


Article

# Evaluation of F10.7, Sunspot Number and Photon Flux Data for Ionosphere TEC Modeling and Prediction Using Machine Learning Techniques

Andres Gilberto Machado da Silva Benoit<sup>1,2,\*</sup> and Adriano Petry<sup>1</sup> 

<sup>1</sup> Southern Space Coordination, National Institute for Space Research, São José dos Campos 12227-010, SP, Brazil; adriano.petry@inpe.br

<sup>2</sup> Mechanical Engineering Department, Federal University of Santa Maria, Santa Maria 97105-900, RS, Brazil

\* Correspondence: andres.benoit@acad.ufsm.br

**Abstract:** Considering the growing volumes and varieties of ionosphere data, it is expected that automation of analytical model building using modern technologies could lead to more accurate results. In this work, machine learning techniques are applied to ionospheric modeling and prediction using sun activity data. We propose Total Electron Content (TEC) spectral analysis, using discrete cosine transform (DCT) to evaluate the relation to the solar features F10.7, sunspot number and photon flux data. The ionosphere modeling procedure presented is based on the assessment of a six-year period (2014–2019) of data. Different multi-dimension regression models were considered in experiments, where each geographic location was independently evaluated using its DCT frequency components. The features correlation analysis has shown that 5-year data seem more adequate for training, while learning curves revealed overfitting for polynomial regression from the 4th to 7th degrees. A qualitative evaluation using reconstructed TEC maps indicated that the 3rd degree polynomial regression also seems inadequate. For the remaining models, it can be noted that there is seasonal variation in root-mean-square error (RMSE) clearly related to the equinox (lower error) and solstice (higher error) periods, which points to possible seasonal adjustment in modeling. Elastic Net regularization was also used to reduce global RMSE values down to 2.80 TECU for linear regression.

**Keywords:** ionosphere modeling; Total Electron Content; machine learning; regression model



**Citation:** Benoit, A.G.M.d.S.; Petry, A. Evaluation of F10.7, Sunspot Number and Photon Flux Data for Ionosphere TEC Modeling and Prediction Using Machine Learning Techniques. *Atmosphere* **2021**, *12*, 1202. <https://doi.org/10.3390/atmos12091202>

Academic Editors: Ljiljana R. Cander and Bruno Zolesi

Received: 6 July 2021

Accepted: 10 August 2021

Published: 16 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The study of ionosphere dynamics has attracted the attention of several research groups related to Space Weather area. This is mainly due to its known effects on electronic equipment, communication and power transmission [1], which are now part of our society's foundations. Among different measurements, ground-based GNSS stations and ionosondes can provide important ionosphere data at fixed locations, while low-earth orbiting satellites can provide atmospheric soundings in broad regions, including the oceans. However, ionosphere Total Electron Content (TEC) maps have gained prominence as one of the most important tools to evaluate current and future ionosphere behavior, since they can provide full diurnal information on large areas around the globe. To provide reliable TEC maps to the scientific community, since 1998 the International GNSS Service (IGS) has combined different sources of TEC maps, and made the final products freely available through its website [2]. Considering the lack of measurements around the globe, the development of ionosphere mathematical models is critical to this task. Besides, the importance of data interpolation for ionosphere models increases with the possibility of forecasting electron density variability over days. Since the Klobuchar model [3], which is still used in single frequency GPS systems, several other interesting empirical models have been proposed [4–7].

Machine learning is a computer science research field in which data analysis can provide the ability to learn without being explicitly programmed. When a deterministic solution is unknown to a problem, we can use machine learning techniques to derive a close-to-optimal solution, using a variety of correlated data. It also stands out in complex problems that traditional concepts are unable to solve [8]. Nowadays, machine learning models are present in a variety of fields, such as self-driving cars, biology systems and remote sensing.

Ionospheric modeling can benefit from machine learning techniques to achieve accurate models, using available features related to sun activity and, therefore, the ionosphere state. The solar radio flux at 10.7 cm (F10.7 index), corresponding to the frequency of 2800 MHz, is known to be an excellent indicator of solar activity [9]. Likewise, photon flux data at different frequency ranges correlate to ion formation and neutralization in the ionosphere. Also, the sunspot number follows the cyclical nature of the sun [9]. In this work, we evaluate the influence of these features for ionosphere TEC modeling and prediction.

## 2. Features Evaluation

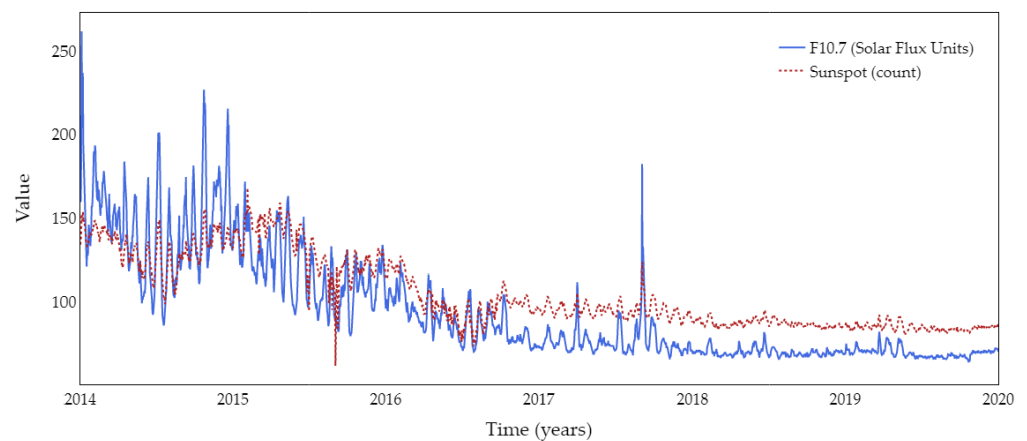
The direct influence of sun activity in ionosphere dynamics is well known. The main source of energy ionizing the molecules in the ionosphere comes from the sun. This energy is unequally distributed at different heights, depending on the solar radiation frequencies that hit the upper atmosphere, which causes variability in electron density. Besides that, earth rotation plays an important role in defining electron content daily change. Geographic regions with peak electronic concentration follow this movement to some extent, which explains periodic components observed in TEC. Similar behavior related to seasonal variation can also be observed. Thus, it is expected that sun activity measurements strongly correlate to the observed TEC change, and these features could be explored to model the quiet ionosphere dynamics. Obviously, other factors are important for defining the full variation of the ionosphere, in particular, the earth's magnetic field, the equatorial ionization anomaly (EIA) [10,11], special phenomena like Coronal Mass Ejections (CMEs) [12] that may cause significant ionospheric disturbances and plasma bubbles [13,14].

### 2.1. F10.7

The 10.7 cm wavelength solar radio emissions (F10.7) originate in the sun's chromosphere and have been measured since 1947. As mentioned in Section 1, these are widely used as an indicator of solar activity.

### 2.2. Sunspot Number

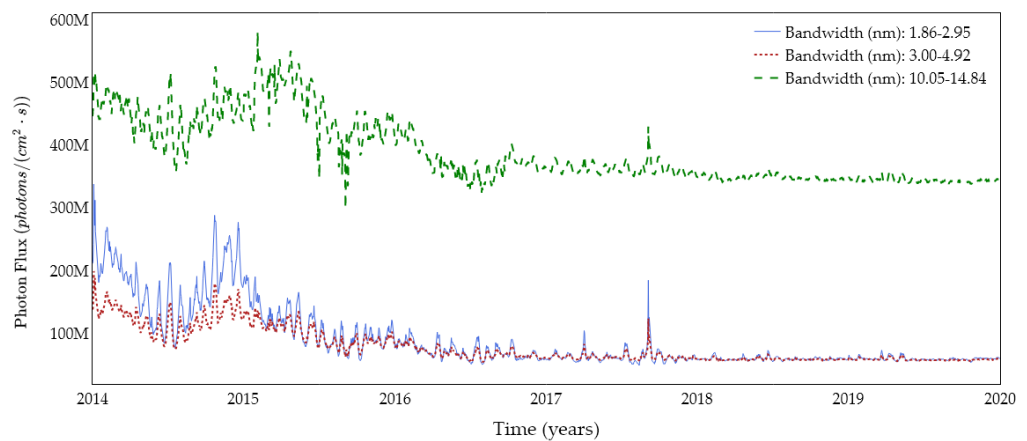
Sunspots are known as temporary phenomena of the Sun's atmosphere. They can be seen as darker spots in surrounding areas, and their occurrence varies cyclically [15]. Actually, sunspots are the sites of strong magnetic fields which are generated by the dynamo effect inside of the Sun. This feature has an important correlation with F10.7, as can be seen in Figure 1.



**Figure 1.** Variation of F10.7 and sunspot number from 2014 to 2020.

### 2.3. Photon Flux

The photon flux data estimate the number of photons striking the atmosphere in a given time per unit area, and impact the number of electrons which are produced in the ionosphere. They can be used to calculate the power density at different bandwidths. In this work, data for different bandwidths are used, and Figure 2 shows the variation for three of them.



**Figure 2.** Variation in photon flux data for 3 different bandwidths.

### 3. Ionosphere Modeling

The ionosphere modeling procedure presented is based on the assessment of a six-year period (2014–2019) of data. Global TEC maps were provided by IGS [2] from the National Aeronautics and Space Administration (NASA) website with a spatial–temporal resolution of  $2.5^\circ \times 5.0^\circ \times 2$  h for latitude, longitude and time, respectively. The correspondent daily F10.7 index, sunspot number and photon flux data were obtained using the Solar Irradiance Platform tool [16]. Each geographic location in TEC maps was evaluated independently, and the corresponding sequence of TEC values available for a given day is first passed to the frequency domain by applying the well known Discrete Cosine Transform (DCT) [17], which shows better spectral compaction and energy concentration properties than Discrete Fourier Transform (DFT), defined as:

$$X_k = C_k \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x_n \cos \frac{\pi}{N} \left(n + \frac{1}{2}\right)k, \quad (1)$$

where  $X_k$  is the  $k$ th frequency component,  $x_n$  is the  $n$ th data point to transform and  $N$  is the number of data points. Also, the inverse discrete cosine transform (IDCT) can be defined as:

$$x_n = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} C_k X_k \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad \text{for } i = 0, 1, \dots, N-1, \quad (2)$$

where

$$C_k = \sqrt{\frac{1}{2}} \quad \text{for } k = 0$$

$$C_k = 1 \quad \text{for } k = 1, 2, \dots, N-1.$$

Then, the DCT frequency components are associated to the correspondent features. This process was repeated daily for the period of analysis. To illustrate the relation between each one of the 12 DCT frequency components (obtained from the 12 TEC data available in a day) and F10.7, sunspot number and photon flux data at bandwidth 1.86 to 2.95 nm, Figures 3–5 show the results for a specific location (latitude 22.5° S and longitude 50° W). From observation of the Figures, it can be seen that low-frequency components concentrate most of the information, as expected. Plots of higher order DCT components show a more flat dispersion around the horizontal axis, where a simple linear regression would present declivity close to zero, thus less correlated to the features.

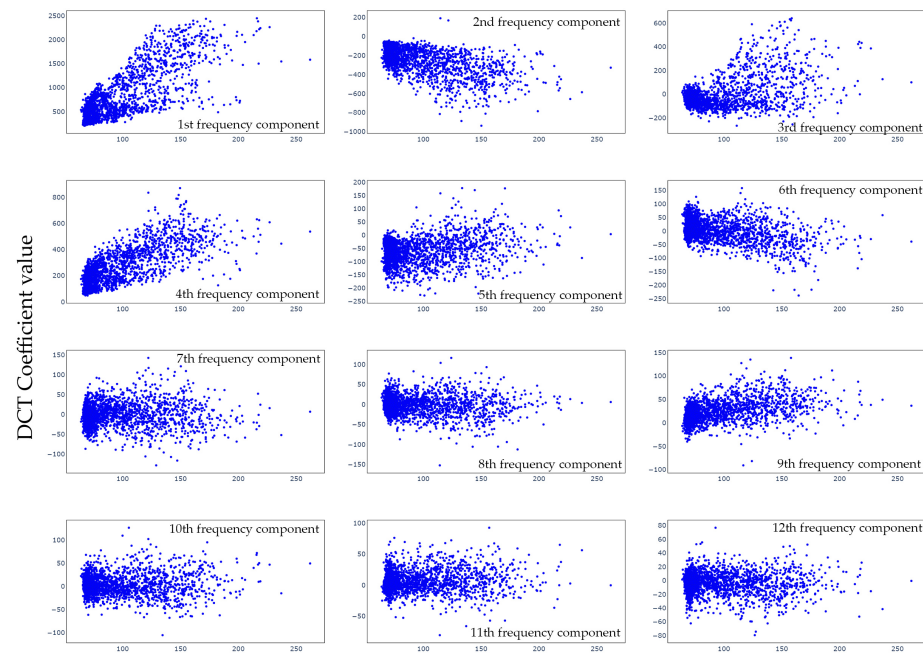


Figure 3. Amplitude of TEC frequency components versus F10.7.

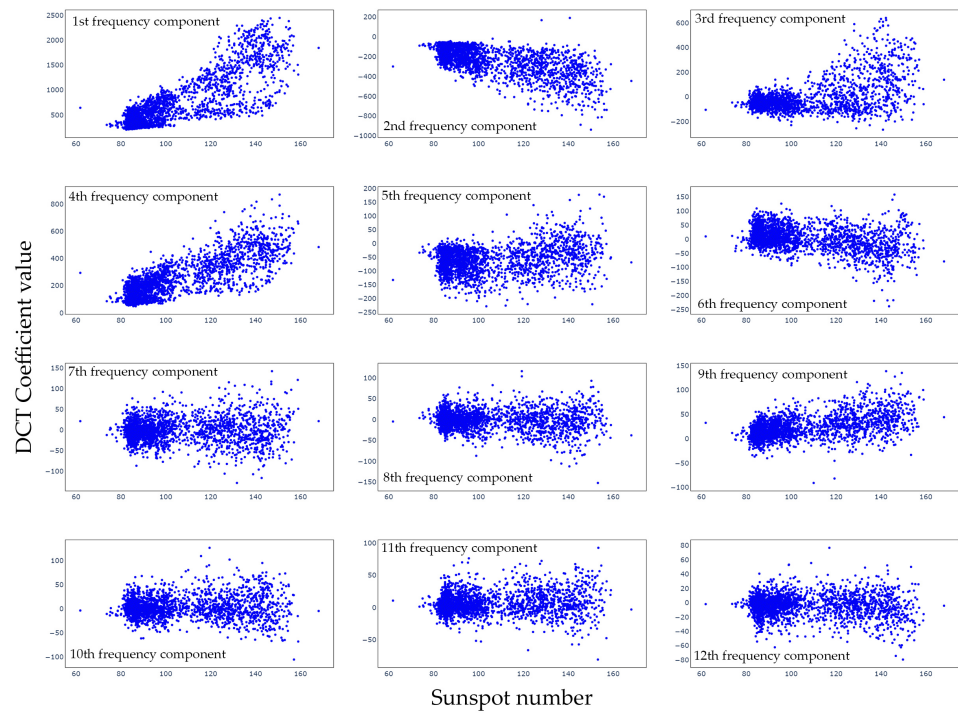


Figure 4. Amplitude of TEC frequency components versus sunspot number.

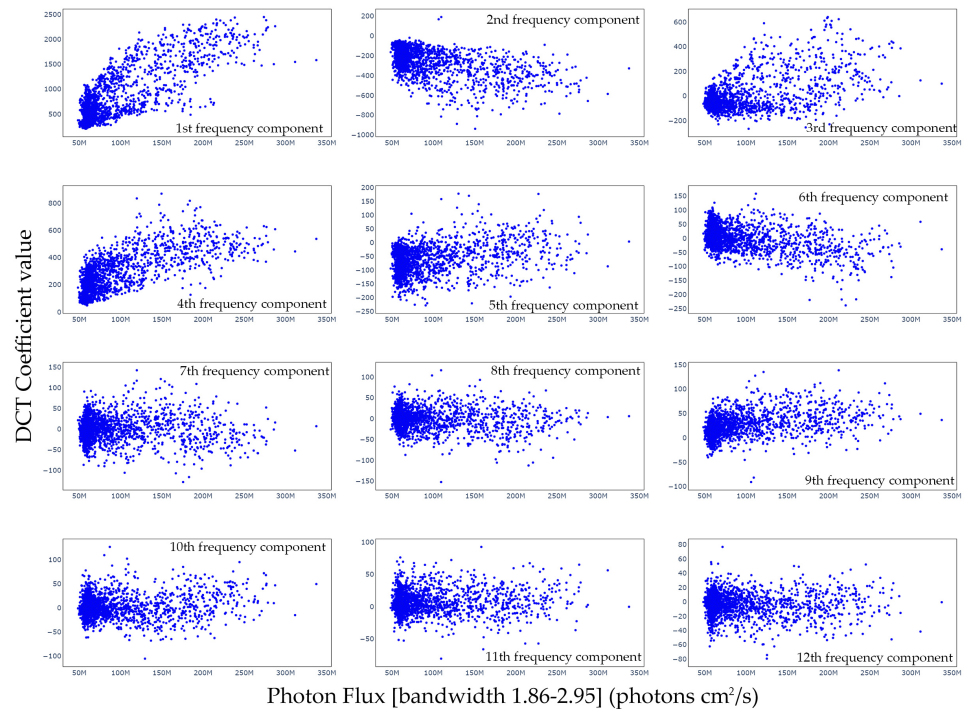


Figure 5. Amplitude of TEC frequency components versus Photon Flux at 1.86–2.95 nm bandwidth.

After passing TEC to the frequency domain, a multi-dimension regression technique can be used to estimate the best fit to the data. The resulting curves (one for every DCT coefficient, in every geographic location) constitute the ionosphere model used in the experiments. There are different types of regression methods that can be used, and our experiments focused on linear, polynomial and Support Vector Machine (SVM) [8]—which is not a proper regression method but can be adapted to be used as such.

### 3.1. Linear Regression

The multi-dimensional linear regression model [18] can be defined as

$$Y(X) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_k X_k + \epsilon, \quad (3)$$

where  $\theta_0$  is the linear coefficient,  $X = X_1, \dots, X_k$  are the feature values and  $k$  is the number of total features used.  $Y(X)$  is defined as the predicted value, in our context the TEC frequency component value estimated for a set of features, and  $\epsilon$  is the model error of estimation. Considering  $n$  observations, the model can be represented in a compact matrix notation [18]:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},$$

$$X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{pmatrix},$$

$$\theta = \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_k \end{pmatrix} \quad \text{and} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

and the Equation (3) can be written in a simpler form as

$$Y = X\theta + \epsilon. \quad (4)$$

The main key of this regression is to estimate the optimal  $\theta$  coefficients which minimize a cost function. The most popular approach used is the least squares [19], that estimates  $\theta$  to minimize the residual sum of squared errors, according to Equation (5).

$$\hat{\theta} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (5)$$

### 3.2. Polynomial Regression

Polynomial regression [18] makes use of linear model approach, but considers an  $m$ th degree polynomial fit to data by adding a power of each feature as a new feature, and then applying the multiple linear regression. An example for feature data is shown in matrix notation in Equation (6).

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 & X_1^2 & \dots & X_1^m \\ 1 & X_2 & X_2^2 & \dots & X_2^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & X_n^2 & \dots & X_n^m \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{pmatrix} + \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}, \quad (6)$$

where  $n$  the number of data available and  $\theta_1, \dots, \theta_m$  are the regression coefficients to be estimated to minimize a cost function, which can be calculated with the least squares method presented in Section 3.1.

### 3.3. Support Vector Machine

Support vector machines (SVMs) [20] have been used successfully in supervised data set classification. Linear SVMs are effective when data sets show an approximate linear distribution yet contain some outliers. In several cases a non-linear class separation, instead of a hyperplane, is more adequate. However, it is the reversed SVM [21] that can be used to implement linear and non-linear regression involving a higher number of features while limiting margin violations. The width of the margin is controlled by a hyperparameter  $\epsilon$ , wherein a larger number of  $\epsilon$  would introduce more error in the solution [8], otherwise, a lower number would penalize every error and require more computational resources. To implement a polynomial regression with SVM, a kernelized SVM model was used, and a new hyperparameter  $C$  was added to control the regularization of the model with  $L_2$  regularization (Section 3.4). Lower values of  $C$  add more regularization to the model, decreasing the overfitting but requiring more computational power. In this work we used  $\epsilon = 0.4$  and  $C = 50$ , following [21], to balance the error introduced in the model, overfitting and regularization.

### 3.4. Regularization Methods

In multiple dimension regression, when several features are used, the probability of overfitting the model increases. In addition, when the main goal is to explain a phenomenon, a model with a low number of features, like three or four, can be more useful than a model with hundreds of features [22]. Therefore, regularization can be used in the model to decrease the overfitting by adding to the cost function of the model a regularization term in the training set. This forces the algorithm not to adjust and to maintain the weight of  $\theta$ s as the smallest possible [8].

$L_1$  norm regularization, also known as Least Absolute Shrinkage and Selection Operator (LASSO) regression adds to the cost function the module of  $L_1$  norm of the coefficients vector [23]. Hence, if the performance of the model is the mean squared error (MSE), the LASSO cost function can be described by Equation (7) [24].

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^k |\theta_i|, \quad (7)$$

where  $k$  is the number of coefficients (features),  $\alpha$  is the hyperparameter which controls how much the model is regularized. For instance, if  $\alpha = 0$  we have a regular linear regression. Also, it is important to notice that the intercept (or polarized term) is not regularized [8], so the summation index can start from a different number depending on the indexing method.

The  $L_2$  regularization (also known as Ridge regression), is based on the same principle but adds to the cost function the half of squared  $L_2$  vector norm:

$$J(\theta) = MSE(\theta) + \frac{1}{2} \alpha \sum_{i=1}^k |\theta_i|^2 \quad (8)$$

It is important to mention that for SVM,  $\alpha$  is inversely proportional to  $C$ . Lastly, another regularization method which combines the  $L_1$  and  $L_2$  penalties by a mixing hyperparameter  $r$ , called Elastic Net, can be used. The cost function for Elastic Net regression can be seen in Equation (9).

$$J(\theta) = MSE(\theta) + r\alpha \sum_{i=1}^k |\theta_i| + \frac{1-r}{2} \alpha \sum_{i=1}^k |\theta_i|^2 \quad (9)$$

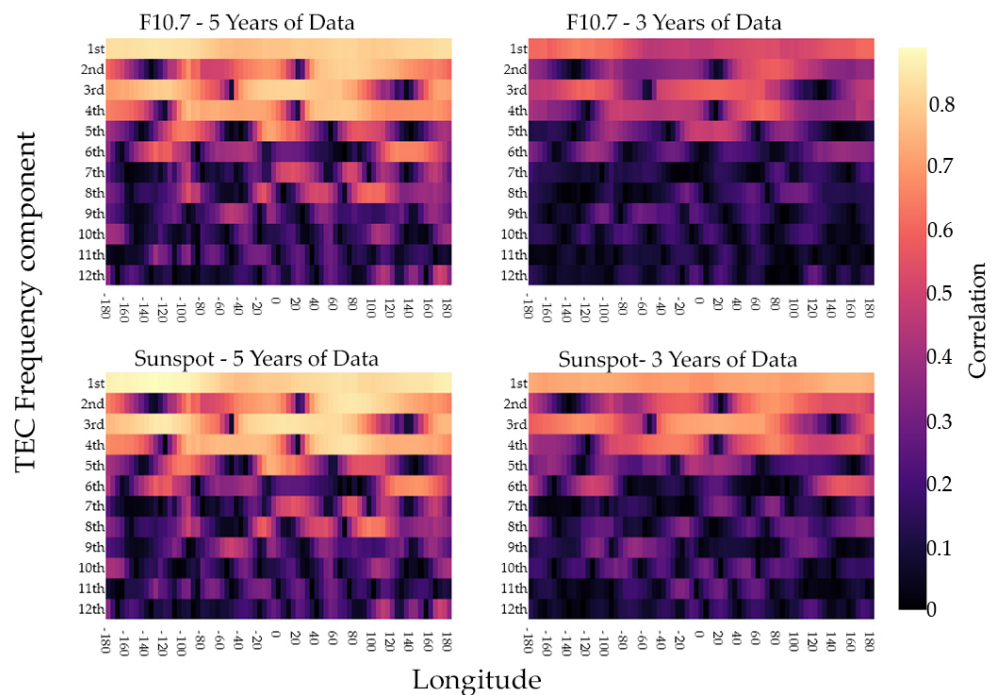
In this work the Elastic Net regression seems more appropriate for linear and polynomial regressions, once the features used show important correlation, as shown in Section 4.1. In such cases Elastic Net is preferred over  $L_1$  or  $L_2$  [8]. The values of  $r = 0.02$  and  $\alpha = 0.03$  were obtained by walk-forward validation and grid search.

#### 4. Model Analysis and Experiments

In machine learning, it is important to analyze the available data to better recognize a possible solution to the problem [21]. The features used in this work are a time series showing complex behavior. The choice of the most promising models could benefit from the evaluation of correlation between features and TEC frequency components, from the learning curves used during training and also from a qualitative comparison of reconstructed TEC maps.

##### 4.1. Features Correlation Analysis

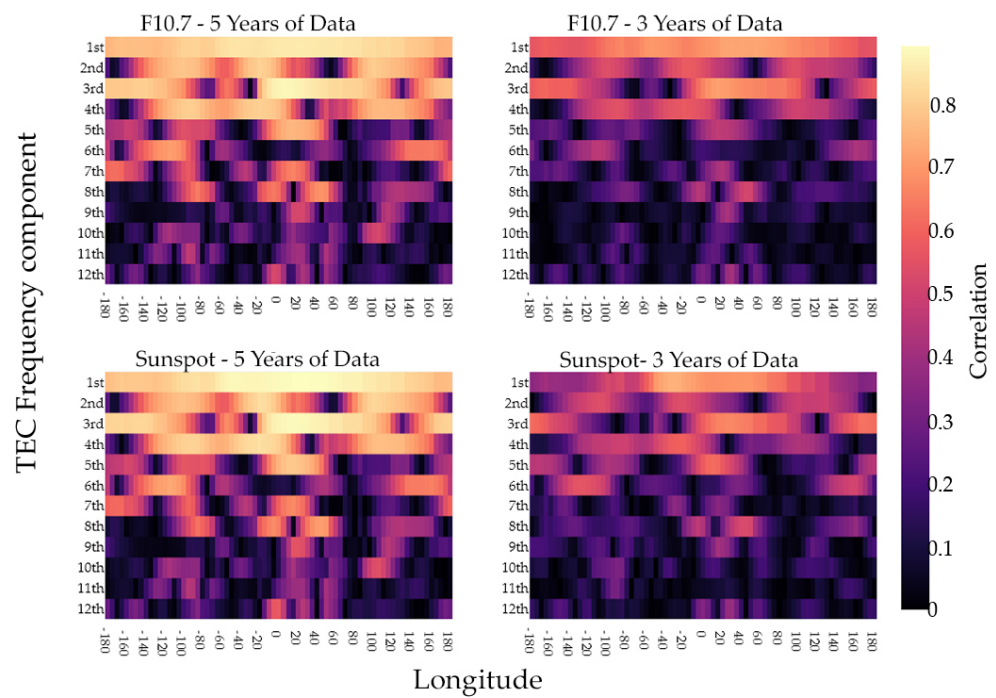
The TEC frequency components values were compared to the features F10.7 and sunspot number at two different geomagnetic latitudes:  $22.5^\circ$  N and  $22.5^\circ$  S. The Pearson correlation coefficient, combined for a period of 3 and 5 years of data, is shown in Figures 6 and 7 for each longitude. Both figures presented close results, which were similar to other latitudes. The amount of data needed in modeling is related to the complexity of the problem, and we can verify that the 5-year period data (2014–2018) seems more adequate for our analysis than the 3-year period (2016–2018). Also, the dependence of low-frequency components was substantial, which agrees with the prior analysis of Figures 3–5.



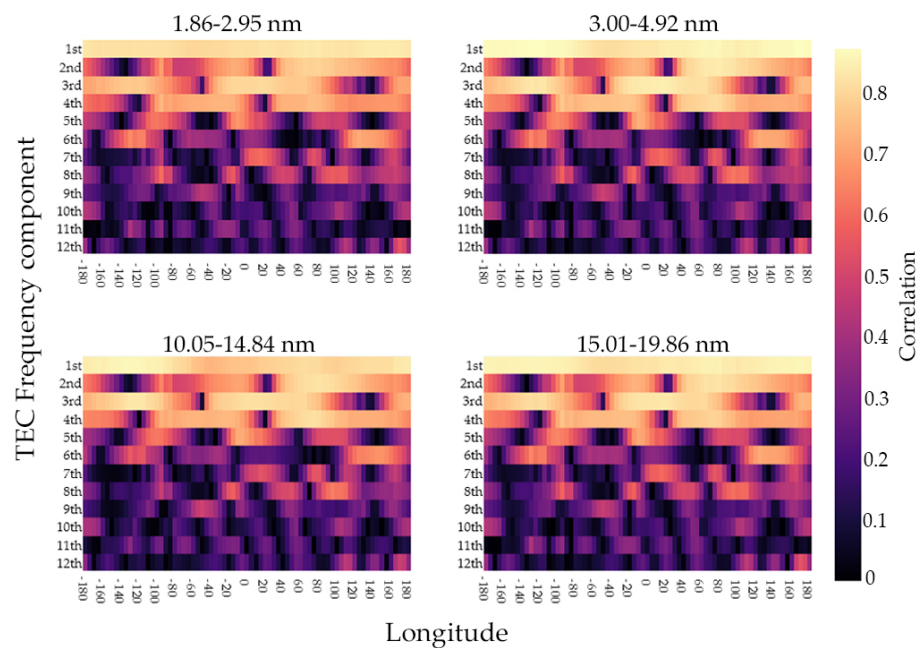
**Figure 6.** Correlation heatmap for F10.7 and sunspot at geomagnetic latitude  $22.5^\circ$  S. TEC frequency components range from 0 to 0.25 cycles-per-hour.

The evaluation of photon flux data is also shown in Figure 8 for four different bandwidths at latitude  $22.5^\circ$  S using 5 years of data. Although similar heat maps were obtained for different bandwidths in photon flux data, F10.7 and sunspot, which indicates a very strong correlation between them, we can see important differences when comparing three and five years of data. Therefore, it is expected that having five years of information available may result in better TEC modeling.





**Figure 7.** Correlation heatmap for F10.7 and sunspot at geomagnetic latitude  $22.5^\circ$  N. TEC frequency components range from 0 to 0.25 cycles-per-hour.



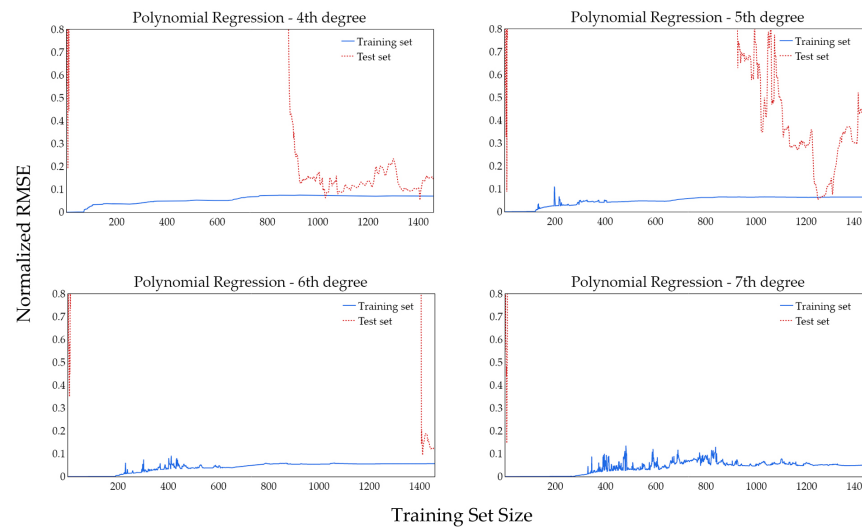
**Figure 8.** Correlation heat map of 5 years of data for photon flux at geomagnetic latitude  $22.5^\circ$  S. TEC frequency components range from 0 to 0.25 cycles-per-hour.

#### 4.2. Learning Curves

Learning curves are a graphical representation of the model performance as a function of the size of training set [8]. They can be evaluated to select the most promising models to be used and to reject models which present significant overfitting. In this work, the metric to evaluate the model performance was the root mean squared error (RMSE), largely adopted in regression models [18]. To generate the curves, normalized training data from 2014 to 2018 was used, and the candidate models were trained and evaluated multiple times, using different sizes of training subsets by the walk-forward method [21]. One year

of data (2019) was used as a testing set, and the scale of RMSE in learning curves was normalized.

Polynomial regression from 4th to 7th degrees presented short RMSE values for the training set. However, the correspondent learning curves for the test set revealed a clearly unstable model behaviour caused by overfitting, as shown in Figure 9 for the 1st DCT frequency component of the model at latitude  $22.5^{\circ}$  S and longitude  $180^{\circ}$  E.



**Figure 9.** Multiple degrees of polynomial regression learning curves for 1st DCT frequency component).

Considering the observed overfitting presented in high order polynomial regression models, our analysis limited polynomial regression up to the 3rd degree. The resultant learning curves for the 1st to 4th DCT frequency components at latitude  $22.5^{\circ}$  S and longitude  $180^{\circ}$  E can be seen in Figures 10–13, using the features F10.7, sunspot number and the two photon flux bandwidths. The strong correlation in photon flux at different bandwidths, previously observed in Figure 8, indicates reduction of new information when more bandwidths are added. As expected, the use of more than two photon flux bandwidths apparently did not lead to significant reduction in RMSE values, but substantially increased the required processing time. In general, these four models presented reasonably well-behaved learning curves and were considered in the following experiments, although the comparison of prediction model learning curves suggests polynomial regression of the 3rd degree is closer to overfitting than the others.

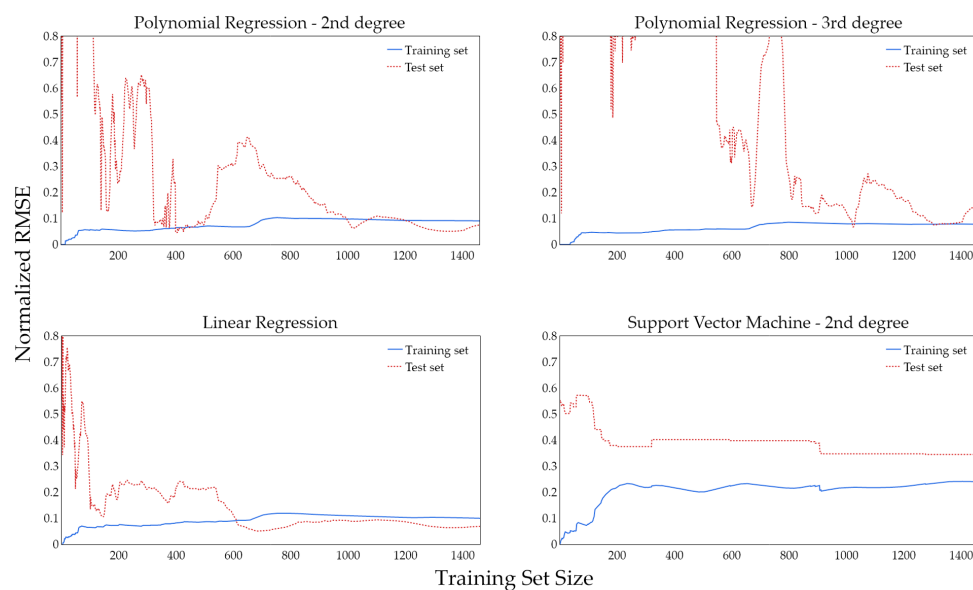


Figure 10. Learning curves of proposed models for 1st DCT frequency component (latitude: 22.5° S, longitude: 180° E).

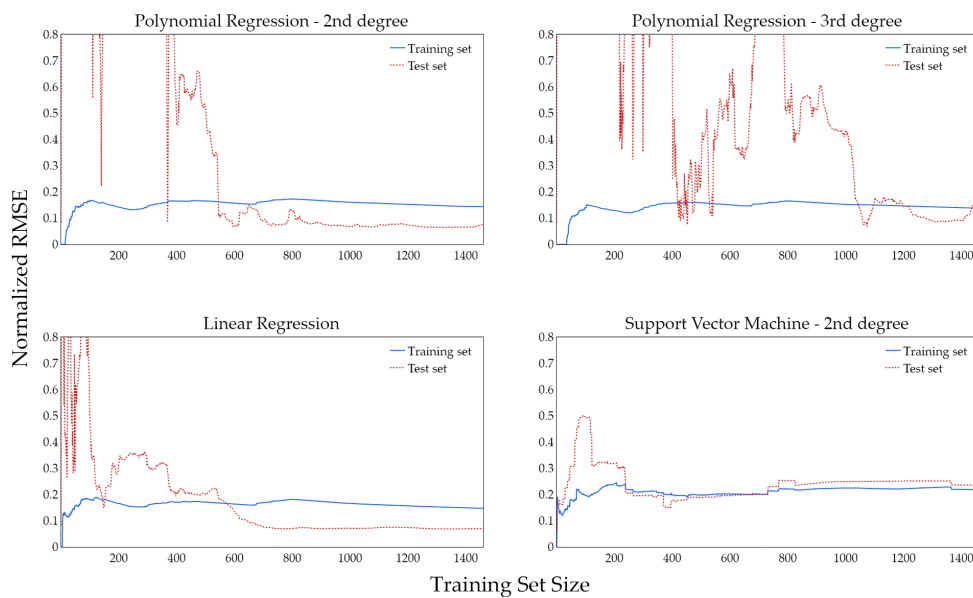
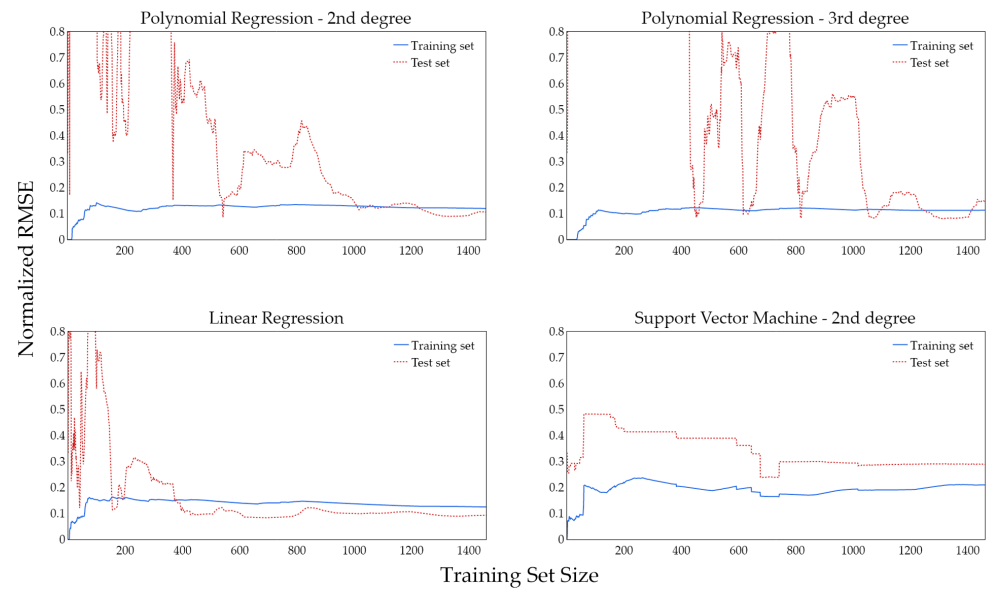
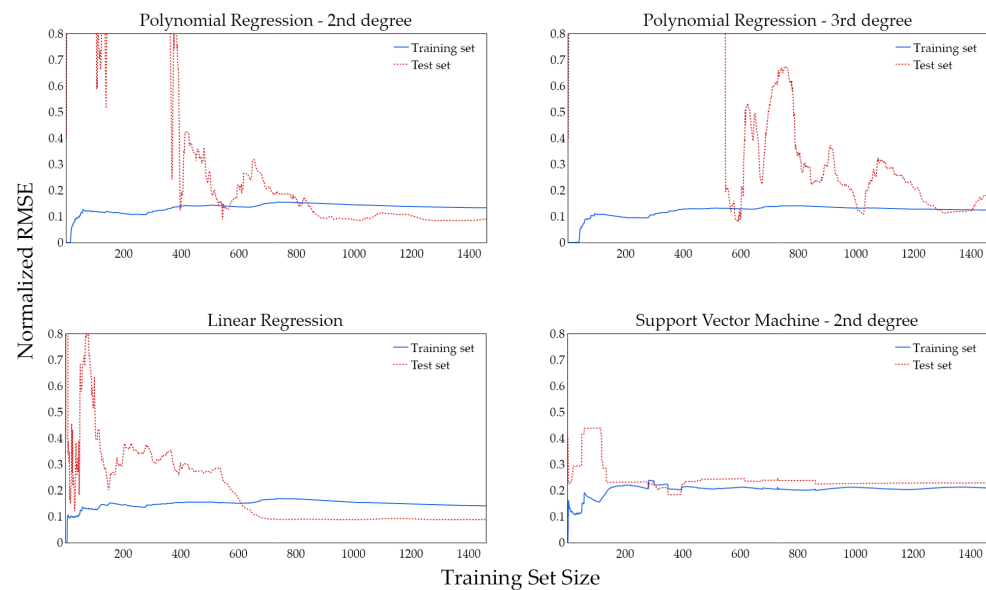


Figure 11. Learning curves of proposed models for 2nd DCT frequency component (latitude: 22.5° S, longitude: 180° E).



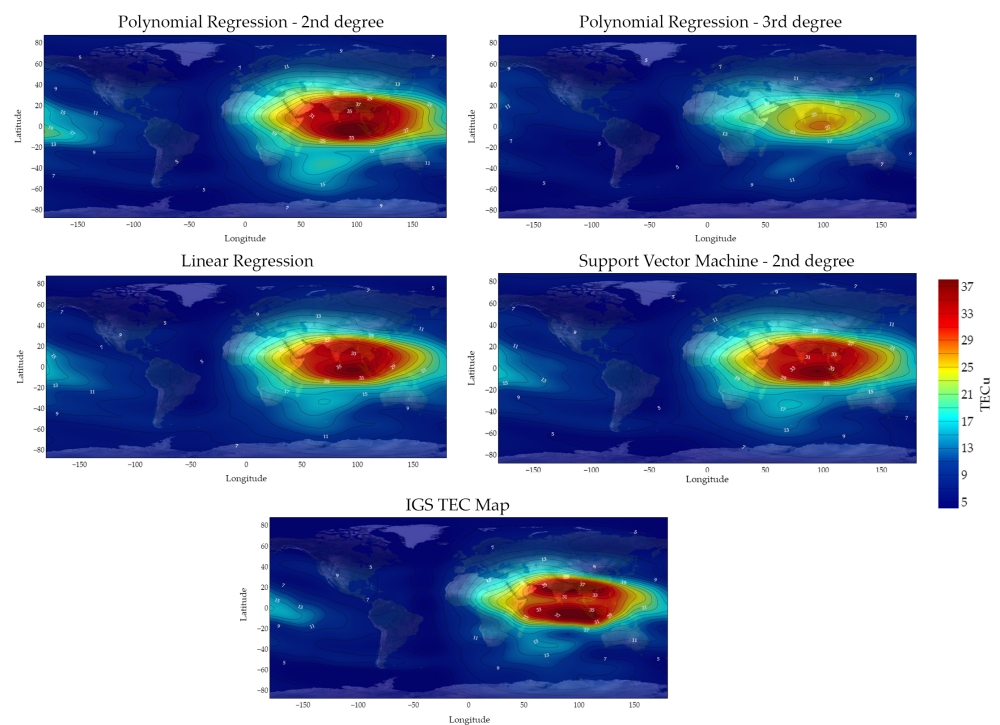
**Figure 12.** Learning curves of proposed models for 3rd DCT frequency component (latitude: 22.5° S, longitude: 180° E).



**Figure 13.** Learning curves of proposed models for 4th DCT frequency component (latitude: 22.5° S, longitude: 180° E).

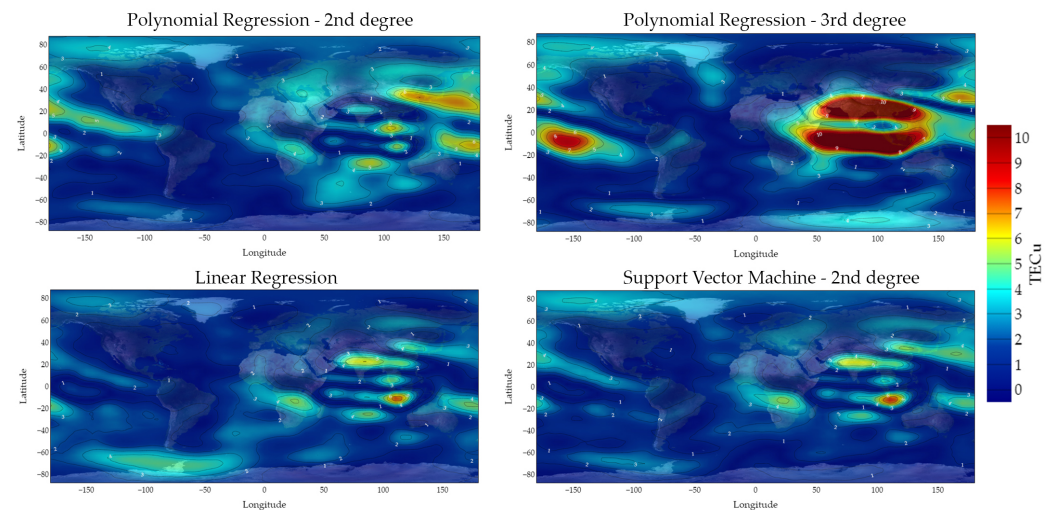
#### 4.3. Reconstruction of TEC Maps

An interesting qualitative analysis can be conducted by using the candidate models to reconstruct the TEC maps, and comparing them to the correspondent TEC map provided by IGS. Figure 14 shows reconstructed global TEC maps for 20 March, with the correspondent IGS TEC map. Clearly the 3rd degree polynomial regression model TEC reconstruction seems more divergent from IGS data than the others, which agrees with learning curve evaluation, presented in the previous section.



**Figure 14.** Candidate regression models TEC Map and correspondent IGS Map. Reconstruction made for 20 March 2019 at 08:00 UTC.

Figure 15 shows the absolute difference between each reconstructed TEC map shown in Figure 14 and the IGS TEC map. It can be noted that the results obtained with linear regression and SVM models are similar, with prediction errors geographically distributed at close locations.



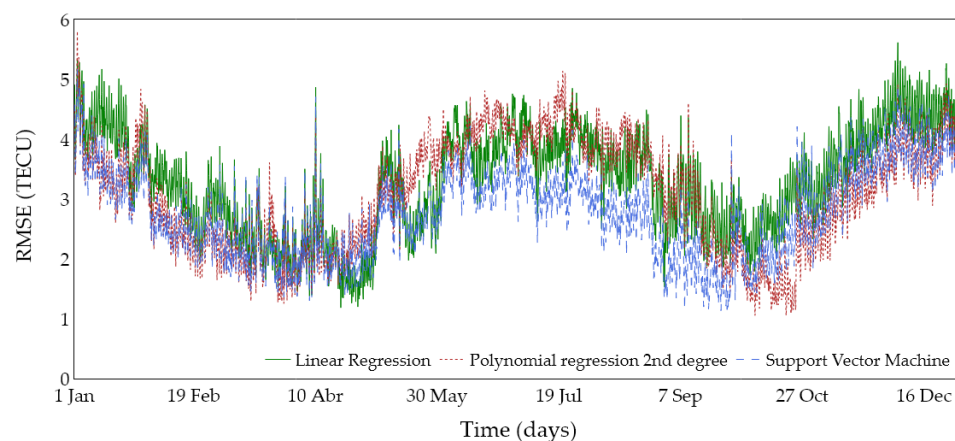
**Figure 15.** Absolute difference comparison between each reconstructed TEC map and IGS TEC map.

The qualitative evaluation, combined with learning curves results, indicates the 3rd degree polynomial regression model seems inadequate for our purposes. Significant errors in prediction were observed close to the ionization crests, symmetrical to the magnetic equator, a critical location for ionosphere modeling.

### 5. RMSE Results

The temporal variation in global RMSE, considering the linear, 2nd degree polynomial and SVM models, for the entire year of 2019 (test set) is shown in Figure 16. Although for

the whole period the variation in RMSE was similar for the three models, it can be noted that there is seasonal variation clearly related to the equinox (lower error) and solstice (higher error) periods. Considering that the models were defined using 5-year training data (not seasonally tuned), it would be expected that their lower errors would occur near the periods where the Earth is equally energized by the sun on both hemispheres—the equinoxes—around 20 March and 22 September. On the other hand, for the periods when the sun energy concentrates in one hemisphere—the solstices—around 21 June and 21 December, it seems the models could benefit from some sort of seasonal adjustment.



**Figure 16.** The temporal variation in global RMSE for 2019 (test set).

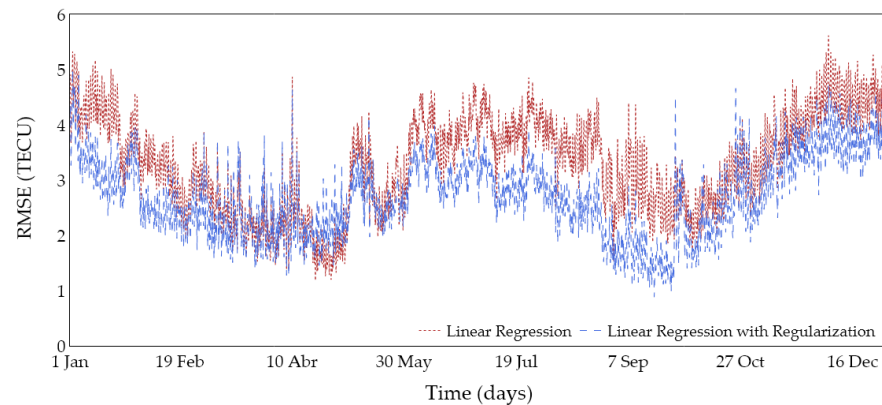
Table 1 refers to the global RMSE values for test set. The results shown are consistent with Figure 16, where it was observed that the errors are similar for the three models.

**Table 1.** Global RMSE for each model.

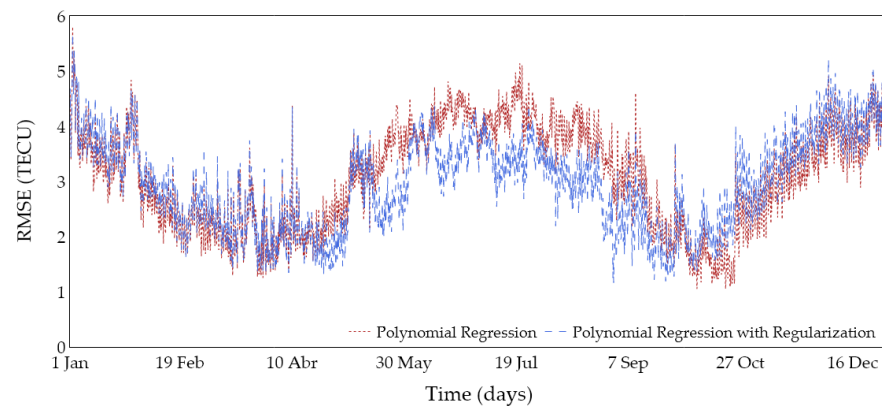
Model	Global RMSE [TECU]
Linear Regression	3.42
Polynomial Regression (2nd degree)	3.23
Support Vector Machine	2.89

### Regularization

Elastic Net regularization procedure, described in Section 3.4, was applied to linear and polynomial models. Error results are shown in Figures 17 and 18, respectively, and summarized in Table 2. Both Figures show error reduction with regularization, mainly for the linear regression model during most of the set duration. For the polynomial model, regularization performed better during periods close to the northern hemisphere summer solstice. However, for the opposite period of the year, regularization was ineffective or slightly worsened the results. It seems the effects of regularization are more noticeable in linear regression, possibly due to the simplicity of the model. Since the SVM model already uses a regularization parameter  $C$ , as discussed in Section 3.3, Elastic Net was not used in this model.



**Figure 17.** RMSE for Linear Regression.



**Figure 18.** RMSE for Polynomial Regression.

**Table 2.** Global RMSE for regularized models.

Model	Global RMSE (TECU)
Linear Regression with regularization	2.80
Polynomial Regression (2nd degree) with regularization	3.07

## 6. Conclusions and Future Work

In this work, we have examined TEC data in the frequency domain using DCT and its correlation with solar features, which was modeled using different approaches. Despite its simplicity, the linear regression model was able to achieve lower RMSE values when Elastic Net regularization was used. Temporal analysis of RMSE further revealed seasonal variation in errors that could be explored in future works. Different models could be achieved using seasonally divided training data. Those models outputs may be combined to obtain better results. Also, the introduction of knowledge about the physics phenomena in the model could lead to more accurate outputs. Further, deep learning could be explored, like basis functions, autoregressive models and recurrent neural networks to build different TEC time series forecasting models.

**Author Contributions:** Conceptualization, A.P.; methodology, A.P. and A.G.M.d.S.B.; software, A.G.M.d.S.B.; validation, A.P. and A.G.M.d.S.B.; formal analysis, A.P. and A.G.M.d.S.B.; investigation, A.P. and A.G.M.d.S.B.; resources, A.P. and A.G.M.d.S.B.; data curation, A.P. and A.G.M.d.S.B.; writing—original draft preparation, A.P. and A.G.M.d.S.B.; writing—review and editing, A.P. and A.G.M.d.S.B.; visualization, A.P. and A.G.M.d.S.B.; supervision, A.P.; project administration, A.P.; funding acquisition, A.P. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul—FAPERGS grant number 20/2551-0000322-1.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors acknowledge the support of Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul—FAPERGS, for research sponsorship.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

TEC	Total Electron Content
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
IGS	International GNSS Service
UTC	Coordinated Universal Time
DCT	Discret Cosine Transform
CME	Coronal Mass Ejection
RMSE	Root Mean Squared Error
NASA	National Aeronautics and Space Administration
SVM	Support Vector Machine
LASSO	Least Absolute Shrinkage and Selection Operator
MSE	Mean Squared Error

## References

- Moldwin, M. *An Introduction to Space Weather*, 3rd ed.; Cambridge University Press: New York, NY, USA, 2008; p. 156.
- GNSS Atmospheric Products—Ionosphere. Available online: <https://cddis.nasa.gov/archive/gnss/products/ionex/> (accessed on 21 August 2020).
- Klobuchar, J.A. Ionospheric Time-Delay Algorithm for Single-Frequency GPS Users. *IEEE Trans. Aerosp. Electron. Syst.* **1987**, *AES-23*, 325–331. [[CrossRef](#)]
- Bilitza, D.; Altadill, D.; Reinisch, B.; Galkin, I.; Shubin, V.; Truhlik, V. The International Reference Ionosphere: Model Update 2016. In Proceedings of the EGU General Assembly 2016, Vienna, Austria, 17–22 April 2016.
- Nava, B.; Coïsson, P.; Radicella, S.M. A new version of the NeQuick ionosphere electron density model. *J. Atmos. Sol. Terr. Phys.* **2008**, *70*, 1856–1862. [[CrossRef](#)]
- Jakowski, N.; Hoque, M.M.; Mayer, C. A new global TEC model for estimating transionospheric radio wave propagation errors. *J. Geod.* **2011**, *85*, 965–974. [[CrossRef](#)]
- Feng, J.; Han, B.; Zhao, Z.; Wang, Z. A New Global Total Electron Content Empirical Model. *Remote Sens.* **2019**, *11*, 706. [[CrossRef](#)]
- Géron, A. *Hands-on Machine Learning with Scikit-Learn and Tensorflow*, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2017; p. 549.
- Geryl, P.; Alvestad, J. *A Formula for the Start of a New Sunspot Cycle*; Springer: New York, NY, USA, 2020.
- Lin, C.H.; Liu, J.Y.; Fang, T.W.; Chang, P.Y.; Tsai, H.F.; Chen, C.H.; Hsiao, C.C. Motions of the equatorial ionization anomaly crests imaged by FORMOSAT-3/COSMIC. *Geophys. Res. Lett.* **2007**, *34*. [[CrossRef](#)]
- Xuguang, C.; Burns, G.A.; Wenbin, W.; Daniell, E.; Martinis, C.R.; McClintock, W.; Inez, S. Batista Observation of postsunset OI 135.6 nm radiance enhancement over South America by the GOLD mission. *J. Geophys. Res. Space Phys.* **2020**, *126*. doi:10.1029/2020JA028108. [[CrossRef](#)]
- Gopalswamy, N. History and development of coronal mass ejections as a key player in solar terrestrial relationship. *Geosci. Lett.* **2016**, *3*, 8. [[CrossRef](#)]
- Chou, M.-Y.; Pedatella, N.M.; Wu, Q.; Huba, J.D.; Charles, C.H.L.; Schreiner, W.S.; Braun, J.J.; Eastes, R.; Yue, J. Observation and Simulation of the Development of Equatorial Plasma Bubbles: Post-Sunset Rise or Upwelling Growth? *JGR Space Phys.* **2020**, *125*, e2020JA028544.
- Karan, D.K.; Daniell, R.E.; England, S.L.; Martinis, C.R.; Richard, W.E.; Burns, G.A.; McClintock, W.E. First zonal drift velocity measurement of equatorial plasma bubbles (EPBs) from a geostationary orbit using GOLD data. *J. Geophys. Res. Space Phys.* **2020**, *125*, e2020JA028173. [[CrossRef](#)]
- Choudhuri, A.R. *Nature's Third Cycle*, 3rd ed.; Oxford University Press: New York, NY, USA, 2015; p. 286.



16. Tobiska, W.K.; Woods, T.; Eparvier, F.; Viereck, R.; Floyd, L.; Bouwer, D.; Rottman, G.; White, O.R. The SOLAR2000 empirical solar irradiance model and forecast tool. *J. Atmos. Sol. Terr. Phys.* **2000**, *62*, 1233–1250. [[CrossRef](#)]
17. Rao, K.R.; Yip, P. *Discrete Cosine Transform, Algorithms, Advantages, Applications*, 1st ed.; Academic Press: Cambridge, MA, USA, 1990; p. 512.
18. Hastie, T.; Tibshirani, R.; Friedman, J. *The Element of Statistical Learning, Data Mining, Inference and Prediction*, 2nd ed.; Springer Science+Business Media: New York, NY, USA, 2017; p. 446.
19. Wasserman, L. *All of Statistics*, 1st ed.; Springer Science+Business Media: New York, NY, USA, 2004; p. 446.
20. Hearst, A. Support vector machines. *IEEE Intell. Syst.* **1998**, *13*, 18–28. [[CrossRef](#)]
21. Murphy, K. *Machine Learning, a Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
22. Hastie, T.; Tibshirani, R.; Friedman, J. *Statistics for High-Dimensional Data*, 1st ed.; Springer Science+Business Media: New York, NY, USA, 2011; p. 575.
23. Bogdan, M.; van den Berg, E.; Su, W.; Candès, E.J. *Statistical Estimation and Testing via the Sorted L1 Norm*; Departments of Mathematics and Computer Science, Wrocław University of Technology and Jan Długosz University: Częstochowa, Poland, 2013.
24. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [[CrossRef](#)]