
# TOTAL ELECTRON CONTENT PREDICTION USING MACHINE LEARNING TECHNIQUES

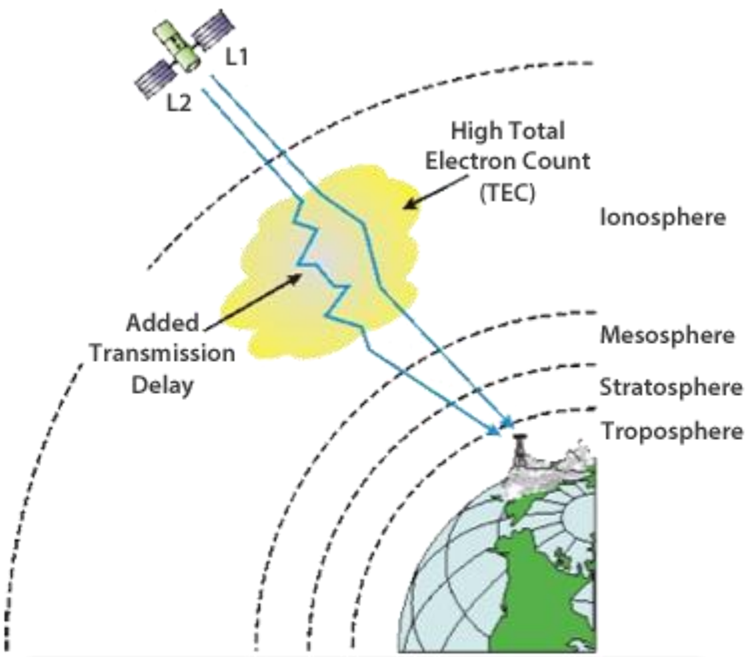
M. Todorović -Drakul[1], M. Samardžić-Petrović, S. Grekulović[1], O. Odalović[1], D. Blagojević[1]

*[1]Department of geodesy and geoinformatics, University of Belgrade, Faculty of Civil Engineering, Serbia*

GeoMLA 2016, Belgrade – June 22-24

# What is TEC and why is it important to modeling?

## Total Electron Content -TEC

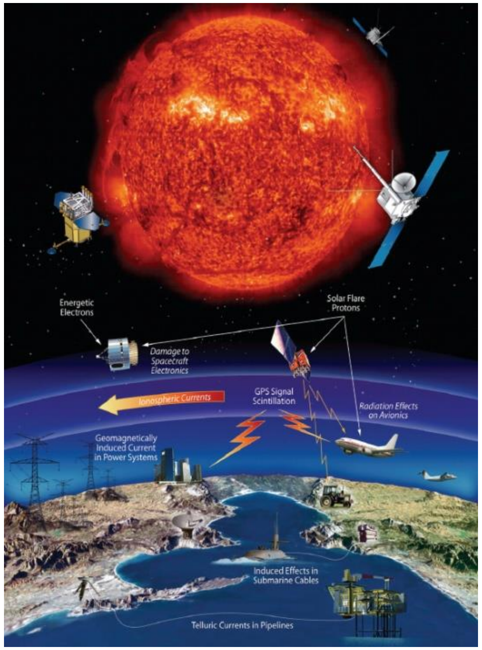



*Effect of ionospheric refraction. The GPS signal are affected in different ways, depending on whether it is a question of codes or phases.*

$$\boxed{TEC = \int n_e(s)dS}$$

$\downarrow$

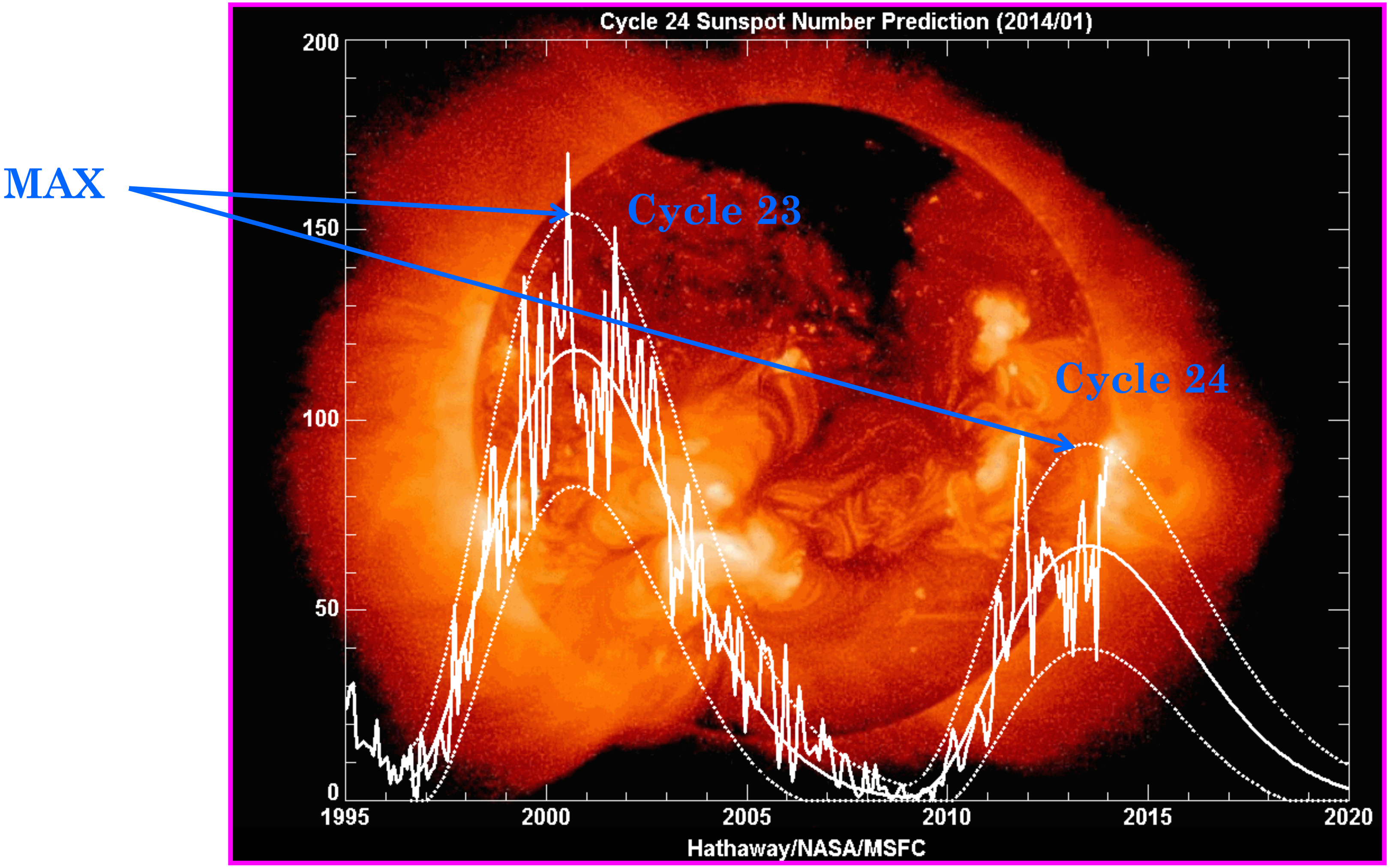
Unit: TECU=$10^{16}$el/m$^2$




http: reflexions.ulg.ac.be

# Variations of TEC values and its extreme

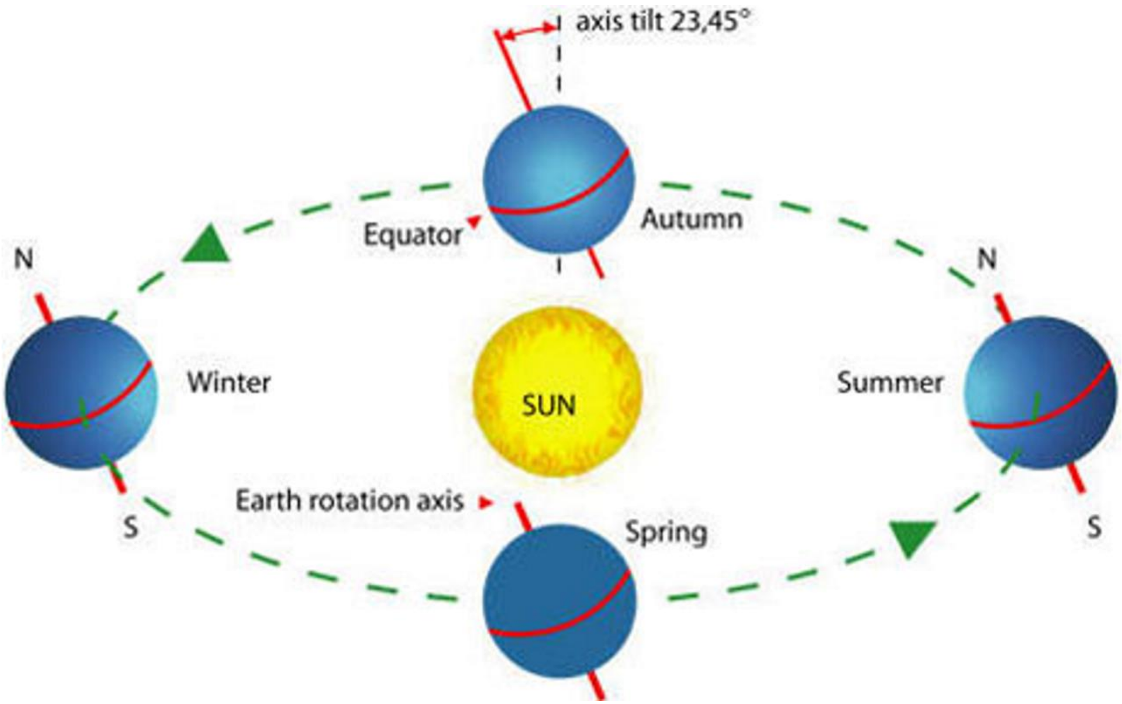
## 11 year cycle of solar activity




http://www.ph.ucla.edu/epi/snow.html

# Variations of TEC values and its extreme

## Earth's revolution around the Sun in the period of the equinox and solstice




**SolarFlux**

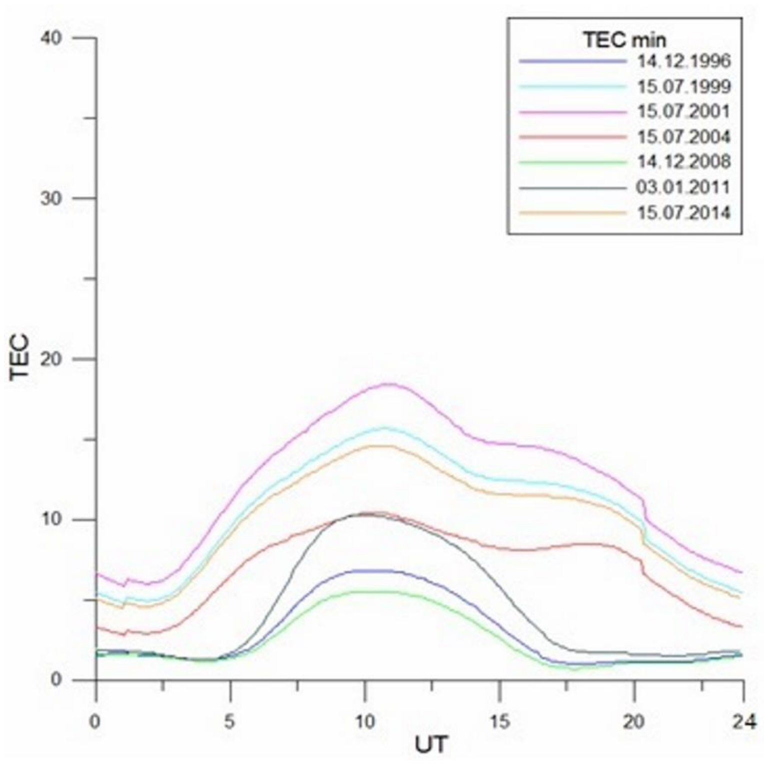
**Ap index**

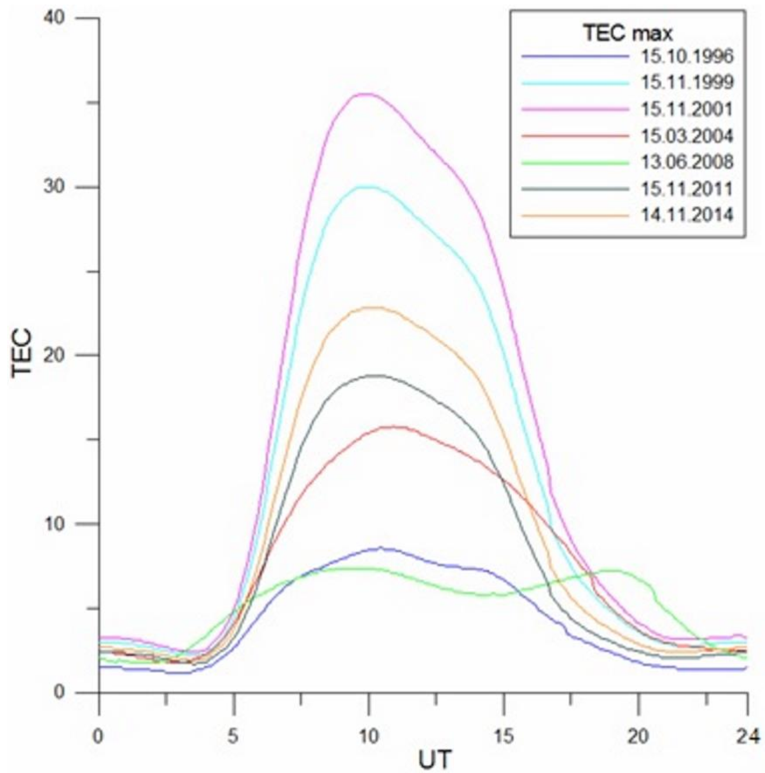
**SunSpotNumber**

http://www.mshid.wordpress.com

# Variations of TEC values and its extreme

## Daily TEC changes for minimum and maximum solar conditions



# Why Machine Learning techniques?

**TEC values** vary spatially and temporally, dependent on many factors - **complex nonlinear** problem.

Machine learning techniques - empiric modeling approaches that have the capability to extract information and reveal patterns by exploring the data.

**Some of ML techniques:**

Decision Tree,

Random Forest,

**Neural Networks (NN),**

**Support Vector Machines (SVM).**

# Objectives

The main objective of this research study is to examine the capability of SMV and NN techniques to model/predict extreme TEC values.

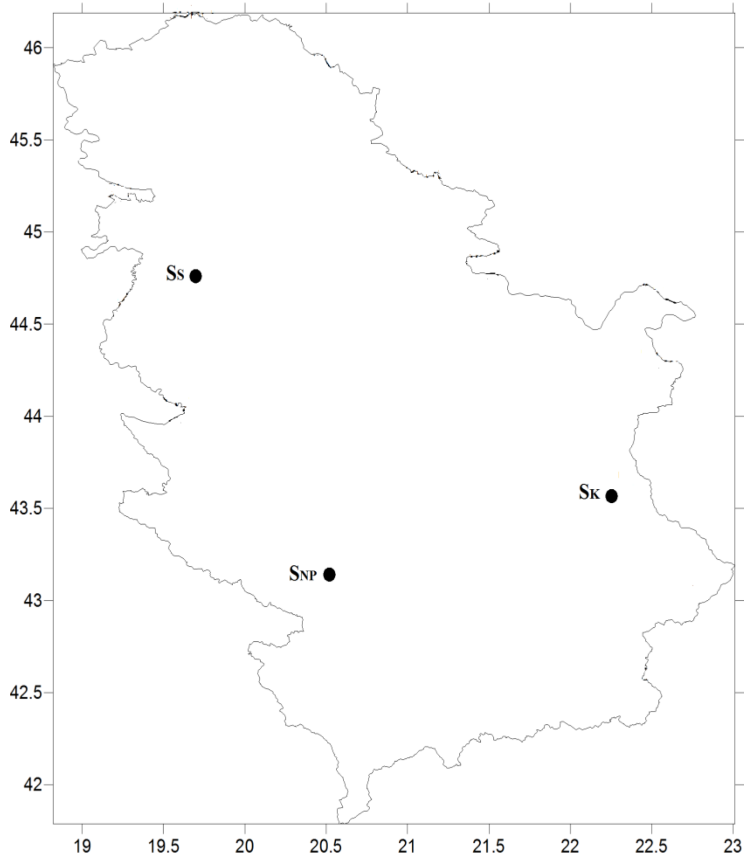
To examine and analyze the capability of SVM and NN to model:

Spatial – temporal TEC values

Spatial TEC values

# Data

## Three base stations: Kanjiža (SK), Novi Pazar (SNP) and Šabac (SS)




- Stations belong to the permanent GNSS network of the Republic of Serbia under the name AGROS (Active Geodetic Reference Base of Serbia)
- In the form of 30-second RINEX files
- Values of TEC are calculated and averaged for 10-14 UT time interval, for five days, for each season and for three years of interest (2013, 2014 and 2015)

# TEC based on GPS observations

Delay $\Delta t = t_2 - t_1$, measurement between L1 and L2 frequencies:

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

$f_1 = 1575.42$ MHz
$f_2 = 1227.60$ MHz

c - speed of light in open space

Vertical TEC (VTEC):

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

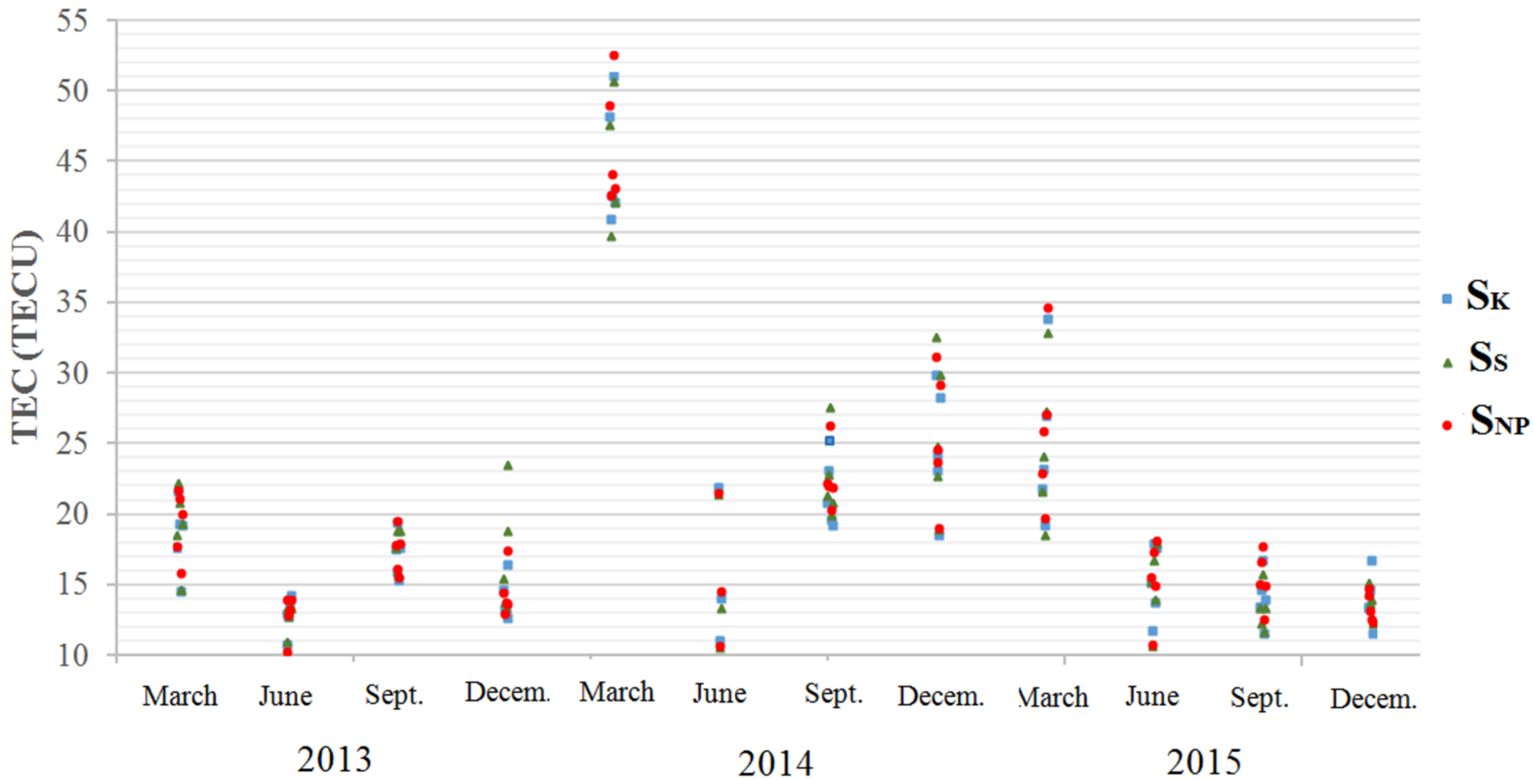
STEC - slanted TEC,
$b_S$ - hardware satellite delay,
$b_R$ - hardware receiver delay,
$e$ - elevation angle of satellites in degree,.

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

$S(e)$ - slant factor
$Re$ - the average Earth's radius in km
$hi$ - the (effective) height of ionosphere over the Earth's surface
$Z$ - the zenith angle

# TEC based on GPS observations

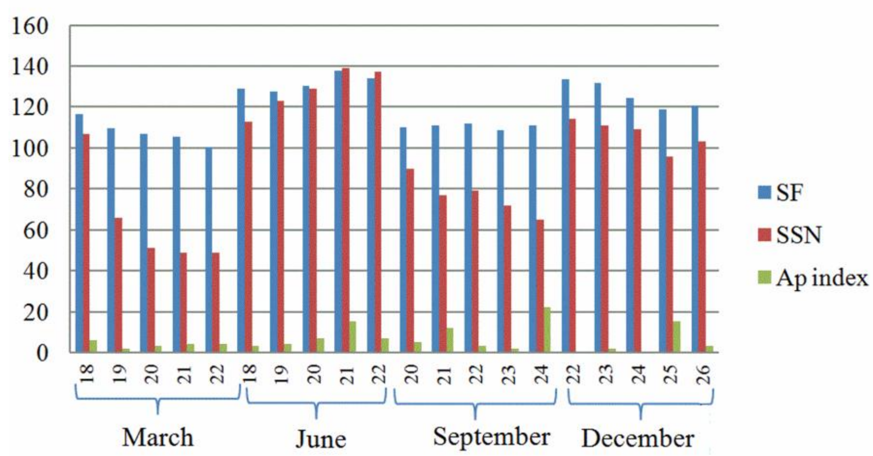
## Distribution of TEC values for examined time intervals based on observation from three stations, SK, SS and SNP



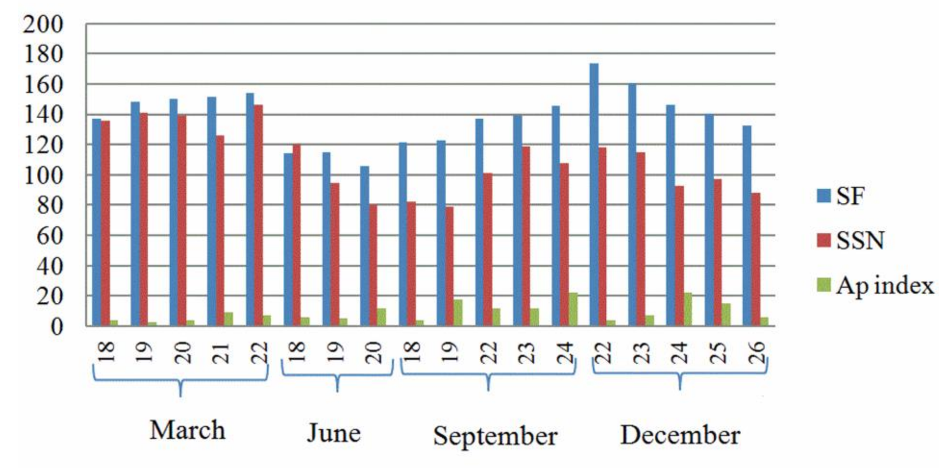
# Data - Indicators of  TEC changes

## Solar flux (SF), Sunspot number (SSN) and Index of geomagnetic activity (Ap index)

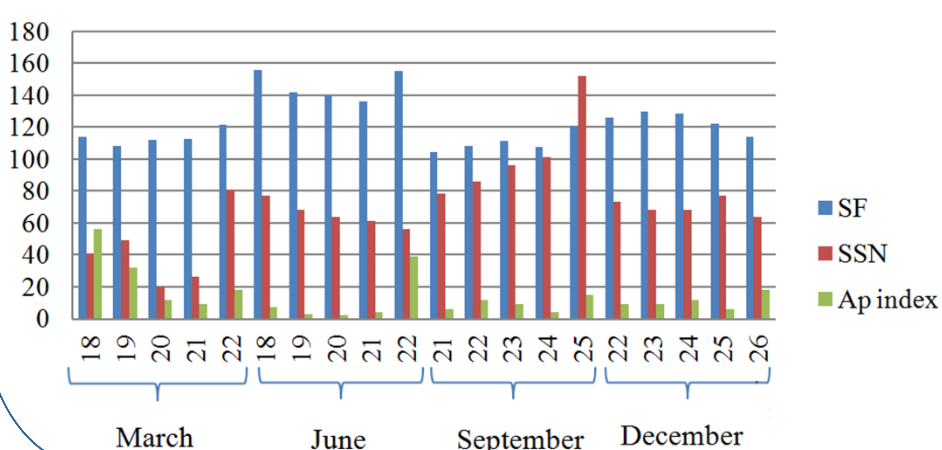
**2013**




**2014**




**2015**



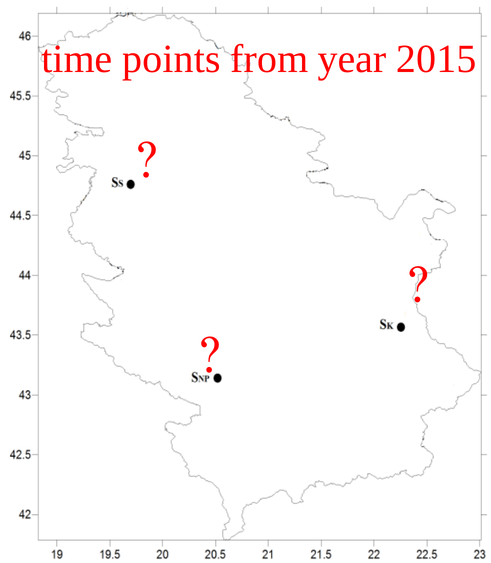

Values were downloaded from NASA's Space Physics Data Facility (http://omniweb.gsfc.nasa.gov/form /dx1.html), for all 12 periods of interest

# TEC based on ML techniques - Datasets

## Spatial – temporal ML TEC model

$TR_1$ training dataset - time points from years 2013 and 2014

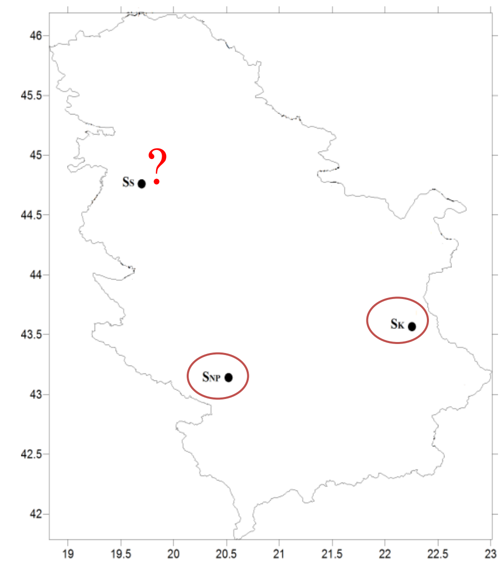
$TE_1$ test dataset - time points from year 2015, actual TEC was used for comparison with the predicted TEC values in the year 2015.




## Spatial ML TEC model

$TR_2$ training dataset - data from stations $S_K$, $S_{NP}$ for all investigated time points

$TE_2$ test dataset - data from station $S_S$, where the actual TEC at all time points was used for comparison with modeled TEC values at station $S_S$



# TEC based on ML techniques - Datasets

| Attributes | Attribute description |
|---|---|
| SF | Solar flux |
| SSN | Sunspot number |
| Ap index | Index of geomagnetic activity – average values between 10 and 14 UT |
| Lat | Geographic latitude |
| Long | Geographic longitude |
| h | Height |
| Month | Time intervals in which TEC value was obtained in regards to winter and summer solstice and autumnal and vernal equinox |

# TEC based on ML techniques - Attribute selection

Finding a subset of most informative attribute:

- Improving accuracy of the model
- Reducing model complexity
- Reducing the time required to train (learn)

Method used for attribute selection:

- Correlation-based Feature Subset – CFS

CFS automatically determines a subset of k (k >> n) relevant attributes that are highly correlated with the target attribute (TEC) but uncorrelated with each other.

Attribute selection by CFS

- Solar flux
- Latitude
- Longitude
- Month

# TEC based on ML techniques

Datasets:

Spatial-temporal: $TE_{1CFS}$ and $TR_{1CFS,}$

Spatial: $TR_{2CFS}$ and $TE_{2CF}$

ML algorithms:

NN: Multi-layer Perceptron (MLP), with softplus activation function

SVM: Kernel function-Radial Basis Function (RBF)

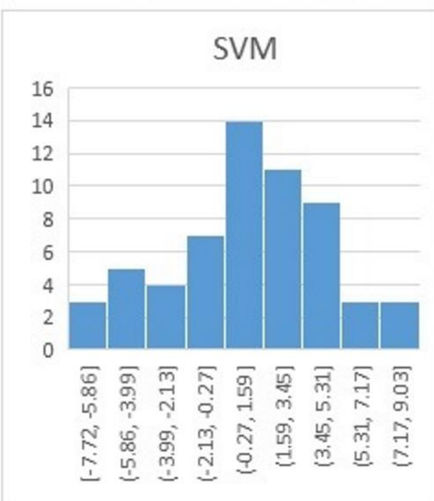
Using training datasets and 10 fold cross-validation, optimal combination of parameters was found for both ML techniques and for both types of models.

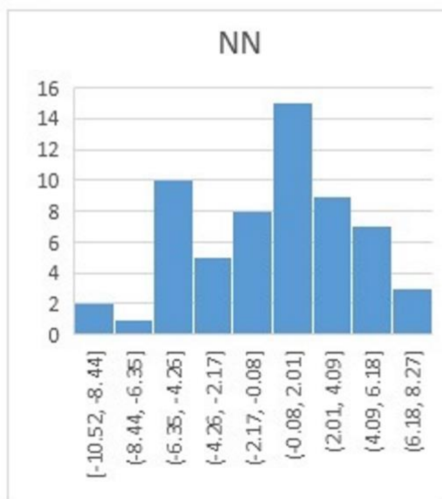
Software: Weka (SMOreg algorithm and MLPreg)

# TEC based on GPS observations

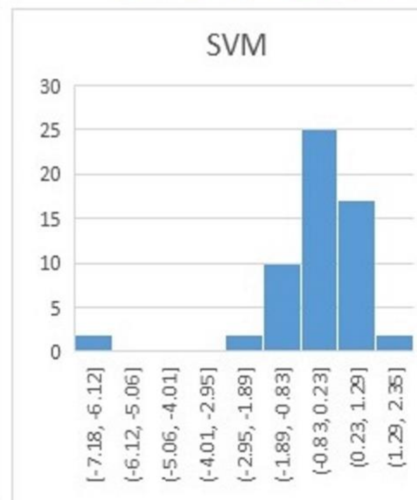
| Datasets/ ML techniques | Quality controls measure | | | | |
|---|---|---|---|---|---|
| | Min | Max | Mean | St.Dev | RMSE |
| $TE_{1CFS}$/SVM | -7.72 | 9.03 | 1.19 | 3.84 | 4.02 |
| $TE_{1CFS}$/NN | -10.52 | 8.27 | 0.03 | 4.10 | 4.10 |
| $TE_{2CFS}$/SVM | -7.72 | 2.23 | -0.36 | 1.54 | 1.58 |
| $TE_{2CFS}$/NN | -4.72 | 3.72 | 0.15 | 1.69 | 1.70 |

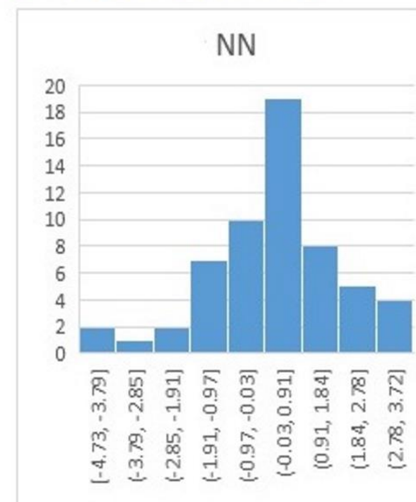
Diffrencise betwine actual and predicted TEC valus for set $TE_{1CFS}$




Diffrencise betwine actual and modeled TEC valus for set $TE_{2CFS}$



# Conclusion

- The SVM and NN techniques are capable to adequately predict and spatially model extreme TEC values
- The differences between the results obtained based on SVM and NN models are small
- Both ML techniques define trend of TEC values and its variations through space more efficiently then through space and time.

In future work our attention will be dedicated to extending the samples.

# THANK YOU !

Contact information:
Miljana Todorović-Drakul: mtodorovic@grf.bg.ac.rs