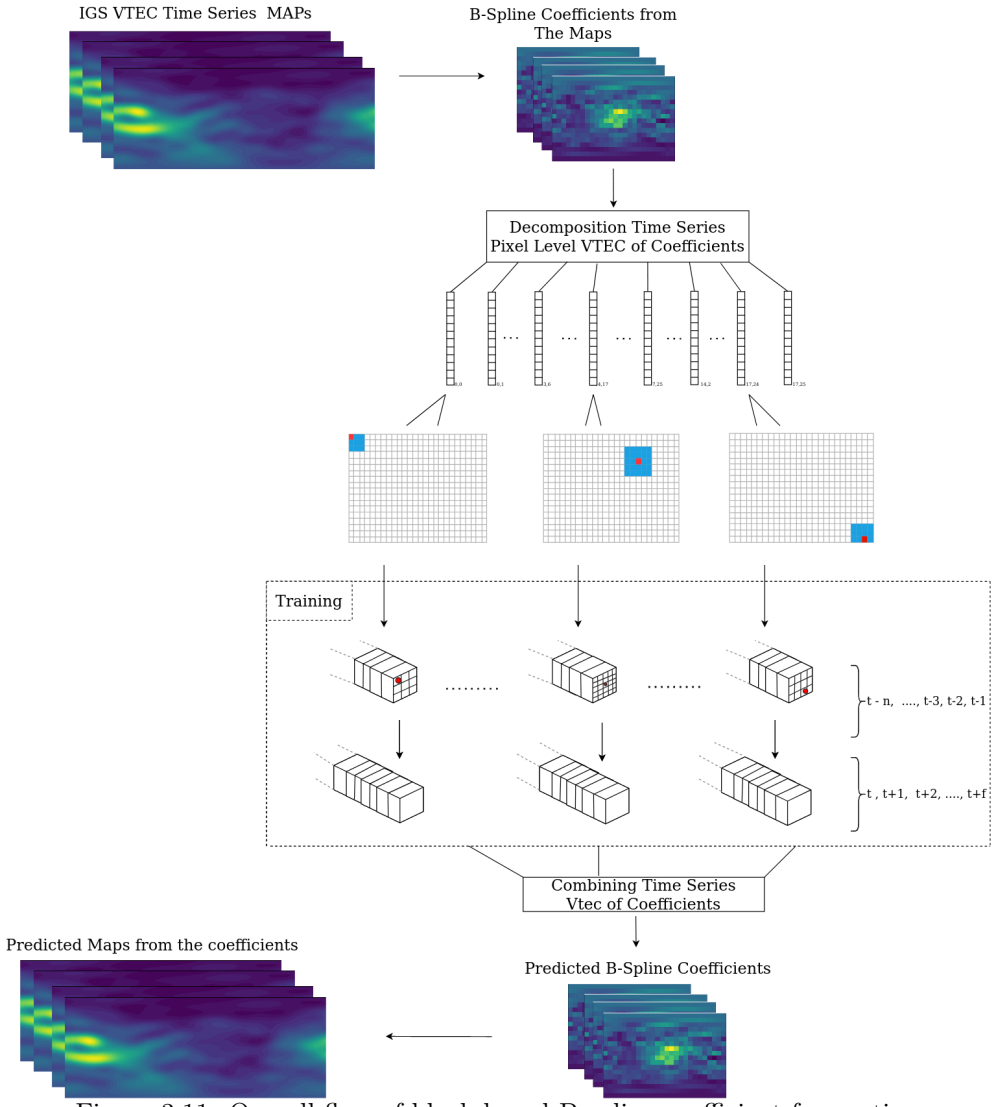

Figure 3.11: Overall flow of block-based B-spline coefficient forecasting

additional features for each coefficient. As shown in the lower part of the Fig. 3.12, two pixels are used from left to right, bottom to top and from the corners, shown in blue. The coefficients shown in red are the target coefficients to be forecast in Fig. 3.12. For instance, there are nine features for the forecasting of a single coefficients without edge effects.

Fig. 3.13 represents training block for each coefficients. The number of features is different in each model according to the location of the window and accounting for edge effects. Since the LSTM layer receives input in three dimensions, the display is [N, time steps, features]. Requested output is a set as [N, future steps]. The Number

of neurons in the Dense layer in the model is determined according to the future steps. For example, in the prediction model of the pixel in the upper left corner, the input becomes [N, time steps, 9].

Due to computer resources available, selection of hyper parameters is mostly based on literature survey from [32; 9]. The optimizer function has been used as Adam in these studies. Activation function of the Dense layer is linear, the Loss function is chosen mean squared error (MSE). After model training for each pixel, the forecast results from each model are combined to reconstruct VTEC maps.

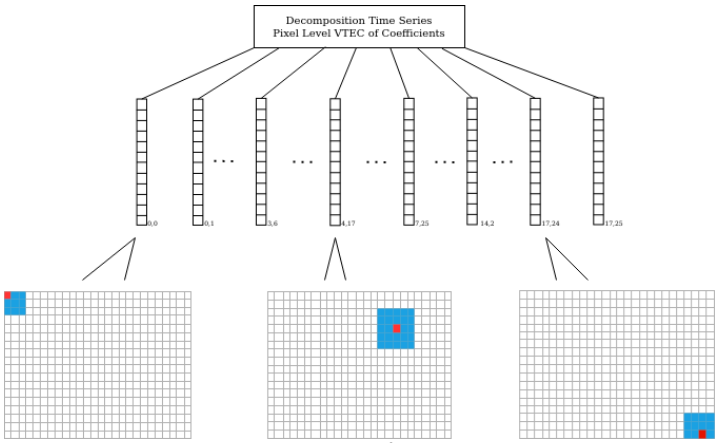

Figure 3.12: Decomposition of time series in pixel level

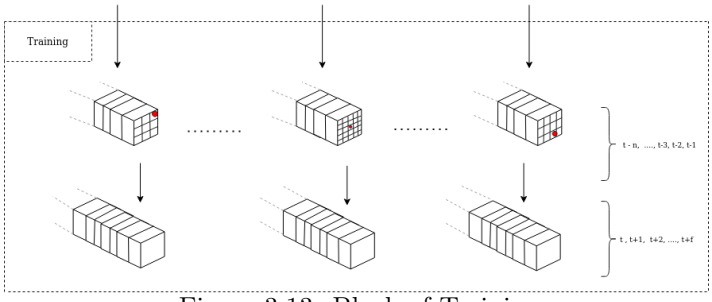

Figure 3.13: Block of Training

3.3 Validation of The Results

Within the model selection phase of each method the models are compared with each other to select the best model. For SARIMA based models, AIC criterion with hyper-parameter search is applied, for LSTM based methods 20% of the training set is used as validation data set and MSE of the validation errors is used. In addition, the best

results are compared with the results obtained from the persistence model with figures of percentage improvement. In Single-Pixel method persistent model is compared with LSTM and SARIMA results. The PCA-based model is compared within itself with respect to the LSTM and SARIMA models for PCs. In the Block-based model, the LSTM results are compared with the single pixel LSTM model results.

For the comparison of the model results, days with both high and low geomagnetic activity is chosen as the validation data. The days are selected in such a way that they were not used in any phase of the training stage. The forecast B-spline coefficients and also the resulting VTEC maps are compared with persistence model generated from ground truth time series data and also IGS global VTEC maps.

The Mean Squared forecast Eror (MSE) is used as a metric for all validation steps. Results for high and low geomagnetic activity are reported separately for each hour and day. The days selected for validation are from the year 2018. Three consecutive days with low storm conditions are selected from January, while three days with severe geomagnetic storm is selected from August utilizing the Dst index.

Mean squared error, which is generally used in VTEC forecasting studies [12], is the average squared differences between the forecast values and the ground truth data:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2 \tag{3.3}$$

where N is number of observations, y_j is actual observation, $\hat{y}_j$ is the forecast data. RMSE is the defined as the square root of MSE:

$$RMSE = \sqrt{MSE} \tag{3.4}$$

An improvement percentage of each forecasting method over persistence model results as calculated as follows:

$$\text{improvement percentage} = \frac{P_{MSE} - [X]_{MSE}}{P_{MSE}} x100\% \tag{3.5}$$

where $[X]_{MSE}$ is the MSE of forecasted VTEC coeficients where $[X]$ represents the method name such as Block-Based Forecasting or PCA-Based Forecasting.

CHAPTER 4

APPLICATION

In this chapter, application and results will be listed for the methodology described before. First, the preparation of the data set is explained. The results of model selection using one year of the data set are listed with associated success figures. The visual and numerical comparison of the best models for each method are listed using a longer training data and independent validation days. The results and lessons learned from the study is discussed in the end.

4.1 B-spline Coefficient Time series

The trigonometric B-spline coefficients are estimated with the help of the Iontrace library and saved as individual files for each hour. The data set contains 24 files for each day, where each file contains the estimated VTEC coefficients representing the IGS final VTEC map for that hour. This data set, which can be considered as a ground truth data, is both used for model training and evaluation of forecasting quality . Although the VTEC coefficients are estimated for almost 20 years, only a few years with high geomagnetic activity is used for training and testing due to the high demand of computational resouces.

4.1.1 Training Dataset

B-spline coefficients are estimated for every hour of the day resulting in a set of twenty-four hours per day. During the model building phase, the dataset is divided into training, testing and validation parts. The training and testing dataset has been

selected between 2013-2015. While the first 80% of the 3-year data is divided for training, the 20 percent in the last section is divided for testing. Dst index has an important role to determine longer training dataset. Temporal variation of Dst index is shown in Fig. 4.1. The days with Dst index below -50 nt are generally considered as active ionospheric condition. For example, we can see the 3rd and 4th months of 2015 having very high activity. This is the reason for longer training data set to contain 2015. Due to computing resources available, the model selection phase uses only one year of B-spline coefficient time series. As can be also seen from the figure, year 2013 also contains a couple of severe geomagnetic storms for which this year has been selected for model selection stage.

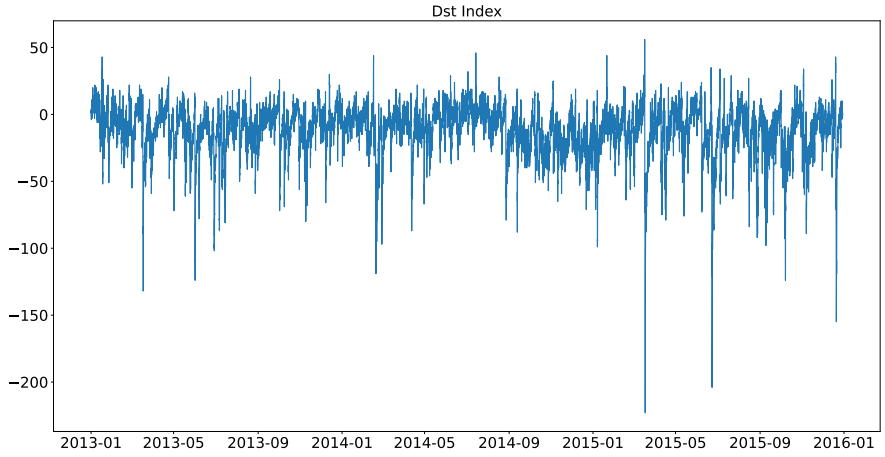

Figure 4.1: DST Index Between 2013 To 2015

VTEC values are less variable in the geomagnetic pole region. The part shown in the Fig. 4.2 belongs to a coefficient in the polar region. The coefficient presented in the Fig. 4.3 belongs to the Equatorial region. Here the temporal variation in ionosphere is more visible. When the two figures are compared, the equatorial region TECU value is spread over a wide range of pixels and shows significantly more variation.

4.1.2 Validation Dataset

For the comparison of selected models of each method proposed in this study, validation days are selected such that geomagnetic activity is high and low. Some days from 2018 are selected for validation, which include both severe storm conditions and quiet days.

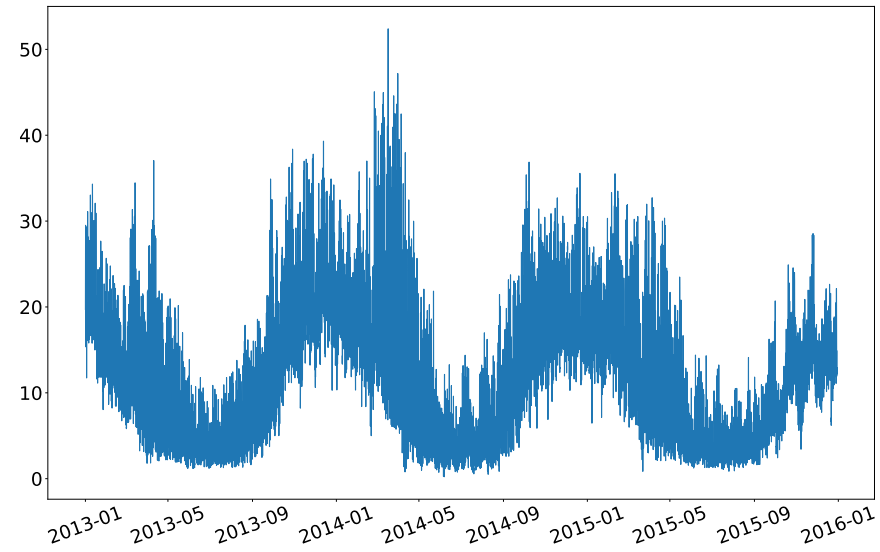

Figure 4.2: VTECs[0,13] B-Spline Coefficient Between 2013 To 2015

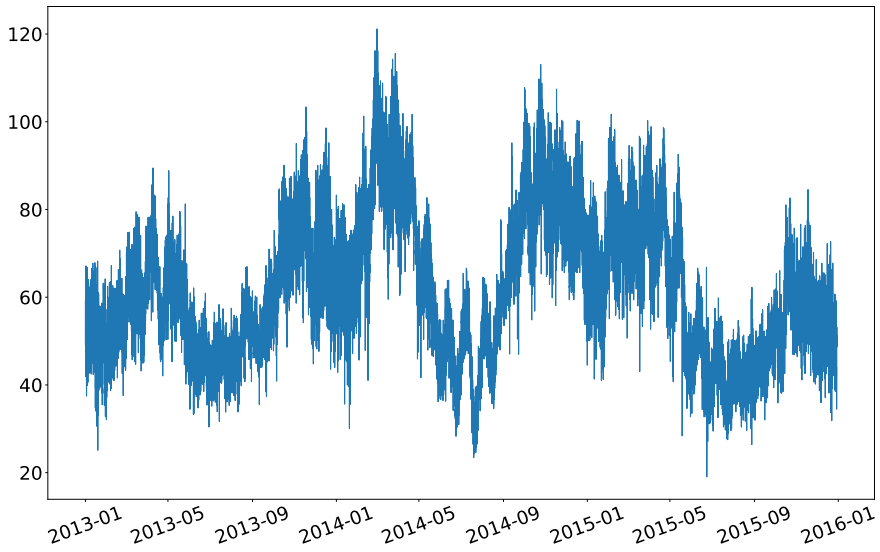

Figure 4.3: VTECs[9,13] B-Spline Coefficient Between 2013 To 2015

The days named as Quiet days are the 3 days of January 15, 16, 17; and those with severe storm days are August 25, 26, 27. Especially on August 26, the Dst index was below -100 nt indicating a severe storm event as shown Fig. 4.4.

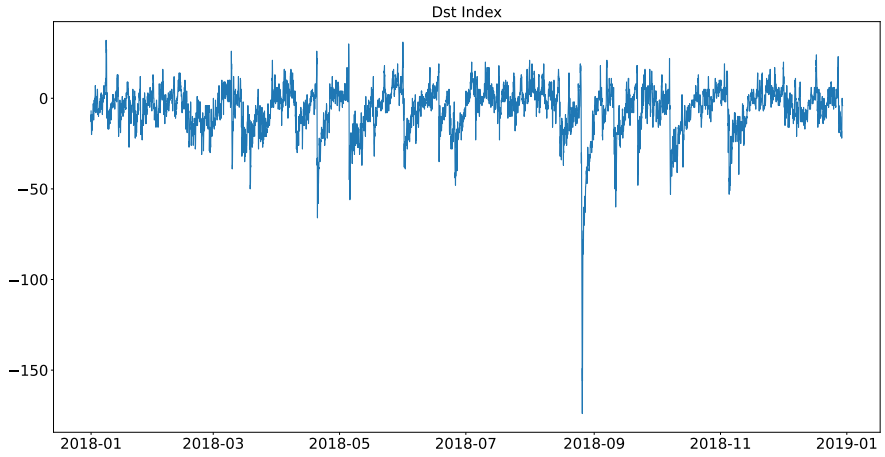

Figure 4.4: DST Index Between in 2018

4.2 Pixel Based Forecasting

A couple of B-spline coefficients are selected from the highly varying equatorial and polar regions with low variance. These coefficients are VTEC[0,13] shown in Fig. 4.2 for the polar region and VTEC[9,13] depicted in Fig. 4.3 for the equatorial region. TEC intensity is high in regions with low latitude which is caused by equatorial anomaly. In regions with high latitude which is polar region, the variance is lower. Thus, these two coefficients can be considered in the decision process of model selection due to low and high variance in their time series. Data from year 2013, which is part of training data, was used for this selection. The model was trained with 80% of this data and tested with 20%. Certain parameters were constructed to determine hyper parameters of each model. These are made specifically for LSTM and SARIMA model selection.

Another consideration in the model building stage is the assumption about the trend. Although some periodic trend is clearly visible from the time series, the physics of the Ionosphere makes the total variation to be finite within the Ionosphere. For further analysis of the linear and seasonal trend, the Python FB Prophet library from Facebook was used. While the models with a constant trend are also trained, the trend and annual seasonality estimated by the library is also tested. Models produced by both subtracting the Prophet trend model and constant trend assumption with the same hyper parameters are trained and their results are presented for comparison in Fig.

4.5 and 4.9.

Some parameters are excluded from parameter selection in LSTM models due to the limiting factor of computational resources. The adaptive moment estimation optimization algorithm (adam) is used as the optimizer. The activation function in the LSTM layers was determined as the rectified linear activation unit (ReLU), the batch size is fixed at 1024.

The Table A.1 lists the name of each model with associated parameters. The parameters changed for the model selection are number of layers, number of neurons to each LSTM layer and the number of previous time steps. Number of layers are selected as 1 and 2. The consideration of many more LSTM layers has been limited based on previous works [12; 32]. The number of neurons was chosen as multiples of two. The reason for starting from 32 is to test how this parameter affects the results without too much complexity in the model. In double-layer models, the number of neurons in the second layer is half that of the first layer. Models were obtained by looking back 6-hours and 12-hours of previous time steps. The selection of the numbers here is to be in the form of multiples of the 24-hour period and so that the complexity of the model do not increase. The changed parameters are applied both 1-hour ahead and 24 hours ahead. In this study, it is aimed to obtain good results in short-term forecasts, which is 1-hour. Because it can be used to eliminate the Ionosphere errors in near-real time positioning. It can be especially important in terms of data assimilation. At the same time, 24-hours forecasts were also tested to determine the 24-hours forecasts quality.

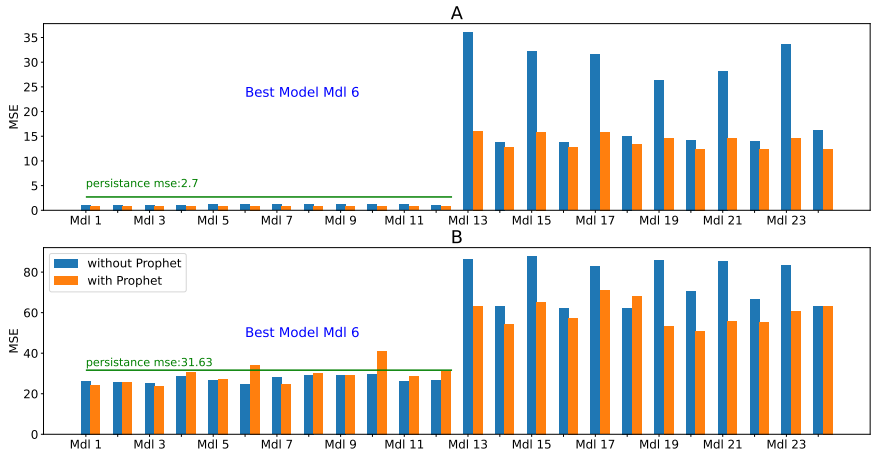

Figure 4.5: MSE Both A [0,13] and B [9,13] Coefficients for Single Pixel based LSTM Model Selection

An overall good model is selected in model selection phase. The models obtained using Prophet did not provide much improvement in the equatorial region compared to the polar region. This may be interpreted as the polar region patterns are much more regular, while the equatorial region has a lot of variance. At the same time, the results obtained without using Prophet are more successful, especially in 1-hour ahead forecasting. For this reason, the final model to be used is selected as the M6 as shown Table A.1 according to Fig. 4.5.

When compared to the persistence model as given Fig. 4.6, the pixel-based model is much better than the persistence model for 1-hour ahead.

Model selection for the SARIMA models was made by looking at the AIC values. ACF and PACF of both coefficients were examined in Fig. 4.7 and Fig. 4.8. According to the results in the figure, there is a strong correlation between lags. For this reason, a difference was taken and ACF and PACFs are calculated as shown in Figs. A.1 and A.2. In order to show whether the data is stationary or not, p-values are checked. "d" is kept as 0 in methods of SARIMA non-seasonal orders due to a borderline result and priority for computer power. As mentioned in the Methodology chapter, a daily periodic data set with seasonaliy equal to 24 is assumed. Therefore, for the seasonal part of SARIMA, "s" is set to 24 and "D" is fixed at 1.

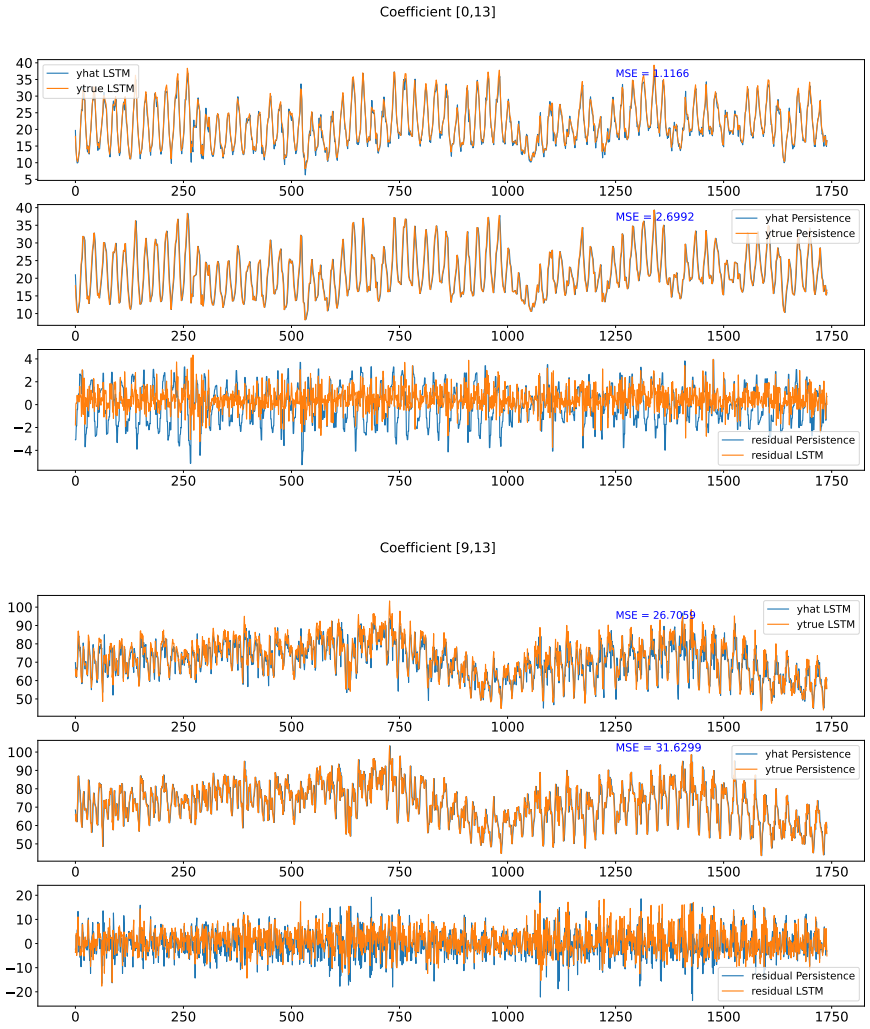

Figure 4.6: Persistance vs Mdl 6 LSTM Model for the Coeff. [0,13] and [9,13]

Figures A.3 and A.4 were used to determine the order ranges of "p", "q" and "P", "Q". PACF chart was effective in determining the p and P orders interval. For this reason, while variables (1,2,3) were selected for p, variables (0,1,2) were selected for P. The ACF chart was effective in determining the q and Q orders range. For this reason, while variables (1,2,3) were chosen for q, variables (0,1) were chosen for P. At the same time, models were produced using trend constant or Prophet as in LSTM models.

Using the parameters of the best AIC value model, the results were produced by giving a value of 1 for d, which was kept constant again. Here, if there was a significant change in the result, it would be taken as d=1 in the model to be used in the overall.

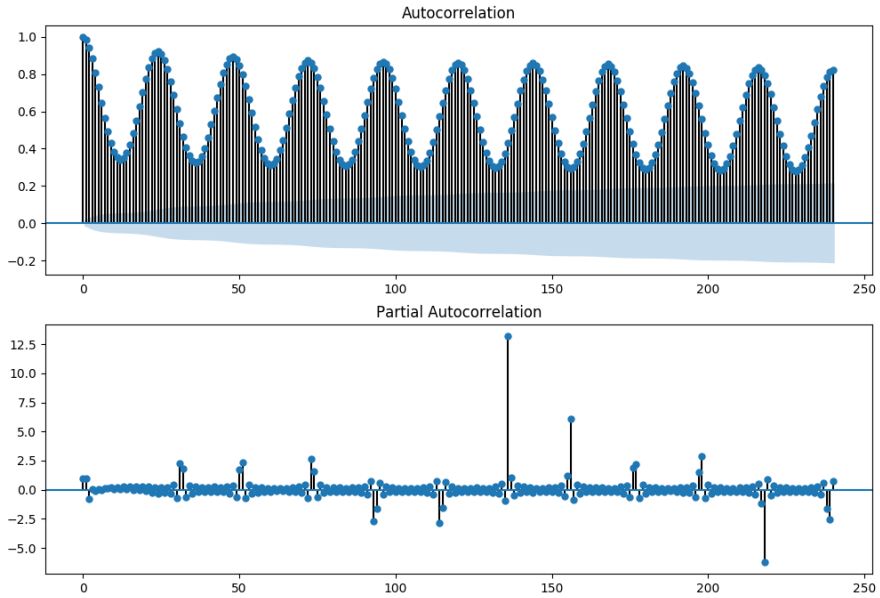

Figure 4.7: ACF and PACF of VTECs[0,13] B-Spline Coefficient

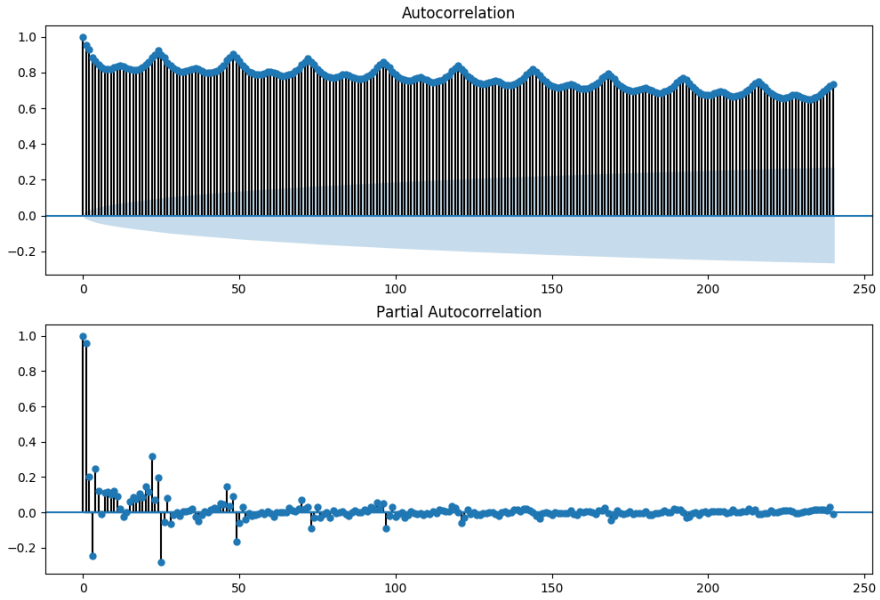

Figure 4.8: ACF and PACF of VTECs[9,13] B-Spline Coefficient

However, d is left as 0 since no significant change is provided. There was no obvious change in AIC values with or without Prophet, thus models were produced without

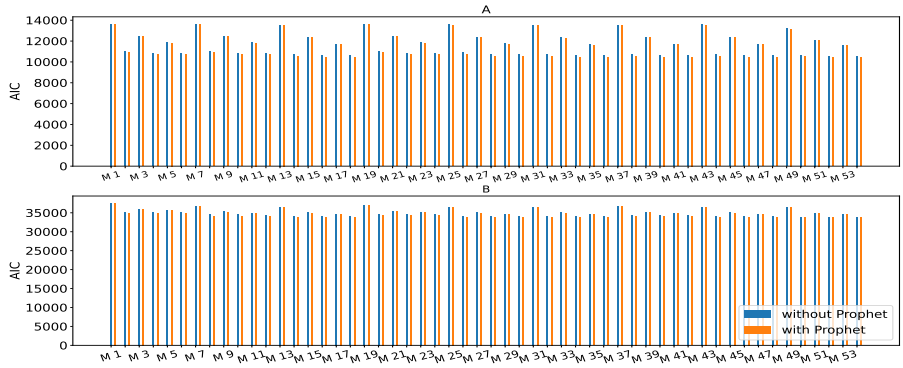

Figure 4.9: AIC Both A : [0,13] and B : [9,13] coefficients for Single Pixel based SARIMA Model Selection

Table 4.1: Comparing MSE for LSTM using Mdl 6 and SARIMA using M 54 for Single-Pixel

Pixel	Hour Ahead	Persistence	MSE of LSTM	MSE of SARIMA
[0,13]	1-hour	2.7	1.137	0.525
	24-hours	16.47	15.058	11.280
[9,13]	1-hour	31.63	24.641	14.057
	24-hours	47.718	62.389	31.863

using Prophet as shown in Fig. 4.9. According to the AIC results, the best result is consistent for both [0,13] and [9,13] coefficients. This model is called M54 from Table A.2 with order of p equal to 3, order of q equal to 3, order of P equal to 2 and lastly order of Q equal to 1, d=0, D=1 and s=24.

According to the Table 4.1, in accordance with our primary aim, models of one hour ahead forecasting are better than persistence. SARIMA results in single pixel are more successful than LSTM model results. The SARIMA model seems more successful in some parts which may be a result of searching for only limited number of parameters in the model selection for LSTM.

4.3 PCA Based Forecasting

Both LSTM and SARIMA are used in the PCA-based model building. For this reason, model selection is required for both LSTM and SARIMA. In this part, firstly, the temporal and spatial mean calculation of the data was made. The temporal mean is

the time series mean of each coefficient. The spatial mean as shown Fig. 4.10, on the other hand, explains the mean of all coefficients at each hour. PCA is applied by subtracting the means from the raw data. The first three PCs are shown in the Fig. 4.11. The number of PCs to be used is determined by explaining 99% variance of the data as shown in the Fig. 4.12. It was decided to use the first 135 PCs by cumulative summation of eigenvalues. Models will be produced for 135 PCs and also for spatial means. The model selection in this section was determined using only the first PC, as it contains the most information.

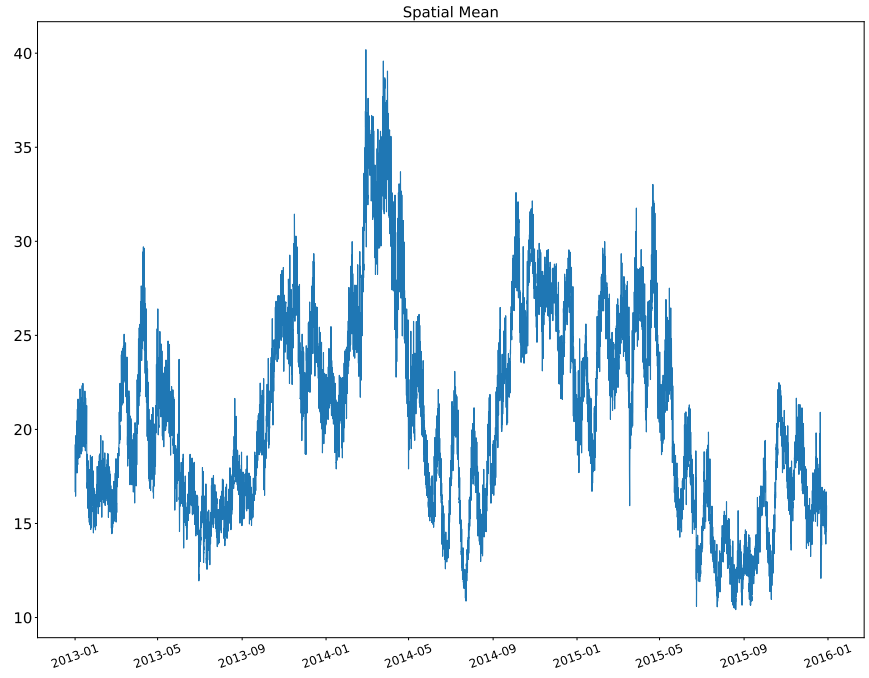

Figure 4.10: Spatial Means Per Hour Between 2013 - 2015

For LSTM model selection phase a similar hyper-parameter search was applied as in the single pixel case. Models with the same hyper-parameter are trained as the trend being constant and with Prophet based trend. The optimizer, loss function, activation functions of LSTM layers and batch size are all kept constant. These are adam, mse, ReLU and 1024 respectively. Changed parameters are number of layers, number of neurons, number of previous time steps. Again, 1-hour and 24-hours were used as feature steps. As seen in the graph in the Fig. 4.14, by looking at the predicted

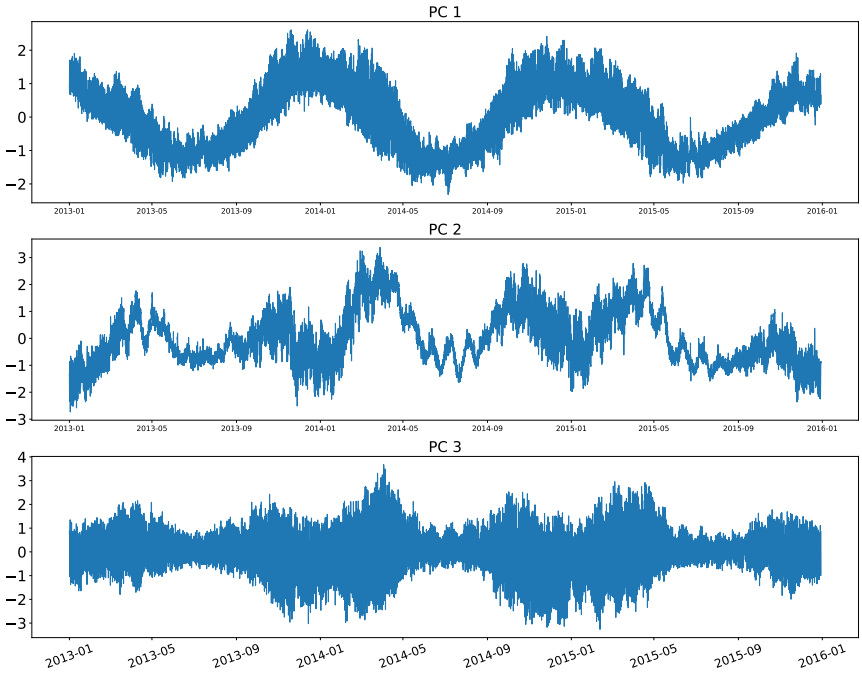

Figure 4.11: First Three Principal Components Between 2013 - 2015

model results of both 1 hour and 24 hours, moving the trend forward with the Prophet gave more accurate results. For this reason, in this part, a model was produced using prophet. The final result for hyper parameters was determined as M6 as shown Table A.1. Both the spatial mean and the first 135 PCs will be carried forward in the future with the hyper parameters of the specified model.

The model selection for the SARIMA model is the same as with Single Pixel-based model selection. The ACF and PACF produced for the first PC are shown in the Fig. 4.13. The range determined for the models is as follows; while set (1,2,3) is selected for p, set (0,1,2) were selected for P. And while (1,2,3) is chosen for q, (0,1) is chosen for P, d=0, D=1 and s=24. According to results, the lowest AIC value is obtained for model M40 from A.2 which corresponds to; p equal to 3, order of q equal to 1, order of P equal to 1 and lastly order of Q equal to 1. At the same time, removing the trend from the data using prophet has not made a significant difference. For this reason, the Prophet library was not used for the SARIMA model.

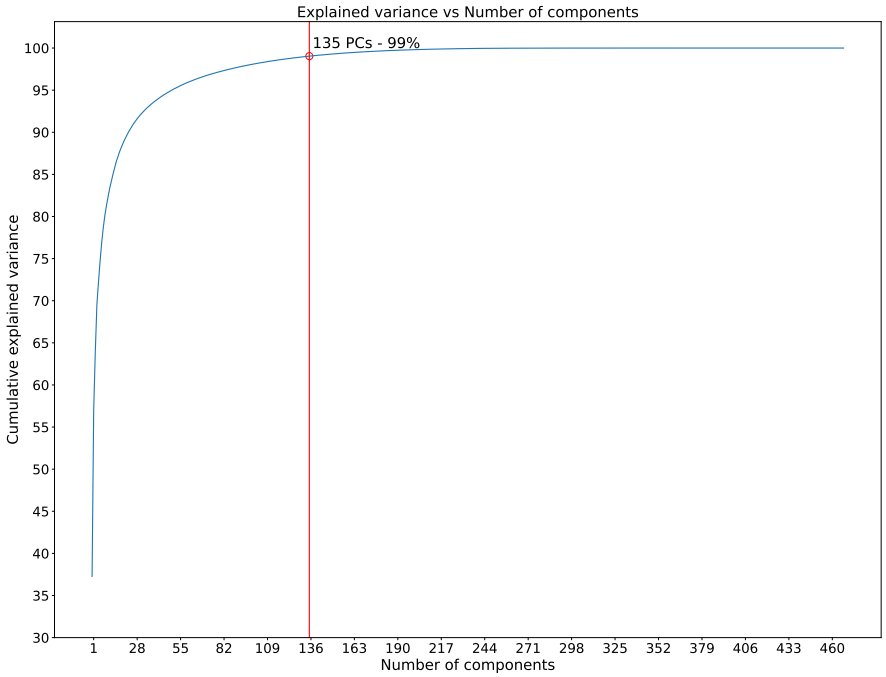

Figure 4.12: Cumulative Explained Variance

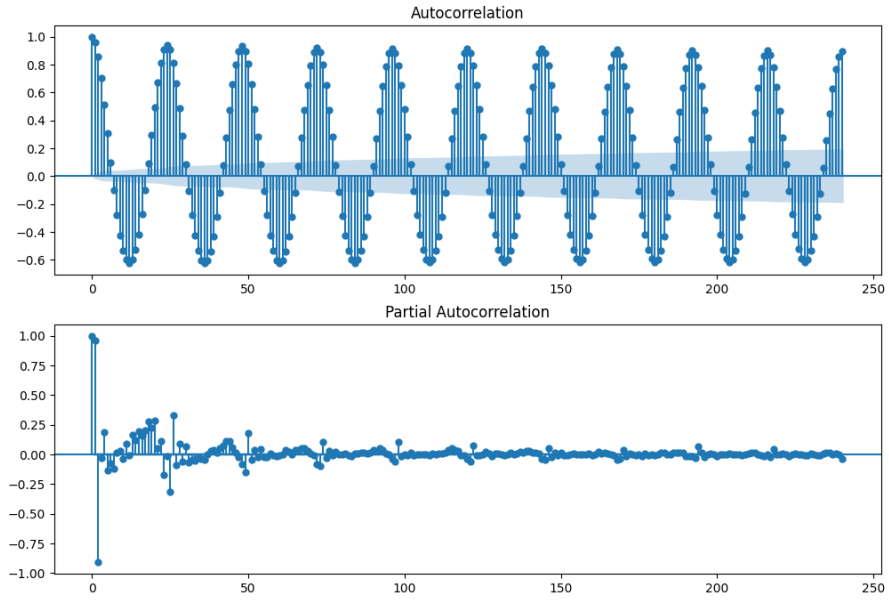

Figure 4.13: ACF and PACF of First PC

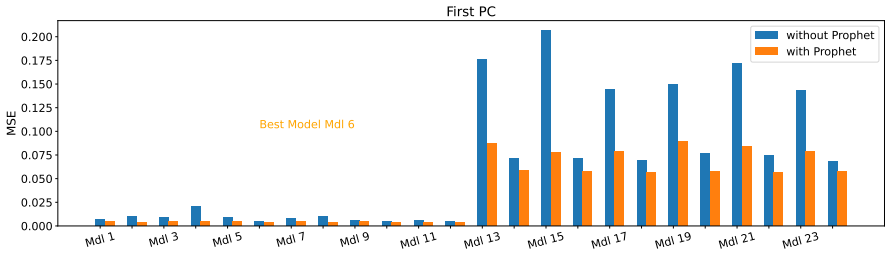

Figure 4.14: MSE First PC for PCA based LSTM Model Selection

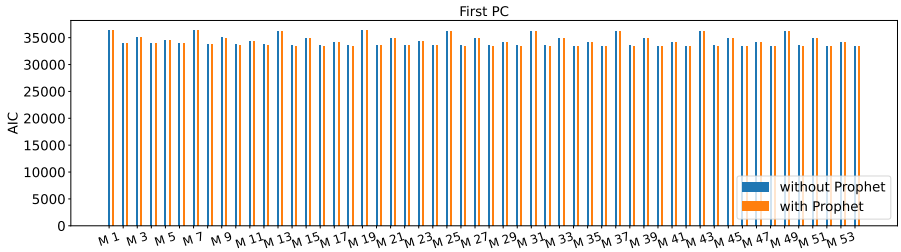

Figure 4.15: AIC First PC for PCA based SARIMA Model Selection

In PCA-based modeling, a comparison Table 4.2 is generated with the final models for PC1. Mdl6 model are used for LSTM and M 40 models are used for SARIMA. According to the results in the table, SARIMA is again more successful for both hours. However, there was no significant difference in MSE values in the 24-hours ahead forecast.

4.4 Block-Pixel Based Forecasting

In this part, model selection was made with the coefficient from the equatorial region [9,13] and the coefficient from the polar region [0,13]. In the selection of the Block-based model, the method in the Single Pixel-based method is applied. The determined hyper parameters were applied to the pixels in the high and low variance regions. Optimizer which is chosen adam, batch size being 1024, loss function which is mse and

Table 4.2: Comparing MSE for LSTM using Mdl 6 and SARIMA using M 40 for PCA based

	Hour Ahead	MSE of LSTM	MSE of SARIMA
PC 1	1-hour	0.00417	0.000934
	24-hours	0.0570	0.0434

activation funciton of LSTM layer being ReLU is constant. The modified parameters are the number of layers, time steps and neurons. Each model, 1-hour and 24-hours ahead, is modeled using trend constant and Prophet. The coefficient used in the Prophet here is directly itself. No trend information is extracted from neighboring pixels around it.

When the results in the Fig. 4.16 are examined, it can be seen that the results obtained using the Prophet are successful in the long term. However, in general, the model in which Prophet was not used was chosen due to both the short-term forecasting being the first priority and the overall success in both coefficients. The model with the lowest mse value appears to be the M6 in Table A.1. It can be seen that M6 has a single layer 128 neurons LSTM layer and time steps of 12.

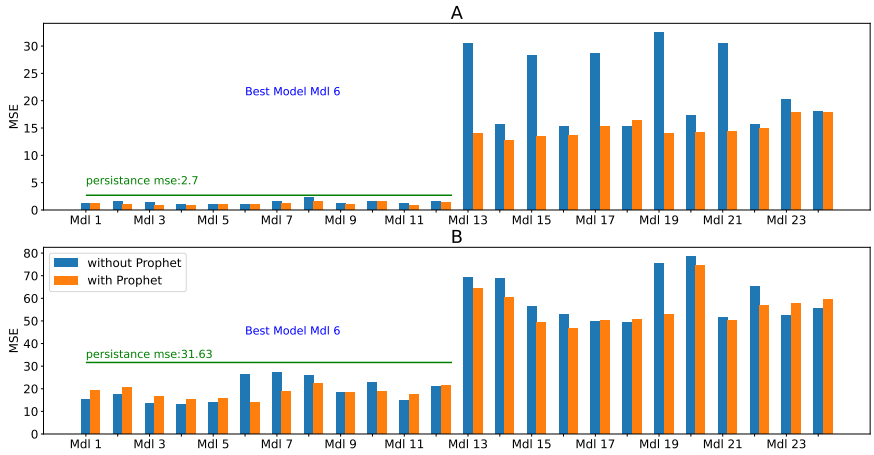

Figure 4.16: MSE Both A [0,13] and B [9,13] Coefficients for Pixel Block based LSTM Model Selection

Block-based model is suggested because it provides improvements over the single pixel based model. It gives more consistent results, especially in modeling of the equatorial region. The parameter sets used for model selection are applied here as well. The results produced using Mdl 6 for Single pixel based method and Block pixel based method are given in Table 4.6.

According to the results in the Table 4.6, MSE values for the coefficient [9,13] showed a significant improvement for Block based model compared to the Single-Pixel method. The fact that the variance in this coefficient is well modeled may indicate that there

Table 4.3: Comparing MSE for Single-Pixel and Pixel-Block based using Mdl 6

Pixel	Hour Ahead	Persistence	LSTM Single-Pixel	Block-Pixel
[0,13]	1-hour	2.7	1.137	1.087
	24-hours	16.47	15.058	15.311
[9,13]	1-hour	31.63	24.641	10.559
	24-hours	47.718	62.389	49.493

may be a significant improvement for other pixels as well.

4.5 Comparison of Predicted Ionosphere Maps

Besides the Single Pixel based method, Block-based based and PCA based models are proposed in this study. In this section, the forecast reconstructed B-spline coefficients estimated by these methods and the resulting VTEC maps will be compared. First of all, the model is trained for all pixels with the best model of the Block-Pixel based method. Before returning to the maps, examinations are made on the reconstructed coefficients. Validation is done on the selected quiet days and storm days. Coefficients are generated by estimating one hour ahead for four days. As seen in the Figures 4.17 and 4.18, the results of this method are compared with the persistence model. In the overall results, it is seen that the results produced using the LSTM-based Block-Pixel method show an improvement over persistence model. However, there is an observable variance in the results, especially in the severe storm hour, Fig. 4.18, on August 26 at 6 am. The variations here may be improved if the the Dst index is considered as an additional feature for forecasting because at the 6th hour of this date, the Dst index is -174 nt. For 6 o'clock of this day, the coefficients are analyzed based on the results from the best methods.

In the Fig. 4.19, the pixel-based comparison of the individual coefficients are shown. Here we examine how the results change in the previous hour and the next hour. As can be seen from the Fig. 4.19, MSE values are very high in pixels in the equatorial region for every 3 hours. The residual values of [14,7] pixels and [24,9] pixels at 3 consecutive hours are quite large. The mean and variance Table 4.4 has been added to look at the bias and variance of the residuals. Although the bias is not very high, the variance is relatively high.

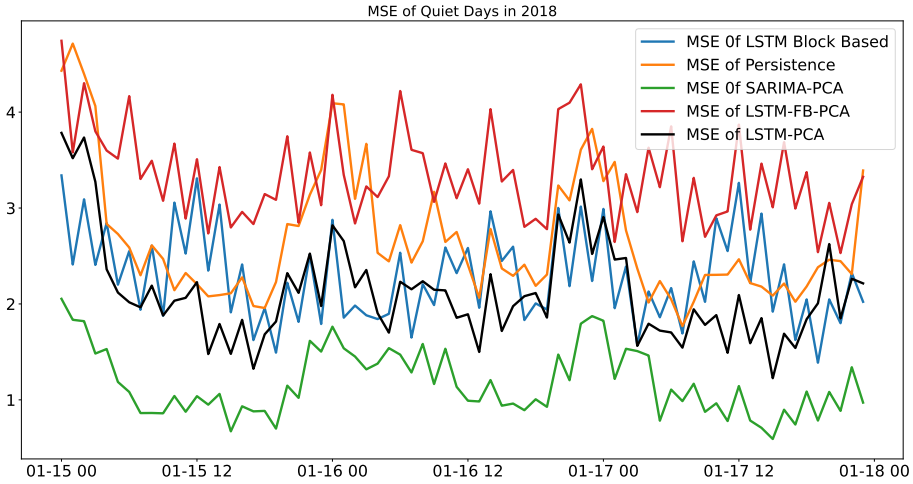

Figure 4.17: MSE of Quiet Days for all models

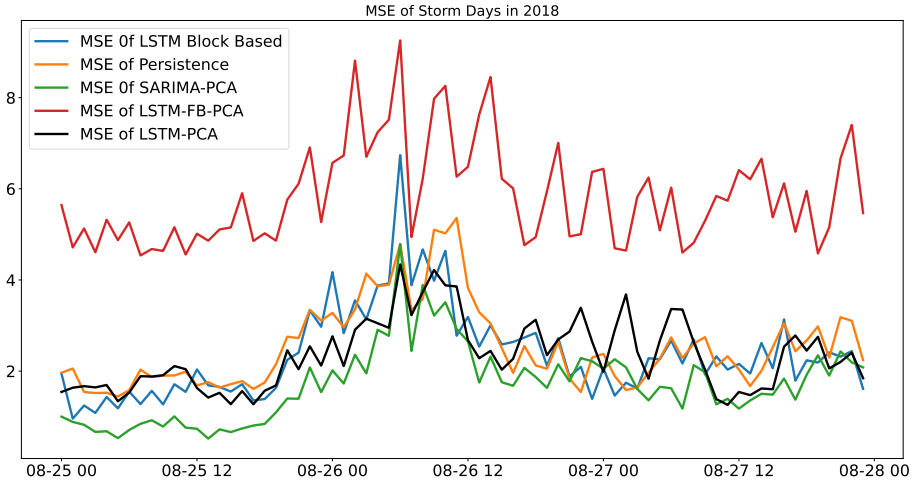

Figure 4.18: MSE of Storming Days for all models

Considering the results on PCA basis, the results obtained with SARIMA in Figures 4.17 and 4.18 which is green line, are more successful for many hours than Block-based model. However, the results are not successful in the LSTM models where the trend is predicted with the Prophet. The PCA LSTM, indicated by the black line in the figure, is trained without Prophet. According to the validation results, it provides more successful results in terms of MSE values than Block-based LSTM.

In the Fig. 4.20, the behaviors of the two methods are examined for the high storm

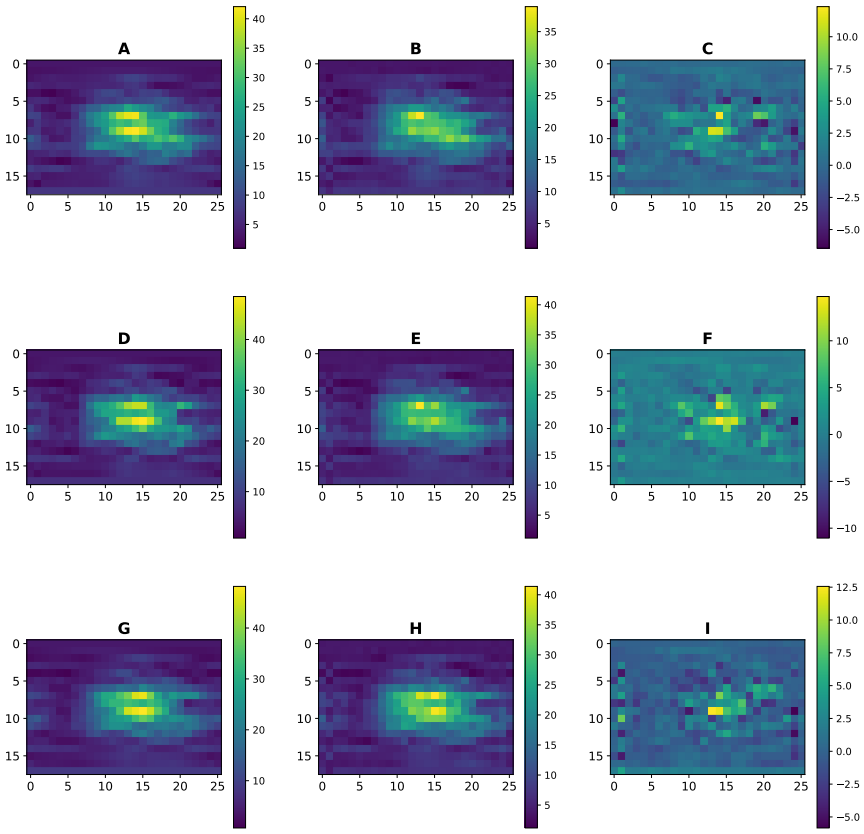

Figure 4.19: Comparison of consecutive hours for Block based forecasting in August the 26th. A) is actual coefficients at 05:00 UT ; B) predicted coefficients at 05:00 UT; C) residual at 05:00 UT; D) actual coefficients at 06:00 UT; E) predicted coefficients at 06:00 UT; F) residual at 06:00 UT ;,G) actual coefficients at 07:00 UT; H) predicted coefficients at 07:00 UT; I) residuals at 07:00 UT

hour (26/08/2018 06:00 UT). In some high-residual pixels of Block LSTM [24,9], PCA SARIMA achieved better results. For some pixels such as [21,8], we can say that better results are obtained with Block LSTM. In PCA-LSTM, a more successful result was produced for [14,7] pixel when compared to the others. This result may indicate a potential of using hybrid or ensamble methods for increasing accuracy of estimates. The reconstructed forecasts of VTEC maps using the PCA based SARIMA models are given in Figs. 4.21, 4.22, 4.23. The results indicate that the forecast maps have RMSE differences with respect to the IONEX grid in the levels of less than 1 TECU.

Table 4.4: Mean and Variance of MSE Residual of 3 Hour in 26.08.2018 using Block-Pixel Method

Hour	Mean	Variance
5	0.324	3.819
6	0.415	6.567
7	0.226	3.833

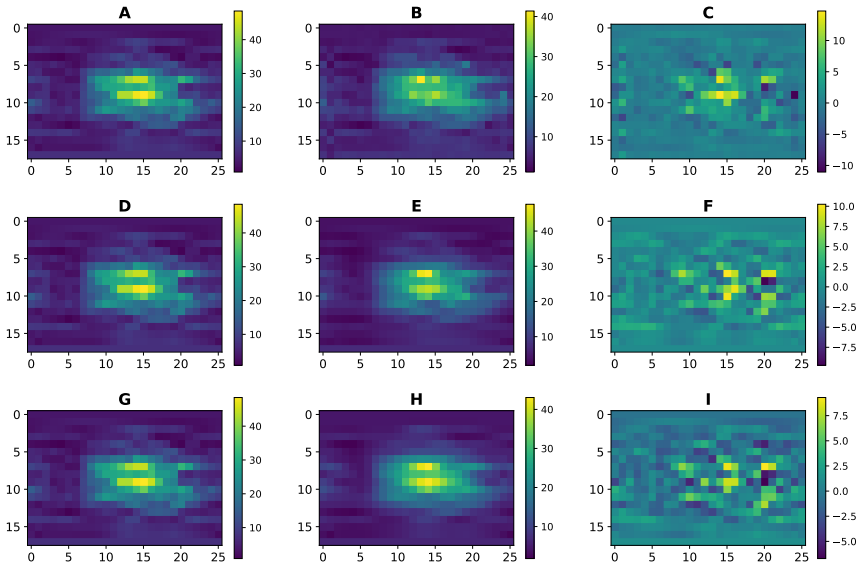

Figure 4.20: Comparison of reconstructed Coefficients from Block-LSTM, PCA-SARIMA and PC-LSTM forecast models in August the 26th. A) is actual coefficients at 06:00 UT ; B) predicted coefficients from Block-LSTM at 06:00 UT; C) residual A-B; D) actual coefficients at 06:00 UT; E) predicted coefficients from PCA-SARIMA at 06:00 UT; F) residual D-E; G) actual coefficients at 06:00 UT; H) predicted coefficients fromPCA-LSTM at 06:00 UT; I) residuals G-H

In addition, for the storm day, the RMSE increases up to 1.6 TECU levels, which is still lower then persistence based models. Considering the accuracy of IONEX grid, such an accuracy may indicate that the developed forecasting method may be used for both data assimilation and real-time positioning applications. Further investigation may reveal the position domain accuracy of the forecast maps.

Validation Days in 2010

In addition, other validation days have been selected from 2010, which corresponds to

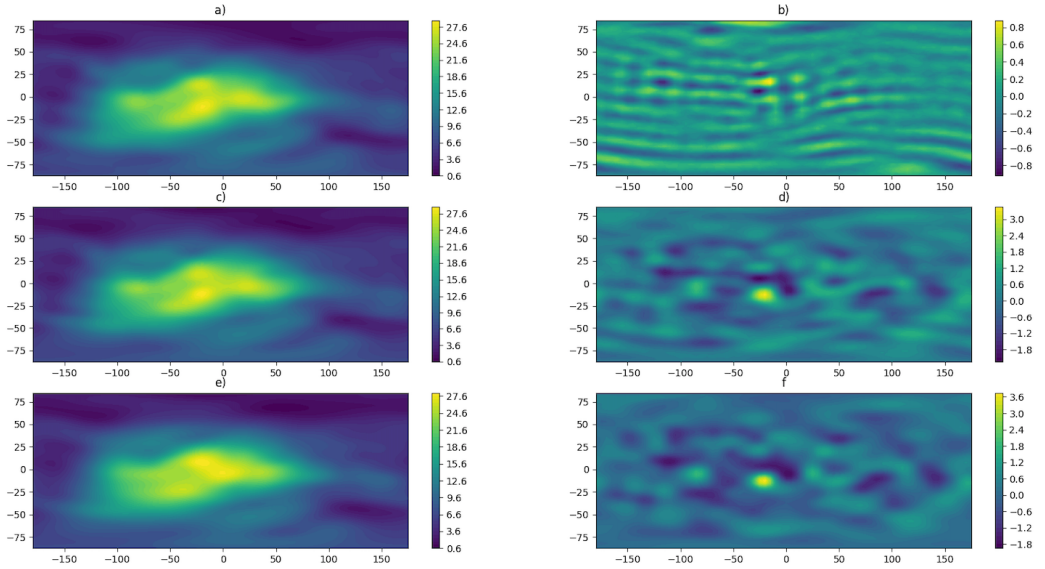

Figure 4.21: Forecasted VTEC map compared with IGS final IONEX Grid at 15/01/2018 15:00 UT. a) The IONEX Grid; b) differences between IONEX Grid and actual B-spline representation (RMSE = 0.1941 TECU);c) VTEC map with actual B-spline representation; d) differences predicted and actual B-spline VTEC maps (RMSE = 0.5167 TECU); e) forecasted VTEC map; f) differences between IONEX Grid and forecasted VTEC map (RMSE = 0.5520 TECU)

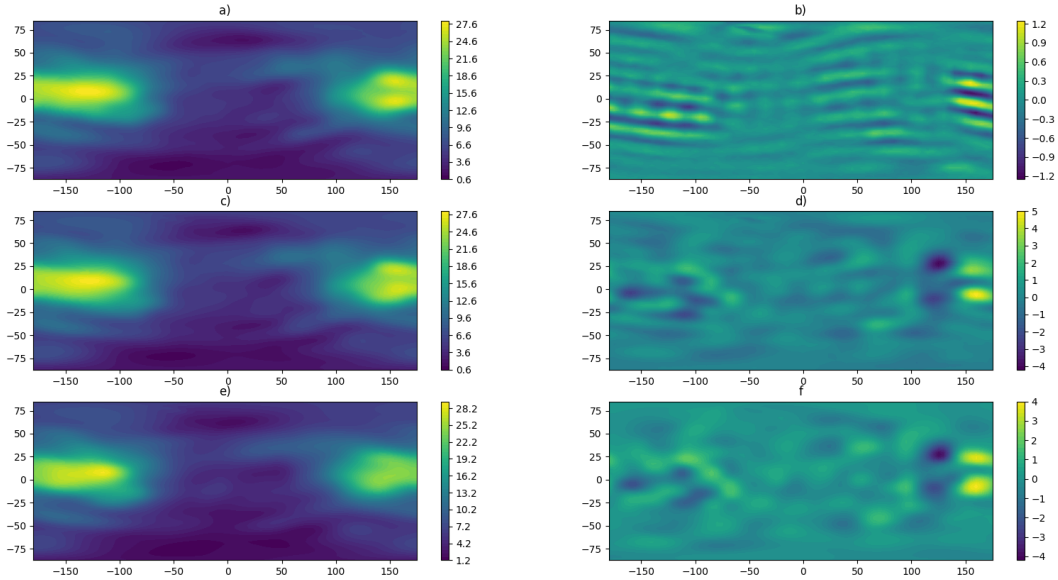

Figure 4.22: Forecasted VTEC map compared with IGS final IONEX Grid at 25/08/2018 01:00 UT. a) The IONEX Grid; b) differences between IONEX Grid and actual B-spline representation (RMSE = 0.2261 TECU);c) VTEC map with actual B-spline representation; d) differences predicted and actual B-spline VTEC maps (RMSE = 0.6542 TECU); e) forecasted VTEC map; f) differences between IONEX Grid and forecasted VTEC map (RMSE = 0.6921 TECU)

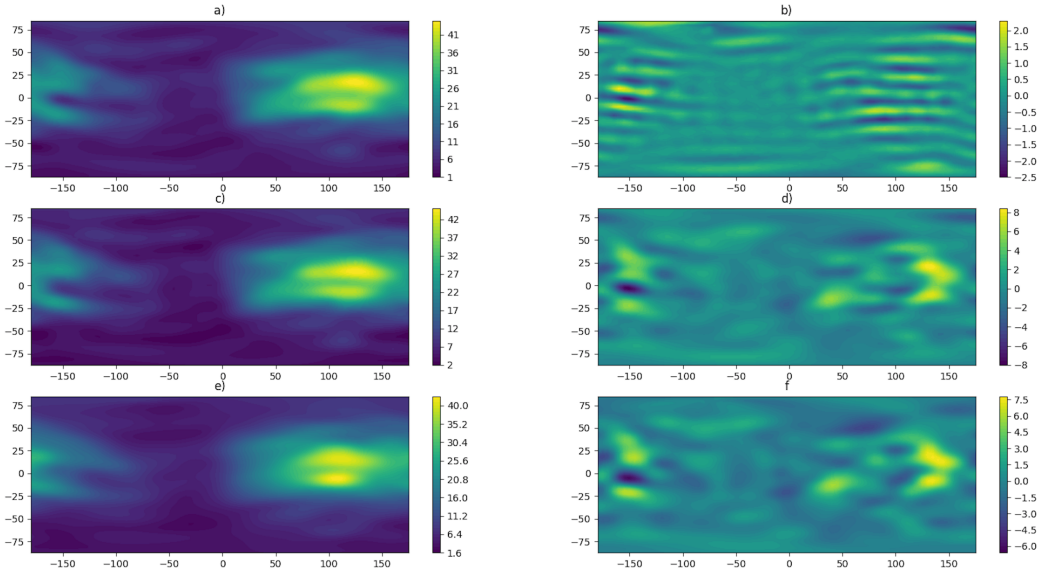

Figure 4.23: Forecasted VTEC map compared with IGS final IONEX Grid at 26/08/2018 06:00 UT. a) The IONEX Grid; b) differences between IONEX Grid and actual B-spline representation (RMSE = 0.5135 TECU);c) VTEC map with actual B-spline representation; d) differences predicted and actual B-spline VTEC maps (RMSE = 1.5619 TECU); e) forecasted VTEC map; f) differences between IONEX Grid and forecasted VTEC map (RMSE = 1.6441 TECU)

Table 4.5: Mean MSE of Validation Days

Days	Persistence	Block-Pixel	Sarima-PCA	LSTM-PCA	fb-LSTM-PCA
Quiet	2.679	2.282	1.174	2.122	3.331
Storm	2.534	2.372	1.708	2.343	5.840

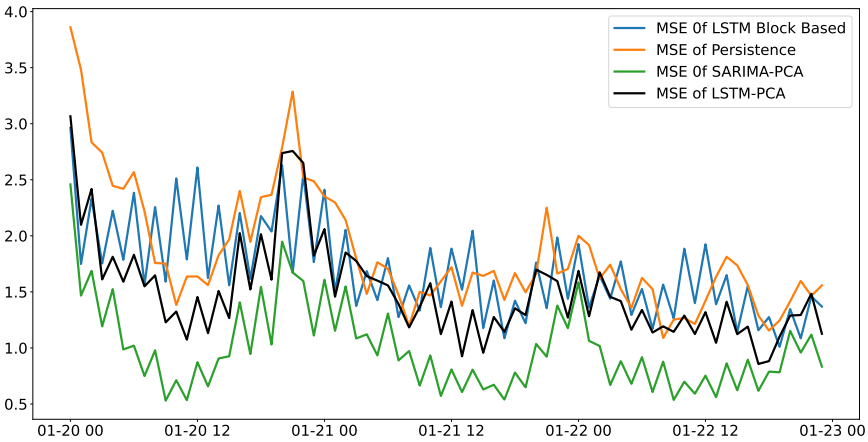

Figure 4.24: MSE of Days of January 20,21,22 2010 for four models

Table 4.6: Mean MSE of Choosing Days in 2010

Days	Persistence	Block-Pixel	Sarima-PCA	LSTM-PCA
January 20,21,22	1.847	1.699	0.990	1.503
June 20,21,22	0.912	1.044	0.430	1.025

dates earlier than the train data set. The first 3 days selected are in winter (January 20,21,22). The other ones are from summer (June 20,21,22). PCA-LSTM Prophet model results for trend prediction with Facebook Prophet are not included in this validation section 4.24, 4.25. According to the results, PCA SARIMA results are more successful for both times. The reconstructed forecasts of VTEC maps using the PCA based SARIMA models are given in Figs. 4.26, 4.27. The results indicate that the forecast maps have RMSE differences with respect to the IONEX grid in the levels of less than 1 TECU.

4.6 Discussion

The main purpose of this study is the short-term forecasting of Global VTEC maps. The estimation methods which are Single-Pixel, Block-Based and PCA based have been proposed. Basically, methods are based on the application of LSTM and SARIMA models. The first proposed method is the Single-Pixel method which generates a unique model for each individual pixel. Both SARIMA and LSTM models are considered

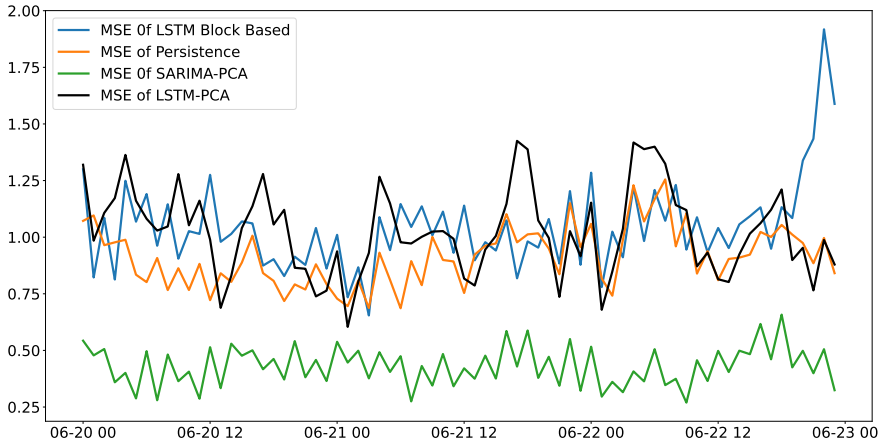

Figure 4.25: MSE of Days of June 20,21,22 2010 for four models

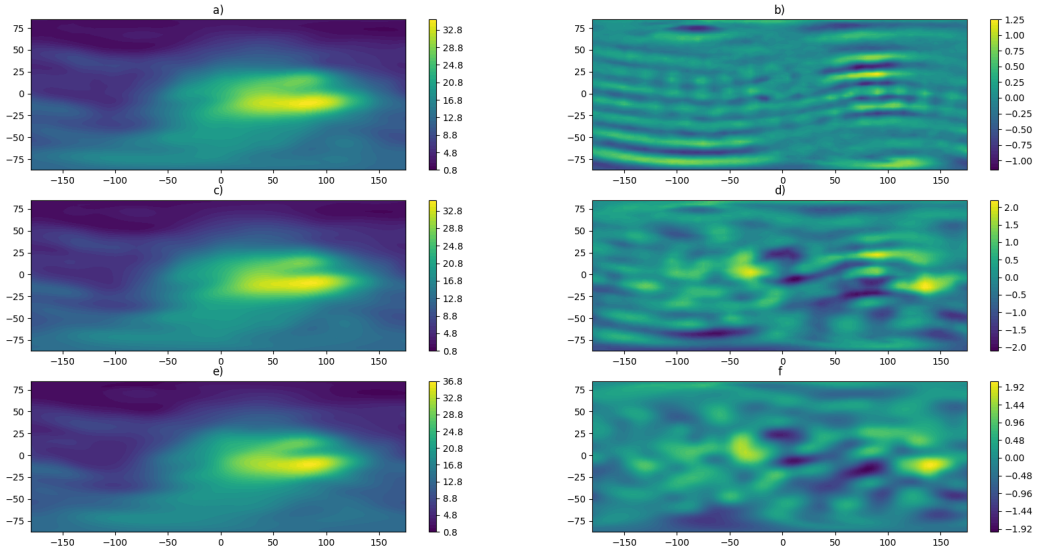

Figure 4.26: Forecasted VTEC map compared with IGS final IONEX Grid at 20/01/2010 11:00 UT. a) The IONEX Grid; b) differences between IONEX Grid and actual B-spline representation (RMSE = 0.2632 TECU);c) VTEC map with actual B-spline representation; d) differences predicted and actual B-spline VTEC maps (RMSE = 0.5918 TECU); e) forecasted VTEC map; f) differences between IONEX Grid and forecasted VTEC map (RMSE = 0.5300 TECU)

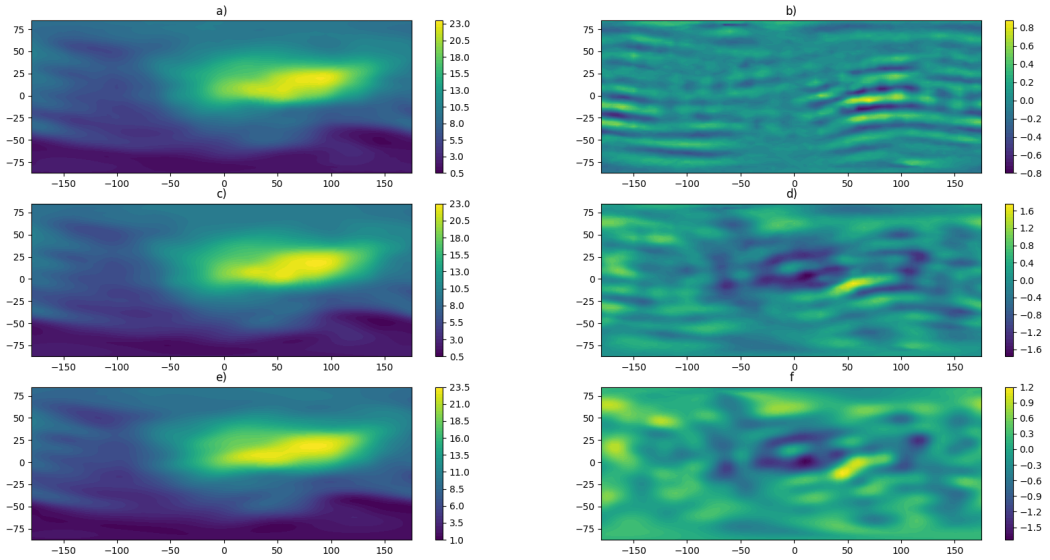

Figure 4.27: Forecasted VTEC map compared with IGS final IONEX Grid at 20/06/2010 11:00 UT. a) The IONEX Grid; b) differences between IONEX Grid and actual B-spline representation (RMSE = 0.1664 TECU);c) VTEC map with actual B-spline representation; d) differences predicted and actual B-spline VTEC maps (RMSE = 0.4268 TECU); e) forecasted VTEC map; f) differences between IONEX Grid and forecasted VTEC map (RMSE = 0.3930 TECU)

here. The results has not been accurate enough according to the model selection results. There is a spatial correlation between the pixels, that is the coefficients have spatial relationship with their neighbours. For this reason, two different methods are proposed. The first of these methods is the LSTM based Block-Pixel method. In this method, neighborhood pixels are taken as features. Especially in regions with high geomagnetic activity, the method achieved better results than Single-Based method. The average MSEs obtained from the model for the days selected for validation are 2.282 and 2.372 in quiet and storm days, respectively. We observed high MSE values in some periods due to high DST index. Overall results showed that the method achieved better results, especially at midnight.

Another method is the PCA-based forecasting method which includes SARIMA and LSTM. Here, it can be said that the results of SARIMA are more successful than LSTM potentially because SARIMA looks back more than 48 hours of previous time steps. In addition, the advantage of dimension reduction from 468 models to training the model for only 135 PCs is appealing in terms of storage, complexity and computing power. The average MSE calculated for the validation days(quiet and storm days) is 1.174 and

1.708, respectively. Here, it can be said that it is relatively more successful in storm days than Block-based method. Trend prediction with Prophet in model selection with LSTM in PCA method was more successful than keeping the trend constant. However, this did not provide improvements in the validation dataset. The average MSE calculated for the validation days (quiet and storm days) is 3.331 and 5.840, respectively. It may be the result of poor mean prediction of trend from Facebook Prophet. At the same time, the PCA-LSTM with Facebook Prophet model has not handled very successful in the storm days. The low success in PCA-LSTM may be a result of poor prediction result of Facebook Prophet. To investigate further, a model was obtained without using the Prophet trend. The average MSE for this model for Quiet days is 2.122, while the average MSE for Storm days is 2.343, which is even better than Block-Based LSTM. In summary, the results in Block-based LSTM and PCA-based LSTM provided an improvement of 15% and 21% on quiet days compared to the persistence model, respectively. For severe Storm days, the improvement reduces to is 6% and 7%. On the other hand, PCA-based SARIMA, results are 56% improvement on quiet days versus 32% improvement on storm days with respect to the persistence model.

For other validation days in 2010, the average MSE calculated for the validation days (January and June days) are 0.990 and 0.430, respectively. Here, it can be said that it is relatively more successful than Block-based method. Model results for PCA-LSTM with FB Prophet were not produced in this section because it does not give successful results. In PCA-LSTM without using the Prophet, The average MSE for this model for January days is 1.503, while the average MSE for June days is 1.025, which is again better than Block-Based LSTM. In summary, the results in Block-based LSTM and PCA-based LSTM provided an improvement of 8% and 18.6% on January days compared to the persistence model, respectively. For severe June days, the improvement could not be obtained for the selected days. On the other hand, PCA-based SARIMA, results are 46% improvement on January days versus 52.8% improvement on June days with respect to the persistence model.

The results from the SARIMA model are more successful than the results from the LSTM. The SARIMA model may have given good results by using more historical data (for both seasonal auto regressive part and also for moving average part). Adding the

historical hours determined for LSTM as a feature can provide improvements in the results. Using the correlation analysis of SARIMA as a feature selection for LSTM based models may also result in better performance. Also the spatial mean can be more non-stationary than other generated data. Therefore, modeling with LSTM can provide improvement here as well.

CHAPTER 5

CONCLUSION

The aim of this study is the short-term forecasting of the global ionosphere, which is an important topic both for applications to real-time single-frequency positioning and navigation and also for monitoring the ionosphere with data assimilation methods. For short-term forecasting, the Trigonometric B-Spline representation of global VTEC is considered. A 20 years of global VTEC B-spline coefficients are generated from final IGS ionosphere grids in IONEX format. For the forecasting of these coefficients models based on LSTM, which is a deep learning method and SARIMA, which is a statistical method are investigated. Also, the PCA method is used thanks to its dimension reduction properties. Two different methods are studied on the basis of trend. The first is to accept the trending as a constant offset, and the other is to calculate the annual and linear trend with the help of Facebook Prophet library. Three different methods are proposed for forecasting. The first is to work with single pixel, the second is to use block-based coefficients and the third is to use PCA transformation. For each method a model selection procedure applied to the best model for forecasting. Single pixel method resulted in the lowest performance among others. For this reason, the model has not been selected for further investigation. In addition, 24-hour ahead forecasting is also studied during model selection in order to have an opinion on model performance for long-term forecasting.

In block and PCA based modeling, 1-hour forecasting quality is tested on the days selected for validation with quiet and active ionospheric conditions. Quiet days include when the dst index above -50 nt, while on the severe storm days the dst index is lower than -120 nt. Looking at the validation results, the results in block-based LSTM

provided an improvement of 15% on quiet days compared to the persistence model. For Storm day, this achievement is 6%. In PCA-LSTM provided an improvement of 21% on quiet days compared to the persistence model. For Storm day, this achievement is 7%. In PCA-based SARIMA, results are 56% on quiet days versus 32% on storm days. The results of PCA-based LSTM using Prophet based trend is under-performing compared to other methods.

Looking at the other validation results in 2010 (January and June days), the results in block-based LSTM provided an improvement of 8% on January days compared to the persistence model. In PCA LSTM provided an improvement of 18.6% on June days compared to the persistence model. In PCA-based SARIMA, results are 46% on January days versus 52.8% on June days.

Although the results provided with PCA based SARIMA model are quite successful, further investigation is required to improve the deficiencies of the LSTM based methods. For example, there are coefficients selected from two different regions. For the days selected of validation, the coefficient [9,13] is well modeled by LSTM may indicate that we can also achieve good results in when some form of feature engineering is incorporated in the model selection phase. For example, other features, namely DST, KP and solar flux indexes can be added to better model the equatorial anomaly. The methods can be selected based on the local forecasting results. In addition some form of hybrid methods or ensamble methods may achieve better forecasting performance.

One of the LSTM-based methods, for example, the Conv2DLSTM method, may also provide better results if necessary computing power can be utilized. Global coefficient maps can be used as a a sequence of images and automated feature extraction power of CNNs can improve the results especially for storm days. New models can be produced with different values for the parameters kept constant in the current study. Or the range of modified parameters can be expanded. Cross validation method, which cannot be done due to lack of computing power, can be incorporated into the model selection phase. Trying different architectures for LSTM models may increase the overall success rate.

REFERENCES

[1] B. Zolesi and L. R. Cander, *Ionospheric prediction and forecasting*. 2014.

[2] S. Schaer, "Mapping and predicting the Earths ionosphere uisng GPS," p. 228, 1999.

[3] M. Durmaz, *Non-Parametric and Semi-Parametric Regional Modeling of the Ionospheric Vertical Total Electron Content Using Ground-Based Gps Observations*. PhD thesis, 2013.

[4] P. Teunissen and O. Montenbruck, eds., *Springer Handbook of Global Navigation Satellite Systems*. Gewerbestrasse 11, 6330 Cham, Switzerland: Springer International Publishing AG, 2017.

[5] E. Erdogan, M. Schmidt, F. Seitz, and M. Durmaz, "Near real-Time estimation of ionosphere vertical total electron content from GNSS satellites using B-splines in a Kalman filter," *Annales Geophysicae*, vol. 35, no. 2, pp. 263–277, 2017.

[6] M. Hernández-Pajares, J. M. Juan, J. Sanz, R. Orus, A. Garcia-Rigo, J. Feltens, A. Komjathy, S. C. Schaer, and A. Krankowski, "The IGS VTEC maps: A reliable source of ionospheric information since 1998," *Journal of Geodesy*, vol. 83, no. 3-4, pp. 263–275, 2009.

[7] N. Cherrier, T. Castaings, and A. Boulch, "Deep sequence-to-sequence neural networks for ionospheric activity map prediction," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10638 LNCS, pp. 545–555, 2017.

[8] C. Wang, S. Xin, X. Liu, C. Shi, and L. Fan, "Prediction of global ionospheric VTEC maps using an adaptive autoregressive model," *Earth, Planets and Space*, vol. 70, no. 1, 2018.

[9] L. Liu, S. Zou, Y. Yao, and Z. Wang, "Forecasting Global Ionospheric TEC Using Deep Learning Approach," *Space Weather*, vol. 18, no. 11, pp. 1–12, 2020.

[10] E. Tulunay, E. T. Senalp, S. M. Radicella, and Y. Tulunay, "Forecasting total electron content maps by neural network technique," *Radio Science*, vol. 41, no. 4, pp. 1–12, 2006.

[11] R. Song, X. Zhang, C. Zhou, J. Liu, and J. He, "Predicting TEC in China based on the neural networks optimized by genetic algorithm," *Advances in Space Research*, vol. 62, no. 4, pp. 745–759, 2018.

[12] W. Sun, L. Xu, X. Huang, W. Zhang, T. Yuan, Z. Chen, and Y. Yan, "Forecasting of ionospheric vertical total electron content (TEC) using LSTM networks," *Proceedings of 2017 International Conference on Machine Learning and Cybernetics, ICMLC 2017*, vol. 2, pp. 340–344, 2017.

[13] I. Srivani, G. Siva Vara Prasad, and D. Venkata Ratnam, "A Deep Learning-Based Approach to Forecast Ionospheric Delays for GPS Signals," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1180–1184, 2019.

[14] I. L. Mallika, D. V. Ratnam, Y. Ostuka, G. Sivavaraprasad, and S. Raman, "Implementation of Hybrid Ionospheric TEC Forecasting Algorithm Using PCA-NN Method," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 371–381, 2019.

[15] M. Reza, G. Razin, and B. Voosoghi, "Ionosphere time series modeling using adaptive neuro - fuzzy inference system and principal component analysis," *GPS Solutions*, vol. 24, no. 2, pp. 1–13, 2020.

[16] R. Niu, C. Guo, Y. Zhang, L. He, and Y. Mao, "Study of ionospheric TEC short-term forecast model based on combination method," *International Conference on Signal Processing Proceedings, ICSP*, vol. 2015-Janua, no. October, pp. 2426–2430, 2014.

[17] A. V. Zhukov, Y. V. Yasyukevich, and A. E. Bykov, "GIMLi: Global Ionospheric total electron content model based on machine learning," *GPS Solutions*, vol. 25, no. 1, pp. 1–9, 2021.

[18] M. Schmidt, D. Dettmering, M. Mößmer, Y. Wang, and J. Zhang, "Comparison of spherical harmonic and B spline models for the vertical total electron content," *Radio Science*, vol. 46, no. 4, pp. 1–8, 2011.

[19] F. Chollet, *Deep Learning with Python*. Manning, Nov. 2017.

[20] C. Mastrangelo, *Time Series Analysis and Forecasting by Example*, vol. 44. 2012.

[21] G. Hajj, L. Lee, X. Pi, L. Romans, W. Schreiner, P. Straus, and C. Wang, "COSMIC GPS ionospheric sensing and space weather," *TERRESTRIAL ATMOSPHERIC AND OCEANIC SCIENCES*, vol. 11, pp. 235–272, MAR 2000.

[22] J. Tang, Y. Yao, L. Zhang, and J. Kong, "Tomographic reconstruction of ionospheric electron density during the storm of 5-6 August 2011 using multi-source data," *Scientific Reports*, vol. 5, no. July, pp. 1–11, 2015.

[23] S. Günter, *Satellite Geodesy*. De Gruyter, 2008.

[24] D. Bilitza, "International Reference Ionosphere 2000 of ionospheric It was and the F peak down to," *Radio Science*, vol. 36, no. 2, pp. 261–275, 2001.

[25] D. Bilitza and B. W. Reinisch, "International Reference Ionosphere 2007: Improvements and new parameters," *Advances in Space Research*, vol. 42, no. 4, pp. 599–609, 2008.

[26] B. Nava, P. Coïsson, and S. M. Radicella, "A new version of the NeQuick ionosphere electron density model," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 70, no. 15, pp. 1856–1862, 2008.

[27] S. Schaer, W. Gurtner, J. Feltens, and E. S. A. Esoc, "IONEX : The IONosphere Map EXchange Format Version 1," pp. 1–15, 1998.

[28] R. W. Schunk, L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, M. Hairston, and B. M. Howe, "Global Assimilation of Ionospheric Measurements (GAIM)," *Radio Science*, vol. 39, no. 1, pp. 211–221, 2004.

[29] M. Schmidt, "Wavelet modelling in support of IRI," *Advances in Space Research*, vol. 39, no. 5, pp. 932–940, 2007.

[30] S. J. Taylor and B. Letham, "Forecasting at Scale," *American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[31] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting - Second Edition*. 2002.

[32] R. Tang, F. Zeng, Z. Chen, J.-S. Wang, C.-M. Huang, and Z. Wu, "The Comparison of Predicting Storm-time Ionospheric TEC by Three Methods: ARIMA, LSTM, and Seq2Seq," *Atmosphere*, vol. 11, no. 4, p. 316, 2020.

[33] X. Li and D. Guo, "Modeling and prediction of ionospheric total electron content by time series analysis," *Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, vol. 2, pp. 375–379, 2010.

[34] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[35] R. B. Harris and R. G. Mach, "GPSTk–An Open Source GPSTk Toolkit," *GPS Solutions*, vol. 11, March 2007.

[36] S. K. Morley, J. Koller, D. T. Welling, B. A. Larsen, M. G. Henderson, and J. T. Niehof, "Spacepy - a python-based library of tools for the space sciences," in *Proceedings of the 9th Python in science conference (SciPy 2010)*, (Austin, TX), 2011.

[37] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[38] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

APPENDIX A

ADDITIONAL RESULTS

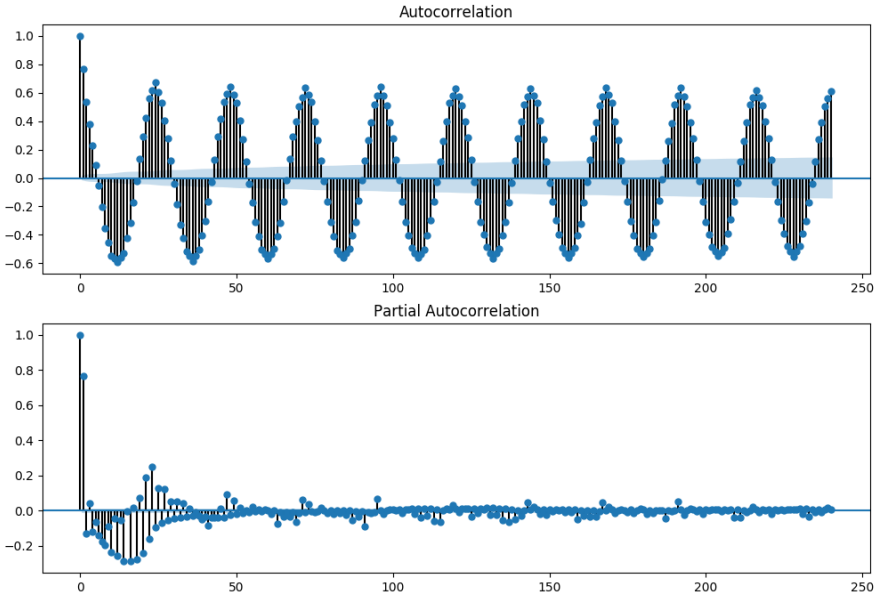

Figure A.1: ACF and PACF of First Difference VTECs[0,13] B-Spline Coefficient

Table A.1: Model Name of LSTM Model Selection

Model Name	(Layers,Neurons,prev,future)
Mdl 1	(1,(32),6,1)
Mdl 2	(1,(32),12,1)
Mdl 3	(1,(64),6,1)
Mdl 4	(1,(64),12,1)
Mdl 5	(1,(128),6,1)
Mdl 6	(1,(128),12,1)
Mdl 7	(2,(32,16),6,1)
Mdl 8	(2,(32,16),12,1)
Mdl 9	(2,(64,32),6,1)
Mdl 10	(2,(64,32),12,1)
Mdl 11	(2,(128,64),6,1)
Mdl 12	(2,(128,64),12,1)
Mdl 13	(1,(32),6,24)
Mdl 14	(1,(32),12,24)
Mdl 15	(1,(64),6,24)
Mdl 16	(1,(64),12,24)
Mdl 17	(1,(128),6,24)
Mdl 18	(1,(128),12,24)
Mdl 19	(2,(32,16),6,24)
Mdl 20	(2,(32,16),12,24)
Mdl 21	(2,(64,32),6,24)
Mdl 22	(2,(64,32),12,24)
Mdl 23	(2,(128,64),6,24)
Mdl 24	(2,(128,64),12,24)

Table A.2: Model Name of SARIMA Model Selection

Model Name	(p,q)x(P,Q)	Model Name	(p,q)x(P,Q)
M 1	(1, 1, 0, 0)	M 28	(2, 2, 1, 1)
M 2	(1, 1, 0, 1)	M 29	(2, 2, 2, 0)
M 3	(1, 1, 1, 0)	M 30	(2, 2, 2, 1)
M 4	(1, 1, 1, 1)	M 31	(2, 3, 0, 0)
M 5	(1, 1, 2, 0)	M 32	(2, 3, 0, 1)
M 6	(1, 1, 2, 1)	M 33	(2, 3, 1, 0)
M 7	(1, 2, 0, 0)	M 34	(2, 3, 1, 1)
M 8	(1, 2, 0, 1)	M 35	(2, 3, 2, 0)
M 9	(1, 2, 1, 0)	M 36	(2, 3, 2, 1)
M 10	(1, 2, 1, 1)	M 37	(3, 1, 0, 0)
M 11	(1, 2, 2, 0)	M 38	(3, 1, 0, 1)
M 12	(1, 2, 2, 1)	M 39	(3, 1, 1, 0)
M 13	(1, 3, 0, 0)	M 40	(3, 1, 1, 1)
M 14	(1, 3, 0, 1)	M 41	(3, 1, 2, 0)
M 15	(1, 3, 1, 0)	M 42	(3, 1, 2, 1)
M 16	(1, 3, 1, 1)	M 43	(3, 2, 0, 0)
M 17	(1, 3, 2, 0)	M 44	(3, 2, 0, 1)
M 18	(1, 3, 2, 1)	M 45	(3, 2, 1, 0)
M 19	(2, 1, 0, 0)	M 46	(3, 2, 1, 1)
M 20	(2, 1, 0, 1)	M 47	(3, 2, 2, 0)
M 21	(2, 1, 1, 0)	M 48	(3, 2, 2, 1)
M 22	(2, 1, 1, 1)	M 49	(3, 3, 0, 0)
M 23	(2, 1, 2, 0)	M 50	(3, 3, 0, 1)
M 24	(2, 1, 2, 1)	M 51	(3, 3, 1, 0)
M 25	(2, 2, 0, 0)	M 52	(3, 3, 1, 1)
M 26	(2, 2, 0, 1)	M 53	(3, 3, 2, 0)
M 27	(2, 2, 1, 0)	M 54	(3, 3, 2, 1)

Table A.3: Block LSTM Validation MSE of 2 Day Global B-Spline Coeff. in 2018

Day	Hour	LSTM Model	Persistence model
16	0	2.8774	4.0924
	1	1.8567	4.0804
	2	1.983	3.0951
	3	1.8816	3.6695
	4	1.8423	2.5327
	5	1.8968	2.4435
	6	2.5341	2.8218
	7	1.6489	2.4308
	8	2.2051	2.6536
	9	1.988	3.1693
	10	2.5887	2.6459
	11	2.3209	2.7507
	12	2.5837	2.4056
	13	1.9596	2.0649
	14	2.9661	2.7823
	15	2.4482	2.3699
	16	2.597	2.2926
	17	1.832	2.4096
	18	2.0056	2.1883
	19	1.9419	2.3065
	20	3.001	3.2352
	21	2.1848	3.0791
	22	3.0158	3.6083
	23	2.2388	3.8252
238	0	4.1715	3.278
	1	2.8325	2.9618
	2	3.5527	3.3517
	3	3.1355	4.139
	4	3.8789	3.8634
	5	3.9245	3.8959
	6	6.7396	4.7863
	7	3.8843	3.3371
	8	4.6711	3.5832
	9	3.9835	5.1025
	10	4.64	5.0226
	11	2.7764	5.3602
	12	3.1892	3.8264
	13	2.5382	3.2916
	14	3.0022	3.0473
	15	2.5856	2.5058
	16	2.6402	1.9607
	17	2.7371	2.5494
	18	2.8398	2.127
	19	2.1186	2.0516
	20	2.7205	2.6413
	21	1.8817	1.8382
	22	2.095	1.541

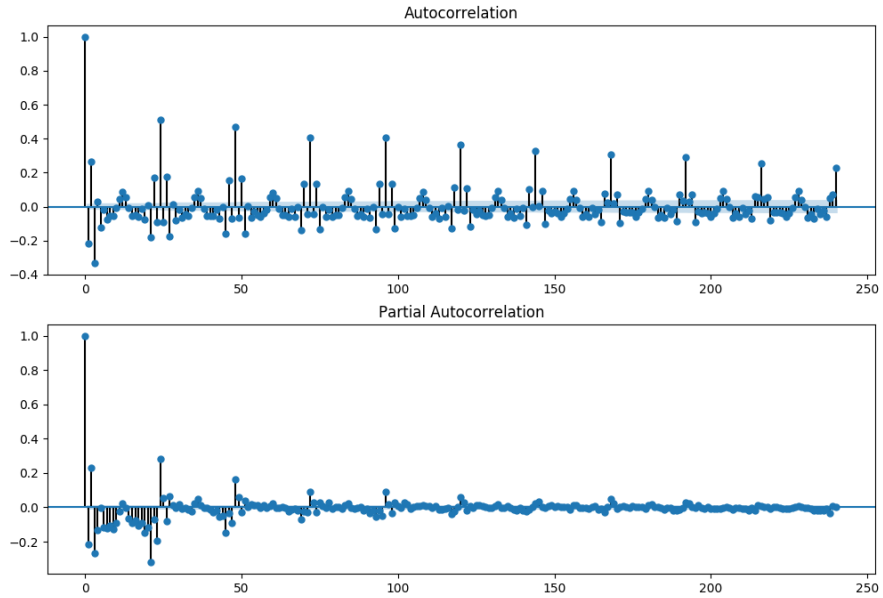

Figure A.2: ACF and PACF of First Difference VTECs[9,13] B-Spline Coefficient

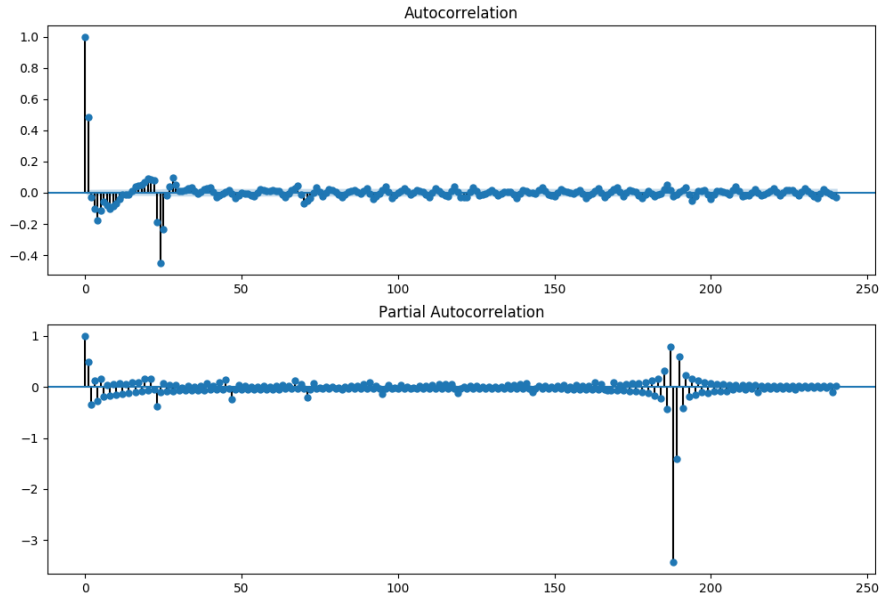

Figure A.3: ACF and PACF of First Difference and Seasonal Difference VTECs[0,13] B-Spline Coefficient

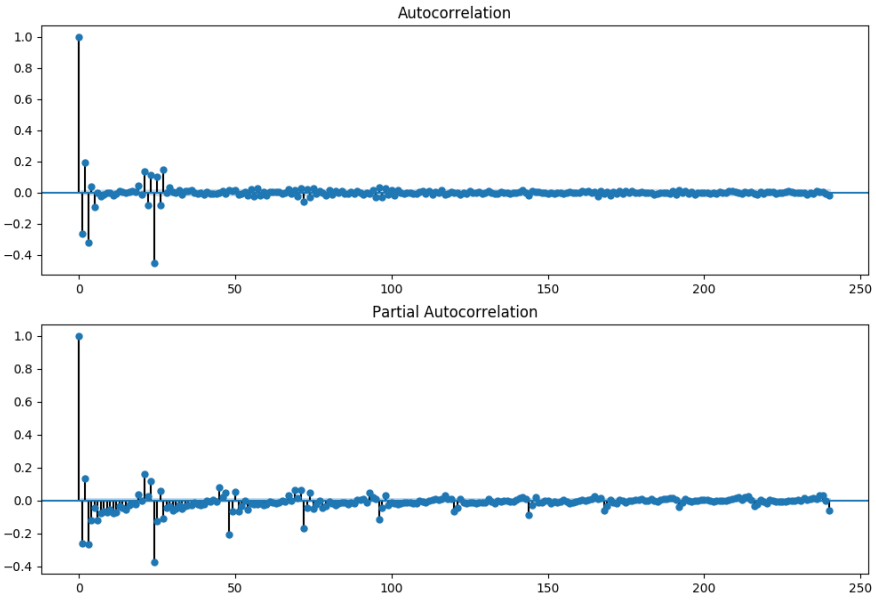

Figure A.4: ACF and PACF of First Difference and Seasonal Difference VTECs[9,13] B-Spline Coefficient