

Accepted Manuscript

This is a PDF file of a manuscript that has been peer reviewed and accepted for publication. Please note that the article has neither been copy edited nor formatted in the Geofizika style yet, and, that it does not have all bibliographic details.

DOI: 10.15233/gfz.2017.34.12

Original scientific paper

UDC

Modelling extreme values of the total electron content: Case study of Serbia

Miljana Todorović Drakul¹, Mileva Samardžić Petrović², Sanja Grekulović¹,
Oleg Odalović¹ and Dragan Blagojević¹

¹ Department of Geodesy and Geoinformatics, Faculty of Civil Engineering,
University of Belgrade, Belgrade, Serbia

² Faculty of Mining and Geology, University of Belgrade, Belgrade, Serbia

Received 25 January 2017, in final form 21 September 2017

This paper is dedicated to modeling extreme *TEC* values at the territory of Serbia. For the extreme *TEC* values, we consider the maximum values from the peak of the 11-year cycle of solar activity in the years 2013, 2014 and 2015 for the days of the winter and summer solstice and autumnal and vernal equinox. The average *TEC* (Total Electron Content) values between 10 and 12 UT (Universal Time) were treated. As the basic data for all processing, we used GNSS (Global Navigation Satellite System) observation obtained by three permanent stations located in the territory of Serbia. Those data, we accept as actual, *i.e.* as a “true *TEC* values”. The main objectives of this research were to examine the possibility to use two machine learning techniques: neural networks and support vector machine. In order to emphasize the quality of applied techniques, all results are adequately compared to the *TEC* values obtained by using International Reference Ionosphere global model. In addition, we separately analyzed the quality of techniques throughout temporal and spatial-temporal approach.

Keywords: TEC, machine learning, ionosphere, GNSS, IRI

1. Introduction

Global Navigation Satellite System (GNSS) techniques have been significantly improved over the past two decades, however, several sources of errors still remain, which may limit the accuracy, practical operation and performances of precise positioning. The ionosphere is the major source of errors in the GNSS positioning (Hofmann-Wellenhof et al., 1992). In this region, ionizing radiation from the Sun causes the existence of electrons, in the quantities that influence radio-waves propagation (Kleusberg and Teunissen, 1996). The number of electrons intercepted by the electro-magnetic waves traveling through ionosphere is known as the Total Electron Content - TEC. It represents an integral of electron density per unit of

volume, along the signal path between the satellite and the GNSS receiver. It is noted in TECU units, with 1 TECU being 10^{16} electrons per square meter of cylindrical cross-section.

The ionosphere is a very dynamic environment, and the electron density may significantly vary per minute at the given location, which leads to temporal and spatial variations in the Total Electron Content. The most significant changes occur due to the geomagnetic activity and the Earth's revolution around the Sun in the period of the equinox and solstice (Wautelet and Warnant, 2014). Indicators of these changes are primarily Solar flux (SF), Sunspot number (SSN), Index of geomagnetic activity (A_p index). Furthermore, the largest signal delay caused by the ionospheric influence occurs in the 10–14h local time interval (Radzi et al., 2013). The testing is conducted for the years 2013, 2014 and 2015 that are at the peak of the 11-year cycle of solar activity. In order to model and predict previously mentioned extreme TEC values, ML techniques were used.

Machine learning (ML) techniques are empiric modeling approaches that have the capability to extract information and reveal patterns by exploring unknown relations between input and output variables (dependent and independent continuous and categorical variables). ML techniques itself include a great number of methods with different kinds of learning algorithms. However, two ML techniques most commonly used in order to model/predict TEC values are Neural Networks (NN) and Support Vector Machines (SVM) (Barrile et al., 2006; Huang et al., 2014; Habarulema and McKinnell, 2012; Okoh et al., 2016, Razin and Voosoghi, 2016). Therefore, the main objective of this research study is to examine the capability of those ML techniques to model and predict extreme TEC values. For that purpose, we separately examined and analyzed the capability of SVM and NN to model both temporal and spatial – temporal variation of TEC values.

In addition, we calculated the examined extreme TEC values using a global ionospheric model International Reference Ionosphere (IRI). It is the international standard empirical model for the terrestrial ionosphere since 1999. The IRI model is one of the most frequently utilized models used for comparison of many different techniques and methods utilized for modeling TEC values, such as: Klobuchar (Klobuchar, 1986; Swamy et al., 2013), NN technique (Sur and Paul, 2013), Holt-Winter method (Elmunim et al., 2016). Therefore, we compared the differences between TEC values based on GNSS and the IRI model and differences between TEC values based on GNSS and model outcomes obtained by ML techniques.

2. Methods

2.1. TEC based on GNSS observation

Ionosphere delay is nearly proportional to the Total Electron Content along the signal path and inversely proportional to the frequency squared. This dispersion property of ionosphere provides for dual frequency GNSS receivers to compensate for the errors of ionosphere delay and measure the TEC.

To compensate ionospheric delay, dual-frequency GNSS receivers use f_1 (1,575.42 MHz) and f_2 (1,227.60 MHz) frequencies. Delay, $\Delta t = t_2 - t_1$ measurement between f_1 and f_2 frequencies is used to calculate TEC along the signal path:

$$\Delta t = \left(\frac{40.3}{c} \right) \cdot \frac{\text{TEC}}{\left(\frac{1}{f_2^2} - \frac{1}{f_1^2} \right)} \quad (1)$$

where c is the speed of light in open space. Considering that using pseudo-ranges provide absolute TEC while using phase differences improves the accuracy, GNSS data provides for the efficient method of estimating TEC values with greater spatial and temporal coverage (Davies and Hartmann, 1997; Igarashi et al., 2001). Having that the frequencies used by the GNSS system are sufficiently high, the signals are minimally influenced by ionospheric absorption and Earth's magnetic field, both in the short and in the long-term changes in the ionosphere structure. Here, the values of slanted TEC were obtained as the sum of slanted TECs, hardware satellite delay b_S and hardware receiver delay b_R . Thus, vertical TEC ($VTEC$) may be expressed as follows:

$$VTEC = \frac{STEC + b_S + b_R}{S(e)} \quad (2)$$

where $STEC$ is slanted TEC, e is the elevation angle of satellites in degree, $S(e)$ is the slant factor against the zenith angle z at the Ionospheric Pierce Point (IPP) and $VTEC$ is vertical TEC in the IPP point (Fig. 1). The slant factor, $S(e)$ (or the mapping function) is defined as (Langley et al., 2002):

$$S(e) = \frac{1}{\cos(z)} = \left(1 - \frac{R_e \times \cos(e)}{R_e + h_i} \right)^{-0.5} \quad (3)$$

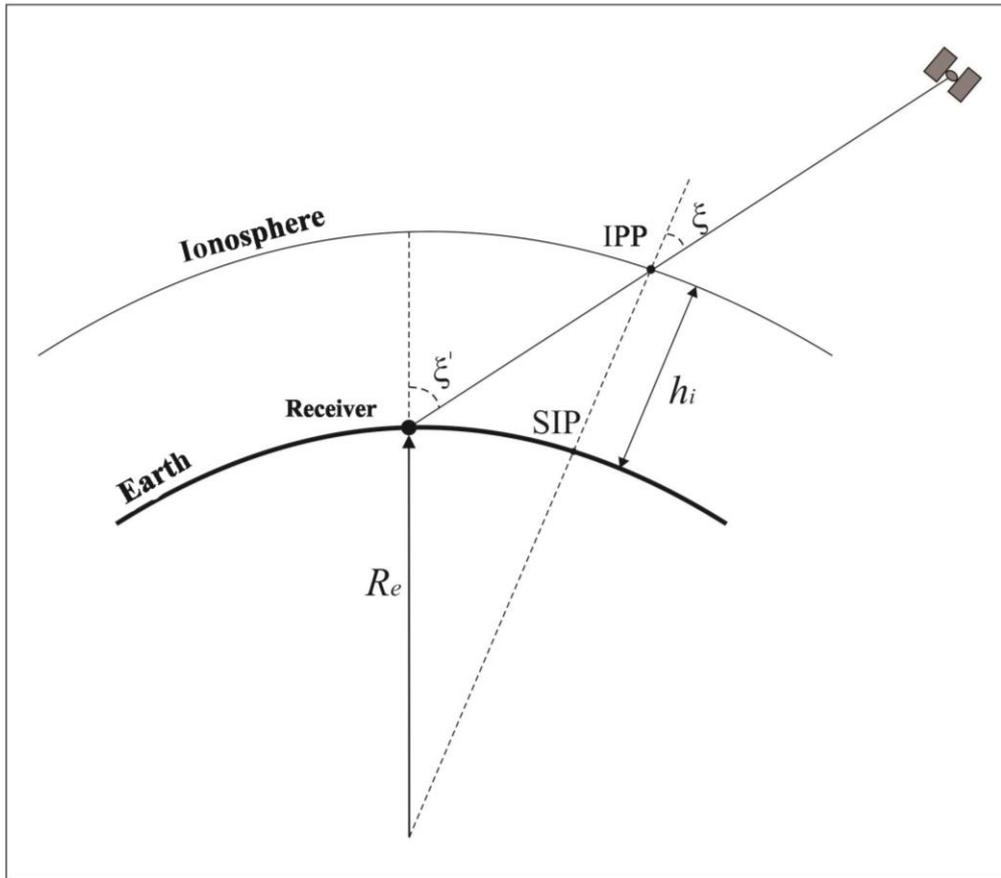


Figure 1. Trigonometric single-layer mapping function (Davies and Hartmann, 1997).

where R_e is the average Earth's radius in km, and h_i is the (effective) height of ionosphere over the Earth's surface.

2.2. TEC based on ML techniques

In this research, temporal and spatial temporal predictions of TEC are derived using two ML techniques, SVM and NN. The prediction of both TEC models is engaged in the form of a supervised learning regression problem, where it is required to determine unknown functions that map attributes, such as Solar flux, Sunspot number, Index of geomagnetic activity or geographic coordinates, to particular TEC values based on the past data for the study area in the case of temporal prediction or based on the data from some other considering GNSS stations in the case of spatial temporal prediction of TEC. For that purpose, we used well known SVM and Radial Basis Function (RBF) (Abe, 2010) as kernel function, and for NN we used Multi-layer Perceptron (MLP) (Rumelhart et al., 1986), with softplus activation function, (Dugas et al., 2001).

The efficient application of most ML techniques including SVM and NN techniques requires selection of an optimal combination of function parameters. Therefore, in order to use these algorithms appropriately it was necessary to find the optimal combination of two parameters; γ of the RBF and penalty C for TEC SVM based models and the number of neurons in a hidden layer for TEC NN based models.

Furthermore, the standard procedure for ML base modelling require the adequate attribute selection (Kim et al., 2003). The main goal of attribute selection is to choose a subset of informative attributes by eliminating those with little or no information relevant to target values, which is in the case of this research attribute related to the TEC. For that purpose, we used Correlation-based Feature Subset (CFS) (Hall, 1999) selection method. The CFS method automatically determines a subset of m relevant attributes ($m < n$, n is the number of all considering attributes), *i.e.* attribute that are highly correlated with the TEC but uncorrelated with each other.

2.3. TEC based on global ionospheric model

Because of the complicated nature of the ionosphere, there have been numerous approaches to model the ionosphere. One of the models that are used to obtain TEC values is IRI model. It is a global standard empirical model of the ionosphere, which for a given location, time and date describes density, electron temperature and content, as well as the temperature and composition of the ions at an altitude of about 50 km to 2 000 km.

IRI is a project under the support of the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) (Bilitza, 2001). Since 1978 (Rawer et al., 1978) when the model was implemented, there were few advanced versions. There is a huge number of reports where it is possible to find a detailed description of the model (Rawer et al., 1978; Rawer et al., 1981; Bilitza et

al., 1990). Those reports contain the databases, algorithms, equations and applications. Their Fortran source codes are available on the IRI website (<http://irimodel.org/>), together with a description of the model. IRI consists of a global model for D, E, F1 and F2 region of the ionosphere. A complete profile of the electron density is obtained by using mathematical functions and merging algorithm for different regions.

3. Data

Data from three base stations have been taken over, in the form of 30-second RINEX (Receiver Independent Exchange Format) files. Examined stations Kanjiža (S_K), Novi Pazar (S_{NP}) and Šabac (S_S), belong to the permanent GNSS network of the Republic of Serbia under the name AGROS (Active Geodetic Reference Base of Serbia) (Fig. 2).

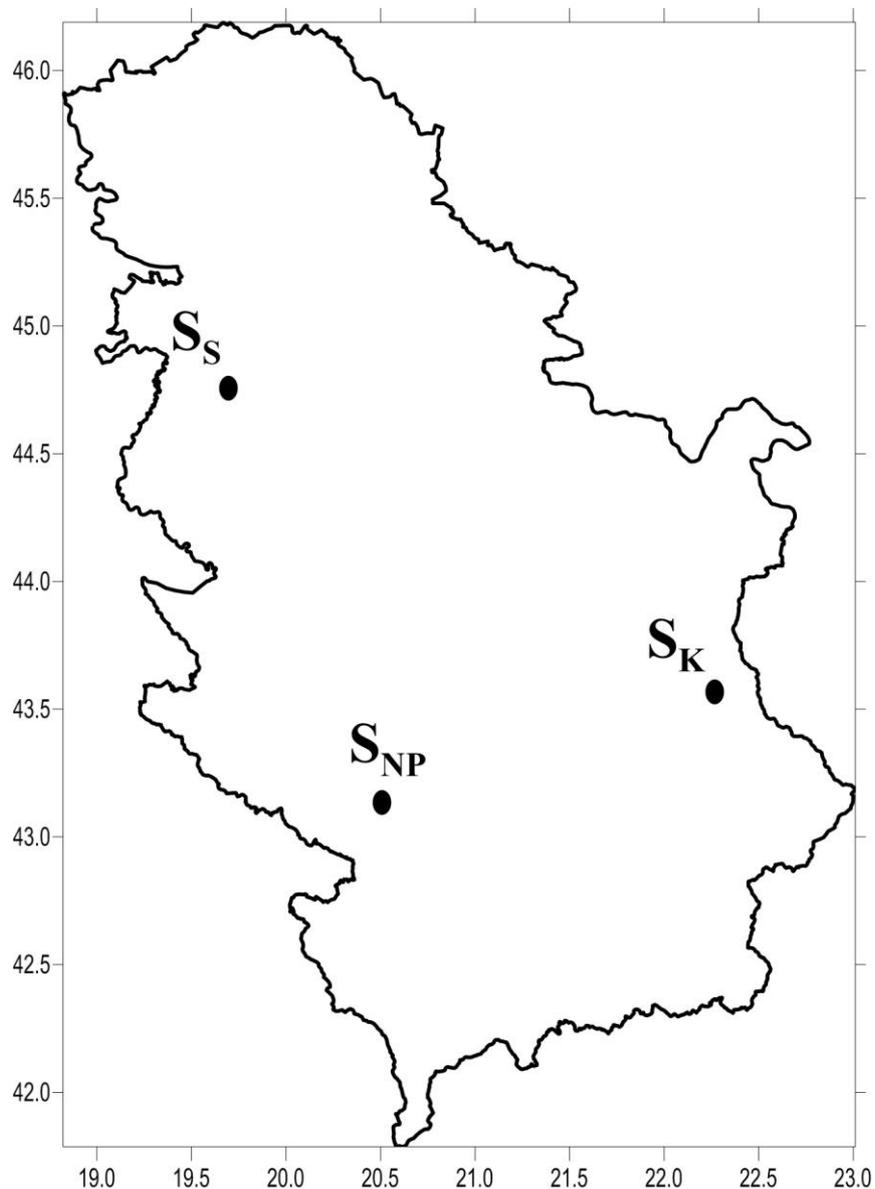


Figure 2. Distribution of AGROS stations, S_K , S_S and S_{NP} , within the territory of Serbia.

Collected data contain five-day observations from March, June, September and December of 2013, 2014 and 2015. The temporal series of TEC measurements have been obtained using equation (2). GNSS TEC Analysis software, developed at Boston University, was used for processing (Seemala, 2014). Phase and code values on both frequencies have been used to eliminate clock and tropospheric effect errors, in order to calculate relative values of slanted TEC (Sardón and Zarraoa, 1997). Afterwards, absolute TEC values have been obtained by removing hardware delays, *i.e.* differential code discrepancies between the satellite (produced by the Data Centre of the Bern University, Switzerland) and the receiver (obtained by minimizing the TEC value between 2:00 AM and 6:00 AM - local time) (Seemala and Valladares, 2011). Trigonometric single-layer mapping function (3) was used to convert TEC to VTEC at AGROS stations and at the IPP point at the altitude of 350 km. Elevation angle was limited to the value of 20 degrees to decrease a potential effect of multiple signal reflection during the tests. Data sampling was done for every 30 seconds.

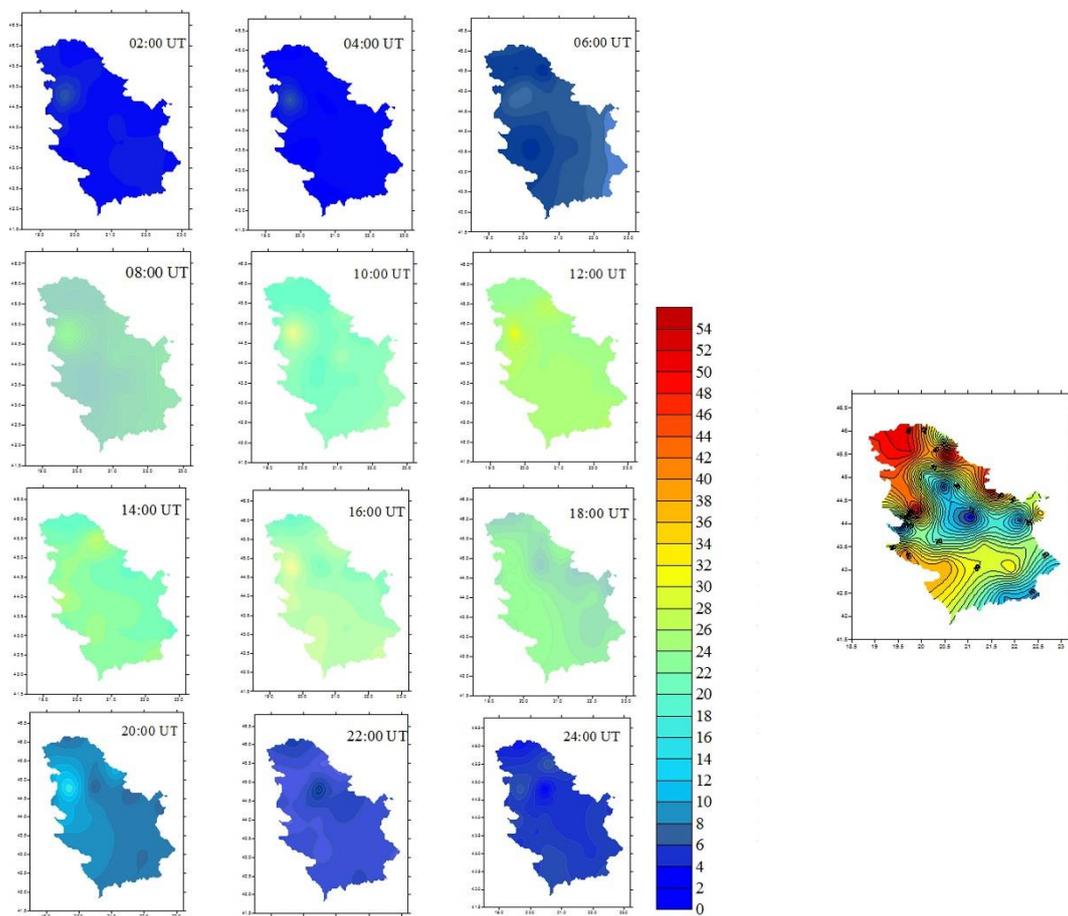


Figure 3. VTEC value over the Serbia for day 22 September 2014.

Changes in TEC were investigated for all the days of interest during 2013, 2014 and 2015. As it is evident on Fig. 3., the TEC reaches the maximum value in the period from 10 to 14 hours

Given that the ionosphere influence on GNSS observations is largest in the period from 10 to 14 h UT, the values of TEC are calculated and averaged for this time interval. Figure 4 shows TEC values for all three stations in the studied time interval. It can be concluded that TEC values for the same time intervals and for all stations vary in the 1 to 7 TECU range. Changes in TEC values during different seasons are significant and vary from 10 to 55 TECU, where the lowest TEC variations are recorded in June and the highest in the March period for all considered years.

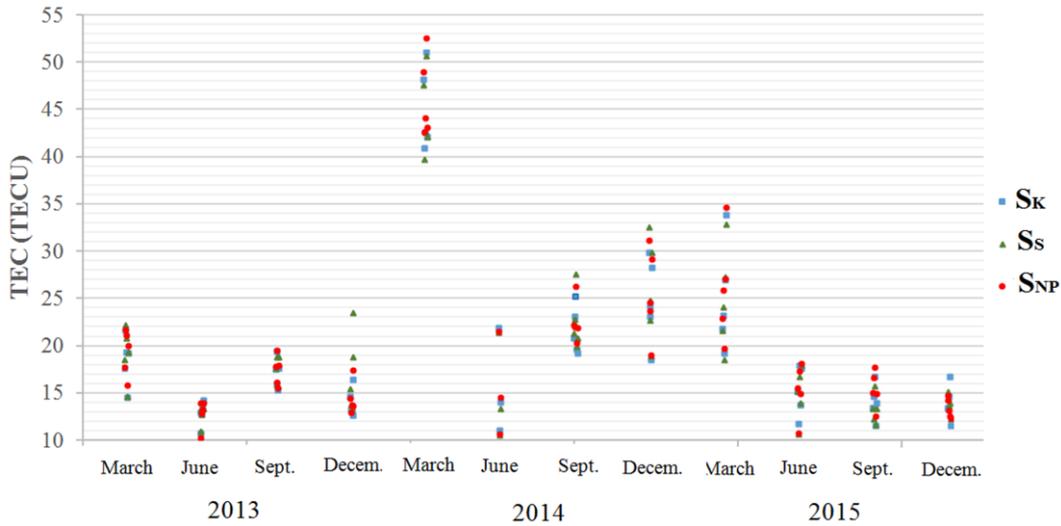


Figure 4. Distribution of TEC values for examined time intervals based on observation from three stations, S_K , S_S and S_{NP} .

Given that the SF, SSN and Ap index are factors that influence ionospheric variations, they are included into the model as attributes. Their values were downloaded from NASA's Space Physics Data Facility (<http://omniweb.gsfc.nasa.gov/form/dx1.html>), for all 12 periods of interest (Fig. 5).

Beside TEC values obtained from GNSS and from ML techniques (SF, SSN and Ap index), TEC values for all examined time points and for all three stations obtained from the global IRI model were also used for this research. All necessary data for the global TEC calculation were provided by the official website of the global model IRI 2012: http://omniweb.gsfc.nasa.gov/vitmo/iri2012_vitmo.html.

4. Results and discussion

Two different research approaches are presented. The first one represents temporal prediction of the TEC values (Fig. 6a.) and the second one spatial - temporal prediction of the TEC (Fig. 6b.). The model for temporal prediction was created based on a data set of all three stations and time points from years 2013 and 2014 and validated based on a data set of all three stations and time points from the year 2015. The model for spatial - temporal prediction was created based on data from stations S_K , S_S for all investigated time points and validated on data from station S_{NP} .

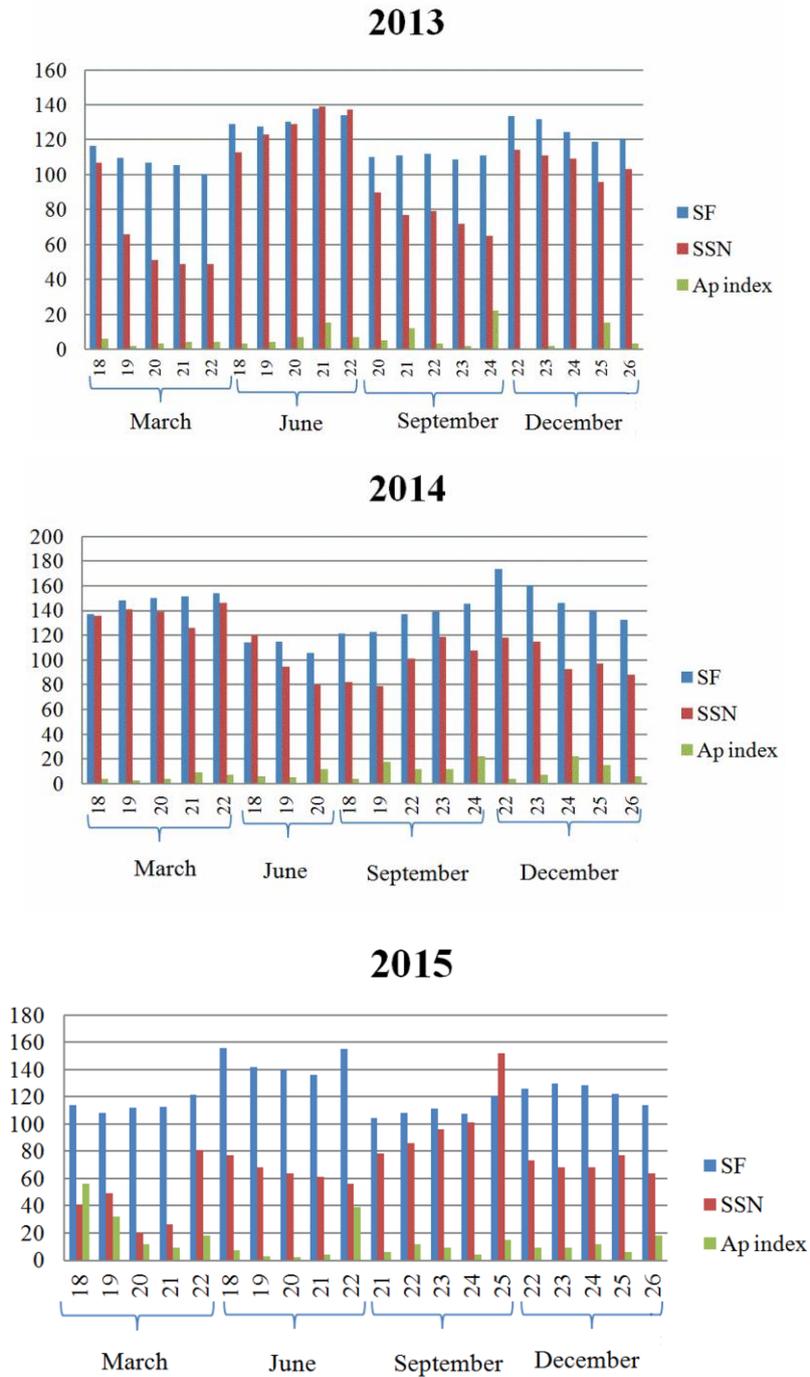


Figure 5. Distribution of SF, SSN and *Ap* index values for examined time intervals at year 2013, 2014, and 2015.

In both mentioned approaches, we used Weka software (Hall et al., 2009) for implementation of CFS and ML techniques. The SOMreg algorithm was used to implement the SVM for regression algorithm and MLPreg algorithm to implement Neural Network.

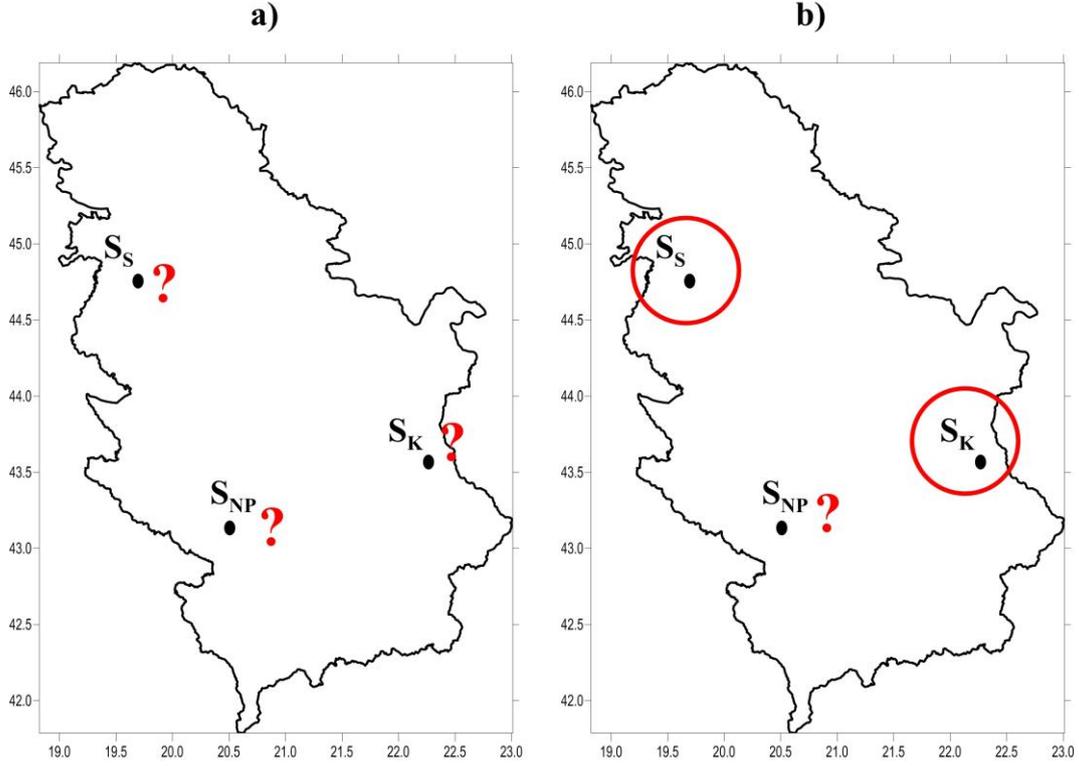


Figure 6. a) Temporal prediction of the TEC and b) Spatial – temporal prediction of the TEC.

4.1. Temporal prediction of TEC

In order to create and validate temporal ML TEC models, two independent data sets were created based on the values of TEC, Solar flux (SF), Sunspot number (SSN), Index of geomagnetic activity (A_p index), and geographic coordinates (latitude-lat, longitude-long, height-h). Furthermore, in order to indicate the time intervals in which TEC value was obtained in regards to winter and summer solstice and autumnal and vernal equinox, the additional attribute was defined, labeled as Month. Each data set (labeled with t , where t can be equal 1 or 2) contains TEC values for all three stations (S_K , S_{NP} and S_S), represented as a row vector of attributes a_{it} , $i = 1, \dots, 7$, including TEC values as the target attribute. The training data set, labeled as TR_1 , contained the time points from years 2013 and 2014 and the test data set, label TE_1 , contained the time points from year 2015, where the actual TEC (TEC based on GNSS observation – TEC_{GNSS}) at year 2015 was used for comparison with the predicted TEC values in the year 2015.

After the creation of data sets, attribute selection was performed on the training data set TR_1 . The CFS automatically determines a subset of four relevant attributes, which are highly correlated with the TEC values and uncorrelated with each other: Month, SF, lat and long. All TEC models were therefore built and validated on training and test data sets that contain only selected attributes, labelled as TE_1^{CFS} and TR_1^{CFS} , respectively. Using training data sets and 10 fold cross-validation, the optimal combination of parameters was found for both ML techniques.

Minimum, maximum, average, standard deviation (*St.Dev.*), root mean square error (*RMSE*) of differences of modeled and actual TEC values, as well as R

squared metric (R^2) are used as models quality control measures. The quality control measures for the best performing models based on SVM and NN and on global model IRI, are present in Tab. 1. Histograms of the differences for those models are presented in Fig. 7. The data in Fig. 7. are labelled as follows:

- $\Delta SVM = TEC_{SVM} - TEC_{GNSS}$,

as differences between TEC obtained by TE_{1CFS}/SVM model and actual TEC values,

- $\Delta NN = TEC_{NN} - TEC_{GNSS}$,

as differences between TEC obtained by TE_{1CFS}/NN model and actual TEC values,

- $\Delta IRI = TEC_{IRI} - TEC_{GNSS}$,

as differences between TEC obtained by global IRI model and actual TEC values.

Considering the range of actual TEC, from 10 to 53 TECU, given the relatively small time interval and small samples, and considering that the objective was to predict the extreme values of TEC, the obtained results indicate that in this research main goal is achieved (Tab. 1). The both techniques are more capable to “learn” and to model the extreme TEC values in the local area (Serbia), comparing, for example, to the global IRI model. The obtained standard deviation and RMSE are equal or less than 4.1 TECU and mean values of differences between actual and modelled TEC values are 1.2 TECU (with a range from -7.7 TECU to 9 TECU) and 0.0 TECU (with a range from -10.5 TECU to 8.3 TECU), for temporal TEC SVM and NN based model, respectively. The same conclusion can be given according to R^2 values where TEC SVM and TEC NN are almost the same (0.94 and 0.95) while R^2 obtained for the global model reach values of 0.86.

When analyzing the distribution of actual and temporal modelled TEC differences (Fig. 7), it can be noted that the results obtained from temporal TEC NN based model are closest to the normal distribution, which indicates that NN have slightly better performance in comparison with SVM. For differences, values of quality control measures gained by applying the IRI model are definitely biased. All differences have a negative sign with a minimum value of -18.6 TECU, maximum of 0 TECU and with an average of -5.0 TECU. Their standard deviation is almost the same as the standard deviation of two other mentioned sets, and that we can say the same for their range. On the other side, their $RMSE$ is about 36% greater comparing to the $RMSE$ of the two other sets. The mentioned characteristics of quality control measures are obvious for the histograms which

Table 1. Best performing temporal models by using TE_{1CFS} data set.

Data sets / ML techniques	Quality controls measure					
	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>RMSE</i>	<i>R²</i>
TE_{1CFS}/SVM	-7.7	9.0	1.2	3.8	4.0	0.94
TE_{1CFS}/NN	-10.5	8.3	0.0	4.1	4.1	0.95
Data sets / Model						
$TE_{1CFS}/IRI2012$	-18.6	-0.0	-5.0	4.0	6.4	0.86

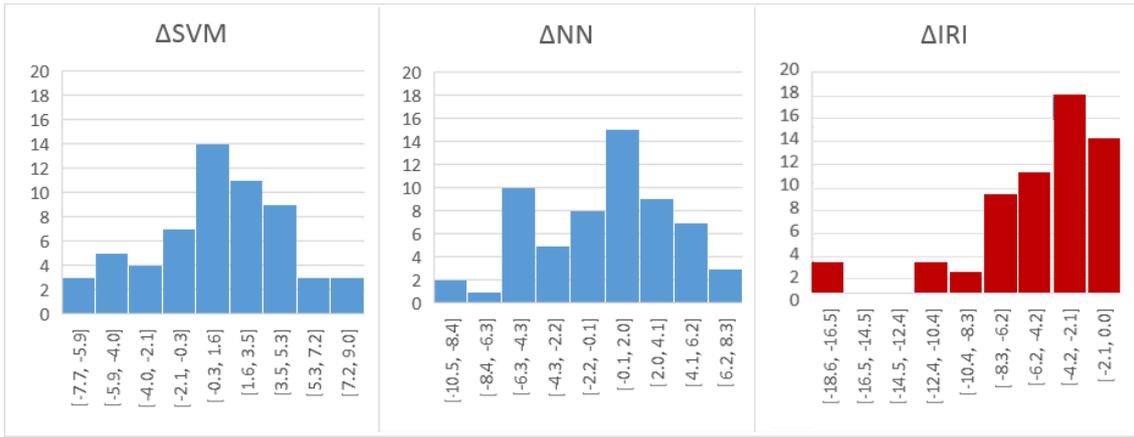


Figure 7. Histograms of the differences between actual and modeled TEC values for best performing SVM and NN techniques and for global IRI model.

are shown also in Fig. 7. Besides this, from the histogram of the IRI differences, we can see that there is a gap in differences distributions (range between, approximately, 12 TECU and 17 TECU).

Furthermore, due to the existence of maximum extreme values in 2015, the data from March 2015 were analyzed separately. The differences are presented in Tab. 2.

Table 2. Quality control measures of models for each station on March 2015.

Date	Station	TEC _{GNSS}	TEC _{SVM}	TEC _{NN}	TEC _{IRI}	Δ SVM	Δ NN	Δ IRI
March 18, 2015	S _K	21.8	21.4	22.2	16	-0.4	0.4	-5.8
	S _S	21.6	23.8	21.9	15.7	2.2	0.3	-5.9
	S _{NP}	22.9	23.5	22.6	16.2	0.6	-0.3	-6.7
March 19, 2015	S _K	23.2	19.5	19.2	16.0	-3.6	-4.0	-7.2
	S _S	24.1	21.3	19.4	15.7	-2.8	-4.7	-8.4
	S _{NP}	25.9	21.3	19.9	16.2	-4.6	-6.0	-9.8
March 20, 2015	S _K	19.2	20.6	20.9	16.0	1.4	1.7	-3.2
	S _S	18.5	22.8	20.8	15.6	4.3	2.3	-2.9
	S _{NP}	19.7	22.6	21.5	16.1	2.9	1.7	-3.6
March 21, 2015	S _K	26.9	21.0	21.5	15.9	-6.0	-5.4	-11.0
	S _S	27.3	23.3	21.3	15.5	-4.0	-6.0	-11.8
	S _{NP}	27.1	23.0	22.0	16.1	-4.1	-5.1	-11.0
March 22, 2015	S _K	33.9	26.2	29.5	15.9	-7.7	-4.4	-18.0
	S _S	32.8	28.7	28.6	15.5	-4.2	-4.3	-17.3
	S _{NP}	34.7	28.4	29.6	16.1	-6.2	-5.0	-18.6

The maximum extreme values in March 2015 were gained at 22nd of March when the actual TEC values reached the values of about 33 TECU for each station. Both ML techniques predicted those extreme values relatively well, with a maximum difference of -7.7 TECU. For the same stations on the same day, data

gained by using the IRI model have significantly higher differences value of approximately about 18 TECU. This is also easy to note from the histograms shown in Fig. 6 (already mentioned lack of the IRI model data).

4.2. Spatial – temporal prediction of TEC

In order to create and validate spatial – temporal TEC ML models, additional two independent data sets were created. Each of two data sets (labelled with t , where t can be equal 1 or 2) is represented as a row vector of attributes α_{it} , $i = 1, \dots, 7$, based on the values of TEC, SF, SSN, Ap index, lat, long, h , and Month and including TEC values as the target attribute. The first data set used for training, labelled as TR_2 , contained data from station S_K , S_{NP} for all investigated time points. The second data set used for a validation, labelled as TE_2 , contained data from station S_S . The actual TEC at each time point was used for comparison with modelled TEC values at station S_S .

The attribute selection was performed on the training data set TR_2 and the CFS automatically determines a subset of same four relevant attributes, Month, SF, lat and long, as in the case of previously derived selection on the training data set TR_1 . Two new training and test data sets that contain only four selected attributes were created, labelled as TR_2^{CFS} and TE_2^{CFS} , respectively. The optimal combination of parameters for training data set TR_2^{CFS} was found using 10 fold cross-validation and the best performing models are selected for both ML techniques. In order to compare the successes of spatial-temporal best performing TEC ML models and the global IRI model, the same quality control measures are used as for temporal prediction TEC models. The derived values of models quality control measures are presented in Tab. 3. Additionally, the histograms of the differences are presented in Fig. 8. The data in the Fig. 8. are labelled as follows:

- $\Delta_{SVM} = TEC_{SVM} - TEC_{GNSS}$,

as differences between TEC obtained by TE_2^{CFS}/SVM model and actual TEC values,

- $\Delta_{NN} = TEC_{NN} - TEC_{GNSS}$,

as differences between TEC obtained by TE_2^{CFS}/NN model and actual TEC values,

- $\Delta_{IRI} = TEC_{IRI} - TEC_{GNSS}$

as differences between TEC obtained by global IRI model and actual TEC values.

Table 3. Best performing spatial-temporal models by using TE_1^{CFS} data set.

Data sets / ML techniques	Quality controls measure					
	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>RMSE</i>	<i>R²</i>
TE_2^{CFS}/SVM	-7.7	2.2	-0.4	1.5	1.6	0.995
TE_2^{CFS}/NN	-4.7	3.7	0.2	1.7	1.7	0.994
Data sets / Model						
$TE_1^{CFS}/IRI2012$	-35.9	3.2	-6.8	8.3	10.7	0.864

The obtained standard deviation and RMSE are equal or less than 1.7 TECU and mean values of differences between actual and modelled TEC values are -0.4 TECU (with a range from -7.7 TECU to 2.2 TECU) and 0.2 TECU (with a range from -4.7 TECU to 3.7 TECU), for spatial-temporal TEC SVM and NN based model, respectively. Furthermore, considering that differences between actual TEC values per station vary from 1 to 10 TECU at the same time point, it can be concluded that the obtained results of spatial TEC are also satisfactory. Once again, the distribution of TEC differences (Fig. 8) indicates that NN have slightly better performance in comparison to SVM. Furthermore, it is evident that the results obtained from spatial TEC SVM based model contained grouped outliers. As in the case of solely temporal prediction, the results obtained by the IRI model displayed significantly larger values for the quality control measures compared to the same measures derived by SVM and NN techniques. The values of standard deviation and RMSE for IRI model are approximately ten times larger than values of other two derived models. The R^2 values for TEC SVM and TEC NN reach values of almost 1, while for the global model R^2 is on the same level as in the case of temporal prediction of TEC.

In addition, we further analyzed the results derived for the year 2014, considering the existence of the maximum extreme values in examined time points. The differences of the actual and modelled values of TEC are presented in Tab. 4.

Since those maximum extreme values occurred in March 2014, it can be expected that prediction of the IRI model derived maximum differences, Δ IRI values have ranged from -25.9 TECU to -35.9 TECU. Based on the results in Tab. 4., it is evident that the IRI model has significantly higher values of differences. The both ML techniques show that are capable to "learn" large variation in TEC values. It can be concluded that for the examined study area and used set of data, ML technique provides a solution with significantly less difference of spatial temporal prediction of TEC values compared to IRI model.

Comparing the results presented in Tab. 1 and Tab. 3, it can be concluded that both ML techniques define the trend of TEC values and its variations through space and time more efficiently than just through time. Those obtained results

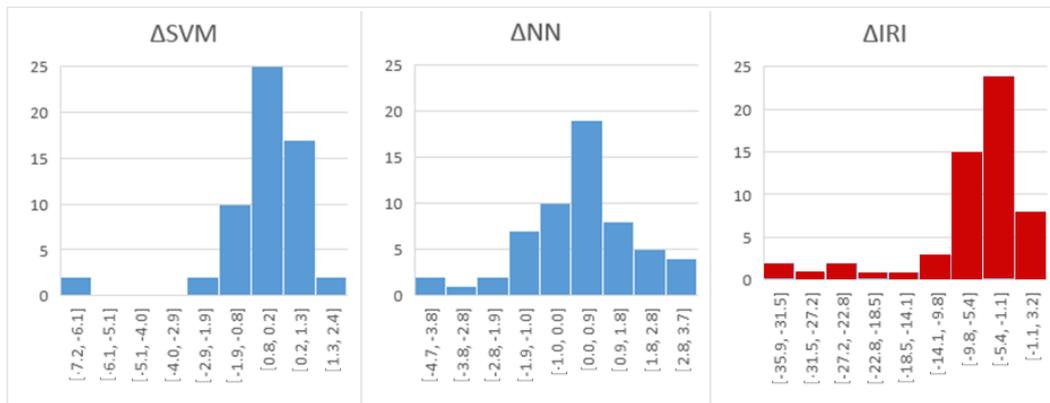


Figure 8. Histograms of the differences between actual and modelled TEC values for best performing SVM and NN techniques and for global IRI model.

Table 4. Models quality control measures for station S_{NP} at 2014.

Date	TEC _{GNSS}	TEC _{SVM}	TEC _{NN}	TEC _{IRI}	Δ SVM	Δ NN	Δ IRI
March 18, 2014	49.0	48.1	48.1	16.8	0.6	-1.0	-32.2
March 19, 2014	42.7	41.0	43.0	16.8	1.2	0.3	-25.9
March 20, 2014	44.1	42.5	44.0	16.8	-0.1	-0.1	-27.4
March 21, 2014	52.6	51.0	50.6	16.7	0.4	-1.9	-35.9
March 22, 2014	43.2	42.2	41.5	16.7	0.1	-1.7	-26.5
June 18, 2014	21.5	21.9	21.7	12.3	0.5	0.2	-9.2
June 19, 2014	10.7	11.1	11.5	12.3	0.6	0.8	1.6
June 20, 2014	14.6	14.0	14.6	12.3	0.7	0.1	-2.3
September 18, 2014	22.3	20.8	22.6	14.0	-0.4	0.3	-8.3
September 19, 2014	22.1	23.0	22.4	14.1	0.3	0.3	-8.0
September 22, 2014	26.3	25.2	26.5	14.2	-2.4	0.2	-12.1
September 23, 2014	20.4	19.6	20.9	14.3	-0.3	0.6	-6.1
September 24, 2014	22.0	19.2	20.3	14.4	-1.6	-1.7	-7.5
December 22, 2014	31.2	29.9	31.5	16.1	-2.6	0.2	-15.1
December 23, 2014	23.7	23.0	24.2	16.0	0.4	0.5	-7.7
December 24, 2014	24.7	24.2	24.2	16.0	-0.6	-0.4	-8.6
December 25, 2014	19.1	18.5	21.7	16.0	-0.4	2.6	-3.1
December 26, 2014	29.2	28.2	26.3	16.0	-1.6	-2.9	-13.2

were expected due to the existence of larger variations of TEC values through time in comparison to the variations through space.

5. Conclusion

This research study examined the possibility of two most commonly used ML techniques, SVM and NN, in order to model extreme values of TEC. For that purpose, we used GNSS observation from three AGROS stations, distributed within the territory of Serbia. The largest variation of TEC follows the changes in solar activity (11-year cycle of Sun activity, winter and summer solstice and autumnal and vernal equinox). Therefore, the average values of TEC between 10 and 12 UT, for five days period, for each season and for three years of interest (2013, 2014 and 2015) were calculated from GNSS observations from each station and used as samples for modelling. Two types of models were examined, the spatial-temporal and just temporal. Therefore, two pairs of appropriate independent training and test data sets were created for each type of model.

The used factors (attributes) of TEC changes were Solar flux, Sunspot number, Index of geomagnetic activity, geographic coordinates of stations and one additional attribute (Month) generated in order to represent the time intervals. In order to remove uninformative attributes, CFS attribute selection method was performed and two new pairs of training and test data sets that contain only four attributes (Solar flux, Month, latitude and longitude) were created.

After performing attribute selection by CFS method and finding the appropriate and optimal combination of parameters for both ML techniques and types of models, the best performing models were analyzed.

Furthermore, in order to examine the performance of each type of TEC model based on NN and SVM techniques we compared the differences between TEC values based on GNSS and the well-known global IRI model and the differences between TEC values based on GNSS and model outcomes obtained by ML techniques. The obtained results for both types of models indicate that used ML techniques provided TEC values significantly closer to TEC values obtained from GNSS observation, compared to TEC values derived from IRI model, especially from the spatial temporal model. Both ML techniques are capable to “learn” relationship between considering attributes and extreme TEC values and two derived TEC values with large variation through examining time intervals.

Generally, the differences between the results obtained based on SVM and NN almost identical, with slightly better values of quality control parameters obtained by NN. The values of quality control measures indicate that both techniques are capable to adequately predict and spatially model extreme TEC values. Based on the results it can be concluded that both ML techniques define the trend of TEC values and its variations through space and time more efficiently than through time. Since that it can be expected that the model can be improved using a larger number of samples and time intervals, in future work our attention will be dedicated to extending the samples.

Acknowledgment – The support for this study was made through the Ministry of Education, Science and Technological Development of the Republic of Serbia (TR36020) awarded to first, third, fourth and fifth authors and through the Ministry of Education, Science and Technological Development of the Republic of Serbia (TR 36009) awarded to the second author.

References

- Abe, S. (2010): *Support vector machines for pattern classification*. Springer, London, 471 pp.
- Barrile, V., Cacciola, M., Morabito, F. C. and Versaci, M. (2006): TEC measurements through GNSS and artificial intelligence, *J. Electromagnet. Wave.*, **20**, 1211–1220, [DOI: 10.1163/156939306777442962](https://doi.org/10.1163/156939306777442962).
- Davies, K. and Hartmann, G. K. (1997): Studying the ionosphere with the global positioning system, *Radio Sci.*, **32**, 1695–1703, [DOI: 10.1029/97RS00451](https://doi.org/10.1029/97RS00451).
- Dugas, C., Bengio, Y., Bélisle, F. and Nadeau, C. (2001): Incorporating second order functional knowledge into learning algorithms, *Advances in Neural Information Processing Systems*, **13**, 472–478.
- Elmunim, N. A., Abdullah, M., Hasbi, A. M., and Bahari, S. A. (2016): Comparison of GPS TEC variations with Holt–Winter method and IRI–2012 over Langkawi, Malaysia, *Adv Space Res.*, **60**, 276–285, [DOI: 10.1016/j.asr.2016.07.025](https://doi.org/10.1016/j.asr.2016.07.025).
- Habarulema, J. B. and McKinnell, L. A. (2012): Investigating the performance of neural network backpropagation algorithms for TEC estimations using South African GNSS data, *Ann. Geophys.*, **30**, 857–866.
- Hall, M. A. (1999): *Correlation-based feature selection for machine learning*. Ph.D. Thesis, Department of Computer Science, University of Waikato, 178 pp.

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009): The WEKA data mining software: an update, *ACM SIGKDD explorations newsletter*, **11** (1), 10–18, [DOI: 10.1145/1656274.1656278](https://doi.org/10.1145/1656274.1656278).
- Hofmann–Wellenhof, B., Lichtenegger, H. and Collins, J. (1992): *Global positioning system, theory and practice*. Springer–Verlag, Berlin, Heidelberg, New York, 4th edition, 389 pp.
- Huang, Z. and Yuan, H. (2014): Research on regional ionospheric TEC modeling using RBF neural network, *Sci. China Technol. Sc.*, **57**, 1198–1205, [DOI: 10.1007/s11431-014-5550-0](https://doi.org/10.1007/s11431-014-5550-0).
- Igarashi, K., Nakamura, M., Wilkinson, P., Wu, J., Pavelyev, A., and Wickert, J. (2001): Global sounding of sporadic E layers by the GPS/MET radio occultation experiment, *J. Atmos. Sol–Terr. Phy.*, **63**, 1973–1980, [DOI: 10.1016/S1364-6826\(01\)00063-3](https://doi.org/10.1016/S1364-6826(01)00063-3).
- Kim, Y. S., Street, W. N and Menczer, F. (2003): Feature selection in data mining, in: *Data mining: Opportunities and challenges*, edited by Wang, J. Idea Group Inc., 80–105.
- Kleusberg, A. and Teunissen, P. J. (1996): *GPS for geodesy*. Springer–Verlag Berlin Heidelberg, 218 pp.
- Klobuchar, J. A. (1986): Design and characteristics of the GPS ionospheric time delay algorithm for single frequency users, in: *Proceedings of PLANS – Position Location & Navigation Symposium*, 4–7 November, Las Vegas, Nevada, 280–286.
- Langley, R., Fedrizzi, M., Paula, E., Santos, M and Komjathy, A. (2002): Mapping the low latitude ionosphere with GPS, *GNSS W.*, **13**, 41 –46.
- Okoh, D., Owolabi, O., Ekechukwu, C., Folarin, O., Arhiwo, G., Agbo, J., Bolaji, S. and Rabiou, B. (2016): A regional GNSS–VTEC model over Nigeria using neural networks: A novel approach, *Geod. Geodyn.*, **7**, 19–31, [DOI: 10.1016/j.geog.2016.03.003](https://doi.org/10.1016/j.geog.2016.03.003).
- Radzi, Z. R., Abdullah, M., Hasbi, A. M., Mandeep, J. S., Bahari, S. A. (2013): Seasonal variation of Total Electron Content at equatorial station, Langkawi, Malaysia, in: *Proceeding of the 2013 IEEE International Conference on Space Science and Communication (IconSpace)*, 1–3 July, Melaka, Malaysia, 186–189.
- Razin, M. R. G. and Voosoghi, B. (2016): Wavelet neural networks using particle swarm optimization training in modeling regional ionospheric total electron content, *J. Atmos. Sol.-Terr. Phy.*, **149**, 21–30, [DOI: 10.1016/j.jastp.2016.09.005](https://doi.org/10.1016/j.jastp.2016.09.005).
- Rumelhart, D., Hinton, G and Williams, R. (1986): Learning internal representation by error propagation, in *Parallel distributed processing: Explorations in the microstructures of cognition*, MIT Press, Cambridge, **1**, 318–362.
- Sardón, E. and Néstor, Z. (1997): Estimation of total electron content using GPS data: How stable are the differential satellite and receiver instrumental biases?, *Radio Sci.*, **32**, 1899–1910, [DOI: 10.1029/97RS01457](https://doi.org/10.1029/97RS01457).

- Seemala, G. K. (2014): *GNSS–TEC analysis application*. Institute for Scientific Research, Boston College, USA, ... pp.
- Seemala, G. K. and Valladares, C. E. (2011): Statistics of total electron content depletions observed over the South American Continent for the year 2008. *Radio Sci.*, **46**, RS5019, [DOI: 10.1029/2011RS004722](https://doi.org/10.1029/2011RS004722).
- Sur, D. and Paul, A. (2013): Comparison of standard TEC models with a Neural Network based TEC model using multistation GPS TEC around the northern crest of Equatorial Ionization Anomaly in the Indian longitude sector during the low and moderate solar activity levels of the 24th solar cycle. *Adv Space Res.*, **52**, 810–820, [DOI: 10.1016/j.asr.2013.05.020](https://doi.org/10.1016/j.asr.2013.05.020).
- Swamy, K. C. T., Sarma, A. D., Srinivas, V. S., Kumar, P. N., Somasekhar Rao, P. V. D. (2013): Accuracy evaluation of estimated ionospheric delay of GNSS signals based on Klobuchar and IRI–2007 models in low latitude region. *IEEE Geosci. Remote Sens. Lett.*, **10**, 1557–1561, [DOI: 10.1109/LGRS.2013.2262035](https://doi.org/10.1109/LGRS.2013.2262035).
- Wautelet, G and Warnant, R. (2014): Climatological study of ionospheric irregularities over the European mid–latitude sector with GPS. *J. Geodesy*, **88**, 223–240, [DOI: 10.1007/s00190-013-0678-4](https://doi.org/10.1007/s00190-013-0678-4).

SAŽETAK

Modeliranje ekstremnih vrijednosti ukupnog sadržaja elektrona na području Srbije

*Miljana Todorović Drakul, Mileva Samardžić Petrović, Sanja Grekulović,
Oleg Odalović i Dragan Blagojević*

Ovaj rad je posvećen modeliranju ekstremnih TEC vrijednosti (*Total Electron Content – ukupan sadržaj elektrona*) na području Srbije. Pod ekstremnim TEC vrijednostima razmatrali smo one koje se javljaju u toku maksimuma Sunčeve aktivnosti koji se periodično ponavljaju svakih 11 godina i to za odabrane 2013., 2014. i 2015. godinu. U navedenim godinama razmatrane su srednje vrijednosti TEC–a tijekom ljetnje i jesenske ravnodnevnice i u dnevnom razdoblju 10–12 UTC (*Univerzalno vrijeme*). Navedene srednje vrijednosti TEC određivane su na osnovu GNSS opažanja (*Globalni Navigacijski Satelitski Sustavi*) tri stanice trajno smještene na teritoriju Srbije. Podaci su u okviru ovih istraživanja tretirani kao ‘uvjetno točne’ vrijednosti. Osnovni cilj istraživanja bio je ispitivanje mogućnosti korištenja dviju metoda strojnog učenja: neuronskih mreže i metode podržavajućih (potpornih) vektora. U cilju naglašavanja kvalitete primijenjenih tehnika svi rezultati su adekvatno uspoređivani sa TEC vrijednostima određenim Međunarodnim Referentnim Ionosfernim globalnim modelom (IRI). Pored navedenog posebno je analizirana kvaliteta metoda kroz vremenski i prostorno–vremenski pristup.

Ključne riječi: TEC, strojno učenje, ionosfera, GNSS, IRI

Corresponding author's address: Miljana Todorović Drakul, Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Boulevard of King Aleksandar 73, 11000 Belgrade, Serbia; tel: +381 11 3370 293; e–mail: mtodorovic@grf.bg.ac.rs