An experiment of predicting Total Electron Content (TEC) by fuzzy inference systems

O. Akyilmaz[1] and N. Arslan[2]

[1]*Department of Geodesy and Photogrammetry Engineering, Istanbul Technical University,*
34469 Maslak-Istanbul, Turkey
[2]*Department of Geodesy and Photogrammetry Engineering, Yildiz Technical University,*
34349 Besiktas-Istanbul, Turkey

The Total Electron Content (TEC) is predicted by fuzzy inference systems for various station-satellite pairs. GPS data from the GRAZ, HFLK, LINZ, MOPI and UZHL permanent stations are processed in order to obtain the vertical total electron content (VTEC) using differenced carrier-smoothed code observations. The quality of the VTEC prediction was studied on 9 and 11 September 2005 (DOY 252 and 254). The predictions were computed for 5, 10 and 15 min intervals. The mean accuracies of predictions are about 0.1, 0.2 and 0.3 TECU for these time intervals. More than 98% of the VTEC is successfully recovered with the proposed prediction method.
Key words: GPS, ionosphere, VTEC, prediction.

1. Introduction

Determining the Total Electron Content (TEC) from dual-frequency GPS observations has become an important application of GPS. The measured TEC along the signal path, also known as slant TEC, is converted to the vertical TEC (VTEC) using mapping functions. The slant TEC is estimated using the geometry-free linear combination by simply differencing the carrier-phase smoothed L_1 and L_2 code observations. The differential code biases (DCB) for the receivers and satellites must also be taken into account when estimating the TEC (Sardon and Zarraoa, 1997; Rideout and Coster, 2006). These mapping functions often assume that the electrons are concentrated in a shell of infinitesimal thickness. Such a single layer model is used in this study (Dach *et al.*, 2007).

There are variety of error sources which affect the estimation of VTEC. For example, Rideout and Coster (2006) extensively discussed mapping function errors, Ciraolo *et al.* (2007) have studied the effects of the multipath on carrier-phase smoothed code observations using co-located receivers and intra-daily DCB variations for a zero-baseline experiment using single difference observations.

Predicting the TEC has also received much attention. For example, Liu and Gao (2004) estimated model parameters and a geometry matrix which describes the dependency of the data on those model parameters that are used for the TEC prediction. Liu *et al.* (2005) applied autocorrelation method for short-term forecasting of ionospheric characteristics. Oyeyemi *et al.* (2006) used neural network technique in order to estimate hourly daily $(f_o F_2)$ ionospheric parameters using long term data for near real-time predictions.

In this study, we use the adaptive network-based fuzzy inference system (ANFIS) prediction method developed by Jang (1993). It uses a supervised learning algorithm to optimize parameters of the fuzzy inference system. The focus is on the efficiency of the proposed method for VTEC prediction.

2. Modeling of the Total Electron Content (TEC)

The carrier-smoothed code observations can be written as

$$\tilde{P}_{1k}^i = \rho_k^i + I_{1k}^i + T_k^i + c\delta_k - c\delta^i + e_{1,k}^i \qquad (1)$$

$$\tilde{P}_{2k}^i = \rho_k^i + \frac{f_1^2}{f_2^2} I_{1k}^i + T_k^i + c\delta_k - c\delta^i + e_{2,k}^i \qquad (2)$$

The symbol ρ_k^i is the geometric distance between the satellite and the receiver, c is the speed of light in a vacuum, δ_k is the receiver clock error, δ^i is the satellite clock error, I_{1k}^i is the ionospheric phase delay, T_k^i is the tropospheric delay, f_1 and f_2 are the carrier frequencies, and $e_{1,k}^i$ and $e_{2,k}^i$ represent the remaining measurement noise and multipath for pseudoranges on L_1 and L_2.

When smoothing the codes it is necessary to analyze and correct the code and phase observations for outliers and cycle slips. This data screening were performed by the Bernese GPS software 5.0. The details of these procedure can be found in Dach *et al.* (2007).

Differencing (2) and (1) gives

$$\tilde{P}_4 = +\alpha \left(\frac{1}{f_1^2} - \frac{1}{f_2^2} \right) F_I(z) E_v + c(\Delta b^i - \Delta b_k) \qquad (3)$$

The symbol α is $40.3 \cdot 10^{16}$ m s^{-2} TECU^{-1}, E_v is the vertical TEC, Δb^i is the differential code bias of the satellite i, Δb_k is the DCB of the receiver k, $F_I(z)$ is the mapping function in the single layer model. TEC is represented by
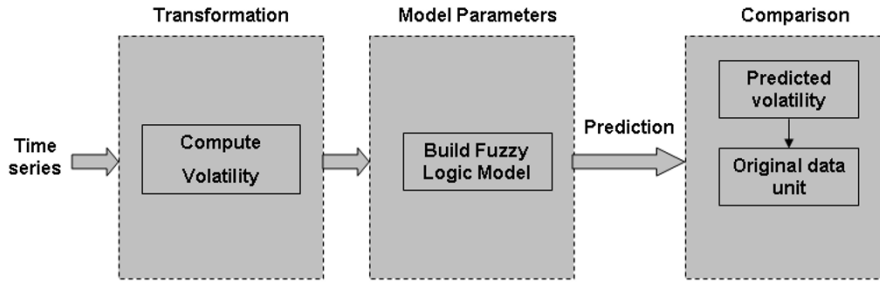
Fig. 1. Architecture of the prediction model.

TEC unit (TECU) where one TECU is equal to 10^{16} el/m^2. Single layer model (SLM) approach can be used in modeling of TEC. In this model, $F_I(z)$ is used to convert the slant TEC(E) to VTEC as follows:

$$F_I(z) = \frac{E}{E_v} = \frac{1}{\cos z'} \tag{3a}$$

and

$$\sin z' = \frac{R}{R+H} \sin z \tag{3b}$$

In these equations, z and z' are the zenith distances at the height of the station and the single layer, R is the radius of the earth, H is the height of the single layer (Wild, 1994). The height of the ionospheric pierce point at single layer selected as 400 km. The satellite DCB values were obtained from the CODE DCB data archive. Daily DCB values for the receivers were estimated by geometry-free linear combination using the Bernese GPS software.

3. Prediction Methodology

Given a time series $X = \{\text{VTEC}(t), t = 1, ..., n\}$, where VTEC($t$) is a value at discrete time t and n is the number of data points in the time series, we wish to predict the value VTEC($t + 1$) at time $t + 1$. The strategy consists of three main steps (Fig. 1).

The first step involves data transformation computing an equivalent volatility index (Popoola *et al.*, 2004),

$$r(t) = \log\left(\frac{\text{VTEC}(t+1)}{\text{VTEC}(t)}\right) \tag{4}$$

Volatility is a measure of the abrupt local changes in the time series. This transformation is carried out for VTEC time series. The benefit of using the volatility time series instead of the estimated TEC data is that it reduces adverse effects of possible linear trends on the predictions, since either fuzzy systems or neural networks do not provide precise predictions for the data that are not within the range of training data sets. Therefore, we use volatility series as training data instead of original VTEC data in order to overcome this problem.

In the second step, the fuzzy logic prediction model is set up using the volatility series. This step includes the generation of the training patterns for the FIS and the estimation of the FIS parameters via a training procedure. For this study,

the training pattern has been composed as follows:

$$\{r(t-3), r(t-2), r(t-1)\} \rightarrow r(t)$$
$$\downarrow \qquad\qquad\qquad \downarrow \tag{5}$$
$$\text{input vector} \qquad\qquad \text{output}$$

It follows that the FIS used for prediction has three input variables and one output variable, i.e., the previous three elements in the volatility time series are used to predict the next element. Each input variable is represented by a fuzzy membership function (MF), which can be of triangular, trapezoid, or gaussian form. In this study, the Gaussian MF is used to represent the input variables. The output variable is represented by MFs which are linear functions of the input variables of the Takagi-Sugeno type FIS (Takagi and Sugeno, 1985). The Takagi-Sugeno type FIS is preferred because it can be trained without much effort.

The parameters of input and output MFs make up the set of parameter of FIS. Using the given training data, the FIS parameters are updated by an iterative least-squares estimation and back-propagation using gradient descent algorithm that minimizes the sum of the squares of the differences between the estimated VTEC and the model (predicted by FIS) output. For the detailed description of the training procedure, the reader is referred to Jang (1993) and Akyilmaz and Kutterer (2004).

In the third step, the output volatility predicted by the FIS must be converted for comparison with the original data using

$$\text{VTEC}(t+1) = \exp(r(t)) \cdot \text{VTEC}(t) \tag{6}$$

For the subsequent predictions, e.g., VTEC($t + 2$), the predicted values (VTEC($t + 1$), $r(t)$) become part of the input to the FIS.

For the quality assesment of the prediction, the RMS error and the relative error (RE) are computed as follows:

$$\text{RMS} = \sqrt{\frac{\left(\text{VTEC}_i^p - \text{VTEC}_i^o\right)^2}{N-1}} \tag{7}$$

$$\text{RE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\text{VTEC}_i^p - \text{VTEC}_i^o|}{\text{VTEC}_i^o} \times 100\% \tag{8}$$

where VTEC_i^p and VTEC_i^o are the predicted and estimated values, respectively, for the ith element in the VTEC series, and N is the total number of the predicted elements in the time series.
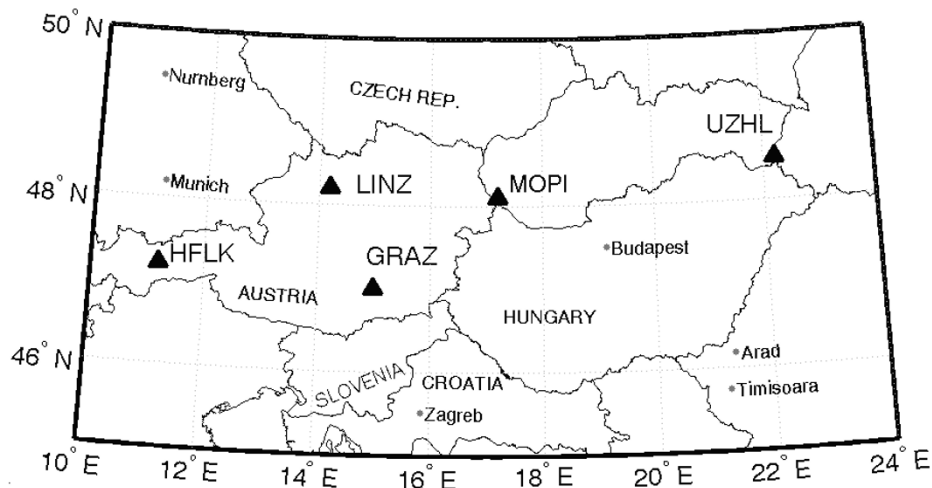
Fig. 2. Locations of permanent GPS stations.

Table 1. RMS of code and carrier-smoothed code multipath.

Station	RMS of code multipath (m)				RMS of smoothed-code multipath (m)			
	DOY 252		DOY 254		DOY 252		DOY 254	
	MP1	MP2	MP1	MP2	MP1	MP2	MP1	MP2
GRAZ	0.41	0.44	0.40	0.45	0.04	0.05	0.04	0.06
HFLK	0.28	0.33	0.27	0.32	0.06	0.08	0.04	0.05
LINZ	0.37	0.48	0.36	0.48	0.04	0.05	0.05	0.06
MOPI	0.30	0.88	0.29	0.88	0.06	0.08	0.06	0.07
UZHL	0.63	1.65	0.61	1.68	0.05	0.07	0.07	0.09

4.   Numerical Example

In this section, we apply the proposed prediction method to predict the VTEC. In the first step, the VTECs were estimated epoch by epoch per satellite-station pair using (3) at five stations located in a mid-latitude region.

Figure 2 shows the geographic location of the stations. The data was obtained from the Scripps Orbit and Permanent Array Center (SOPAC) web site maintained by University of California, San Diego. We use a 24-hour data set in the RINEX format for DOY 252 and 254, a 30-s sampling rate and a 15° elevation mask.

There is a correlation between solar events and changes in the ionosphere. Significant energy input source to the ionosphere is solar wind/radiation. A proxy on solar energetics can be derived from $F_{10.7}$ cm solar flux values and sunspot numbers. The 10.7 cm solar flux is given in solar flux unit (an sfu = 10^{-22} m^{-2} Hz^{-1}). For this study, solar flux readings are obtained at 17:00, 20:00 and 23:00 UT on DOY 252 and DOY 254 from Dominion Radio Astrophysical Observatory in Pentiction (DRAO). The official reading was taken as 99.2 sfu at 17:00 UT on DOY 252 because the other values are extremely high. The solar flux reading was 109.7 sfu at 20:00 UT (local noon) on DOY 254. 10.7 cm solar flux average is 110 sfu on September, 2005 (DRAO, 2008). The solar flux values are around the monthly mean. The international sunspot numbers are 28 for DOY 252 and 34 for DOY 254, respectively (NGDC, 2008).

Since the quality of the observations is very important for data analysis, we used UNAVCO's quality check program to check the code multipath before and after smoothing of code observations (Estey and Meertens, 1999). Table 1 shows the RMS values. The multipath RMS values for each satellite are computed and converted to the mean RMS at each station for both P codes. The threshold values for mean RMS for the P1 code multipath is 50 cm (denoted as MP1) and 65 cm for P2 multipath (denoted as MP2). We note that the multipath is almost the same for consecutive days due to repeating satellite geometry. The RMS of the smoothed codes is small as a result of the carrier phase smoothing.

There are variety of techniques to estimate DCBs. For the scope of the study, we assume that the receiver DCBs are constant in a selected time period for VTEC estimation. Estimated VTEC values were used in training steps. Table 2 shows daily variation of estimated DCBs for receivers with their RMS values. DCBs were estimated with the regional ionospheric model at the same processing run using the geometry-free linear combination equation. DCBs are assumed constant within a day. 1 ns error is equal to 1.875 TECU or 30 cm of delay for L_1. For comparison purposes, the daily DCB values for some of the receivers can be obtained from ftp://ftp.unibe.ch/ aiub/BSWUSER50/ORB/. As it can be seen from Table 2, the difference of the CODE solution and estimated DCB are between 1–1.5 ns at GRAZ and HFLK stations. This discrepancies are depended on the selected parameters in the software panel and the number of the stations used for estimation.

The PRNs 6, 8, 9, 23, 27, 28 and 30 were selected on September 9 and 11, 2005 (DOY 252 and 254). The volatil-

Table 2. DCBs and corresponding RMS values at selected stations on DOY 252 and 254 [ns].

Station	DOY 252				DOY 254			
	Estimated DCB	RMS	CODE DCB	RMS	Estimated DCB	RMS	CODE DCB	RMS
GRAZ	21.324	0.041	22.410	0.029	20.787	0.041	22.281	0.034
HFLK	20.092	0.041	21.261	0.029	19.380	0.041	20.881	0.034
LINZ	18.967	0.041	—	—	18.540	0.041	—	—
MOPI	−12.670	0.041	—	—	−12.765	0.041	—	—
UZHL	2.786	0.041	—	—	2.524	0.041	—	—

Table 3. RMS error statistics for 5, 10 and 15 min VTEC predictions at each station.

Interval		RMS (TECU)						Mean relative error (%)					
		5 min		10 min		15 min		5 min		10 min		15 min	
DOY		252	254	252	254	252	254	252	254	252	254	252	254
Station	GRAZ	0.15	0.15	0.23	0.20	0.31	0.32	1.5	1.6	2.4	2.4	2.7	2.8
	HFLK	0.08	0.13	0.14	0.22	0.24	0.32	1.0	1.5	1.4	1.8	1.8	2.1
	LINZ	0.07	0.08	0.14	0.17	0.20	0.31	0.9	1.3	1.3	1.4	1.7	1.9
	MOPI	0.07	0.08	0.13	0.15	0.18	0.21	0.9	1.0	1.4	1.5	1.7	1.8
	UZHL	0.14	0.11	0.21	0.19	0.30	0.27	1.7	1.6	2.1	1.9	2.3	2.2
	Mean	0.10	0.11	0.17	0.19	0.25	0.29	1.2	1.4	1.7	1.8	2.0	2.2

ity series of VTEC were computed for each satellite-station pair by (4). Then the training data was composed by using first 19 epochs of volatility series of VTEC which gives the best FIS parameters in each satellite arc according to the systematic given in (5). This equation shows the training pattern, not the training data. Training pattern indicates how the training data matrix is composed. In this study, 19 epochs of volatility (computed from the first 20 epochs of VTEC data) training data is found to be sufficient for a high-quality prediction. After the training pattern has been determined, each input variable is represented by two MFs of Gaussian type while the output MFs are linear functions of the input variables. An example of an FIS with two fuzzy rules are in the following form.

Rule 1: If x is A_1 and y is B_1; then $f_1 = p_1 x + q_1 y + r_1$.

Rule 2: If x is A_2 and y is B_2; then $f_2 = p_2 x + q_2 y + r_2$.

The parameters of the input fuzzy sets (membership functions) A_i, B_j ($i = 1, 2$; $j = 1, 2$) are called the premise parameters whereas the parameters of output functions f_i ($i = 1, 2$), i.e., p_i, q_i and r_i are called consequent parameters. In this study, the input membership function $\mu_{A_i}(x)$ is chosen to be the Gaussian function (Eq. (9)) with the maximum value equal to 1 and the minimum value equal to 0,

$$\mu_{A_i}(x) = \exp\left[-\left(\frac{x - c_i}{a_i}\right)^2\right] \tag{9}$$

where $\{a_i, c_i\}$ is the parameter set. As the values of these parameters change, the shape of the MFs vary accordingly.

The composed FIS was trained in the last stage of the process. The training procedure takes less than a second for the selected training patterns using a Pentium III PC. ANFIS editor of MATLAB v 6.5 is just used for the training. In addition, a software was developed as m-functions in order to make computations in MATLAB v 6.5 for predictions. The FIS parameters are available after the training step. Further predictions of the VTEC in the time series can
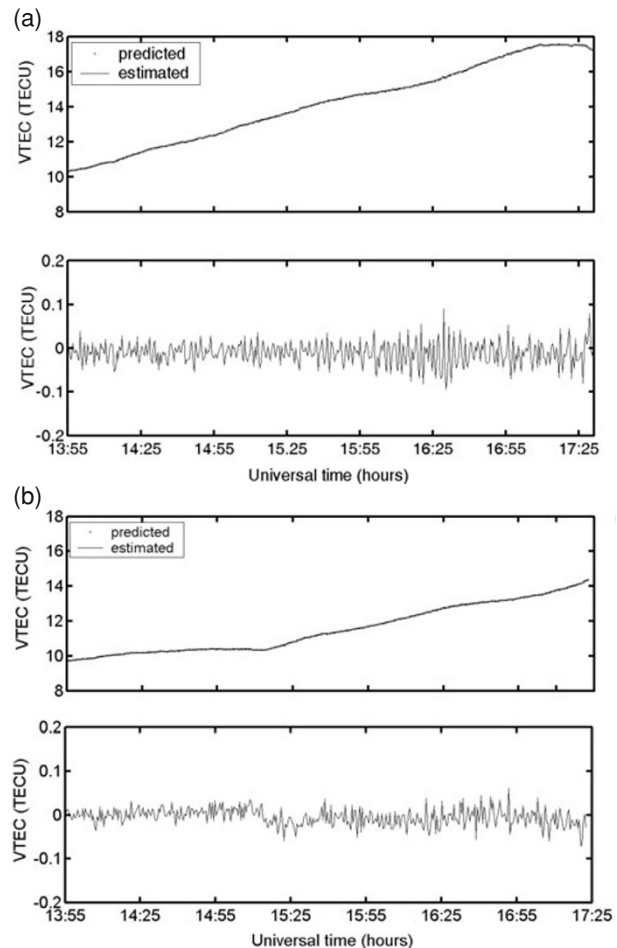

Fig. 3. Estimated and predicted VTEC for PRN 9 (top), prediction error (bottom) at GRAZ station on (a) DOY 252 (RMS = 0.025 TECU, RE = 0.16%) (b) DOY 254 (RMS = 0.018 TECU, RE = 0.13%).

be made using this parameter set. Consecutive predictions are computed using the predicted values as input. Recall that the output of the FIS are the volatilities. Therefore, the predicted VTEC is calculated by (6).
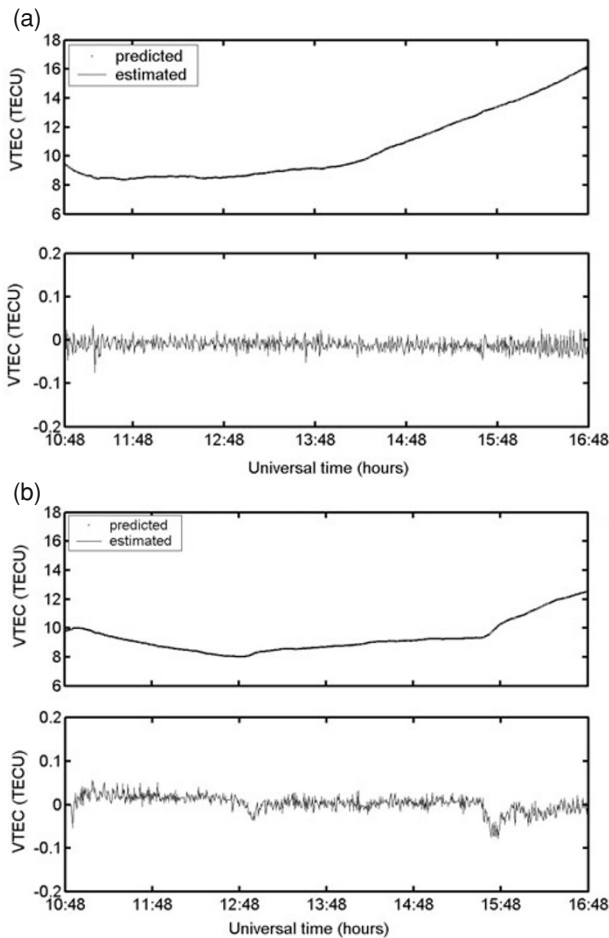
Fig. 4. Estimated and predicted VTEC for PRN 9 (top), prediction error (bottom) at HFLK station on (a) DOY 252 (RMS = 0.020 TECU, RE = 0.14%) (b) DOY 254 (RMS = 0.021 TECU, RE = 0.16%).
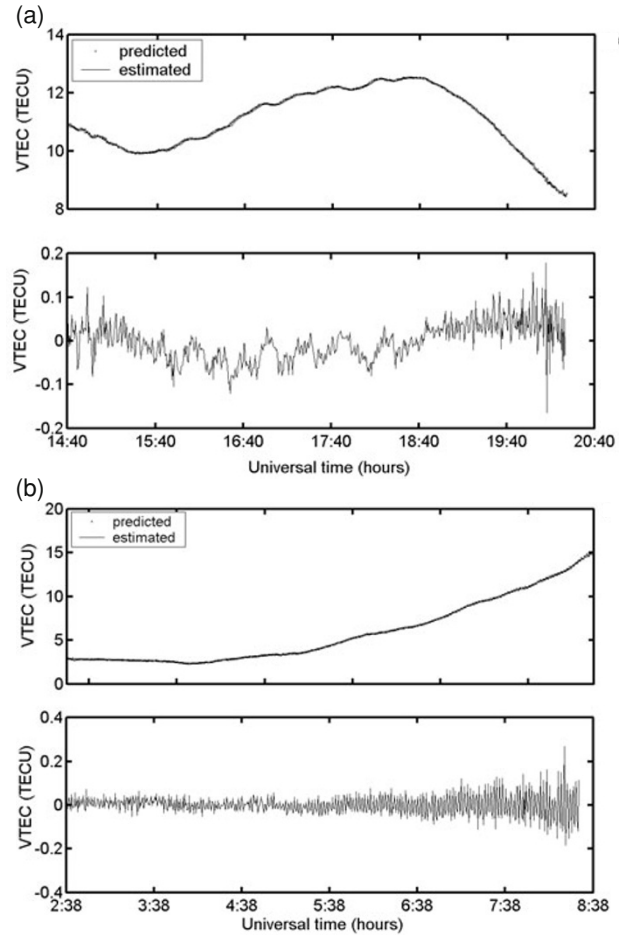
Fig. 5. Estimated and predicted VTEC (top), prediction error (bottom) at UZHL station for (a) PRN 30 and DOY 252 (RMS = 0.043 TECU, RE = 0.31%) (b) PRN 23 and DOY 254 (RMS = 0.045 TECU, RE = 0.61%).

The training patterns and the FIS parameters are kept fixed afterwards for 5, 10 and 15 min predictions. We make predictions at every 30 s successively following the training data. However, we use the predicted value of the 5, 10 and 15th min data point in order to evaluate accuracy information using (7) and (8). Predictions start at 10.5 min (following the data used for training) and after the 15th min prediction is achieved the starting point is shifted along the time axis by one epoch (30 s) for successive predictions, i.e., the next prediction starts at 11th min until 25.5th min. About 200 predictions were made starting from the randomly chosen data points at each station-satellite pair of VTEC time series.

In order to show statistics of VTEC predictions, the RMS errors and the mean relative errors of the VTEC for all observed satellites in each station are summarized in Table 3. The mean RMS values for 5, 10 and 15 min predictions obtained from the five stations are given in the last row. In this table, the VTEC prediction errors at all the stations for 5, 10 and 15 min prediction intervals are less than 0.12, 0.20 and 0.30 TECU for DOY 252 and 254, respectively. The prediction RMS errors have very similar behaviour for the same stations for the selected prediction intervals. As it is expected, the short-term predictions have been found to be more accurate than the long-term predictions. This is be-

cause the predicted values were used as inputs for further predictions and hence the error propagation decreased the prediction quality.

For 5 min prediction intervals, the mean RE has been found to be 1.2% and 1.4% which means that 98.8% and 98.6% of the VTEC was successfully recovered by the prediction technique for DOY 252 and 254, respectively. In addition, the mean RE for 10 min intervals are at the level of 1.7% and 1.8%, and for 15 min intervals are about 2.0% and 2.2% on two succesive days.

In addition to the given statistics in Table 3, Figs. 3–5 are given as the selected examples of the FIS prediction from the analyzed VTEC arc to visualize the prediction procedure. Each figure represents the estimated and the predicted VTEC values on top and the respective prediction errors in bottom frames. The X-axis shows the Universal Time (UT) while the Y-axis represents the VTEC values in TECU for the selected satellite-station pair.

Note that for each sample, the prediction error increases with time. For example, in Fig. 5(b) the amplitude of the prediction error at 7:50 UT is almost three times larger than the one at 4:38 UT. This is mainly due to the FIS predictor, because the volatility index computed from 7:50 UT are not covered by the volatility values in the training data set. This

situation can be regarded as one of the weakness of the proposed method. The input data range is an important parameter in FIS predictions. However, the mentioned method gave reasonable predictions in the scope of the study. All the statistics in Table 3 are very consistent within the selected prediction intervals. This case can be related to the similarity of the ionospheric variations in the local area, especially in the mid-latitude region. For example, the TEC values for PRN 9 at GRAZ (Fig. 3(a)) and HFLK (Fig. 4(a)) are about 10 TECU at 13:55 UT for DOY 252. The differences between the estimated and predicted TEC in Fig. 5(a) are more disturbed with respect to the other time series of TEC which was given in the figures. This can be related with the periodic variation in VTEC in the given time period.

5. Conclusions

The vertical TEC is predicted that allows real-time GPS applications without concerns about latency, using Takagi-Sugeno type fuzzy inference systems. The input parameter for the prediction is VTEC, which can be easily computed from carrier smoothed-code observations. A mean relative prediction error of about 2% shows the efficiency of the proposed prediction method. The mean RMS errors for 5, 10 and 15 min intervals are about 0.1, 0.2 and 0.3 TECU for the selected days, respectively. When the prediction time is increased by a factor of two or three, the RMS values also increases by two or three. This demonstrates a linear relation between prediction times and RMS, which is normal due to the nature of fuzzy prediction methods. Therefore, the RMS error increases linearly by the prediction time. The mean RMS errors from different stations are almost identical. This result is expected because the ionospheric changes in a local area have similar behaviour.

It is important to select the prediction time interval carefully because the prediction error increases with prediction range. Positioning in real-time, including single-frequency applications, can be supported up to 15 min with sufficient accuracy using predicted VTEC with the methodology discussed.

References

Akyilmaz, O. and H. Kutterer, Prediction of earth rotation parameters by fuzzy inference systems, *J. Geod.*, **78**, 82–93, 2004.
Ciraolo, L., F. Azpilicueta, C. Brunini, A. Meza, and S. M. Radicella, Calibration errors on experimental slant total electron content (TEC) determined with GPS, *J. Geod.*, **81**,111–120, 2007.
Dach, R., U. Hugentobler, P. Fridez, and M. Meindl, *User manual of the Bernese GPS software version 5.0*, Astronomical Institute, University of Bern, 2007.
DRAO (Dominion Radio Astrophysical Observatory in Pentiction), http://www.drao-ofr.hia-iha.nrc-cnrc.gc.ca/icarus/www/daily.html, 2008.
Estey, L. H. and C. M. Meertens, TEQC: The multi-purpose toolkit for GPS/GLONASS data, *GPS Solutions*, **3**(1), 42–49, 1999.
Jang, J. S. R., ANFIS: adaptive-network-based fuzzy inference systems, *IEEE Trans. Syst. Man. Cybernet*, **23**, 665–685, 1993.
Liu, Z. and Y. Gao, Ionospheric TEC predictions over a local area GPS reference network, *GPS Solutions*, **8**, 23–29, 2004.
Liu, R., Z. Xu, J. Wu, S. Liu, B. Zang, and G. Wang, Preliminary studies on ionospheric forecasting in China and its surrounding area, *J. Atmos. Solar-Terr. Phys.*, **67**, 1129–1136, 2005.
NGDC (National Geophysical Data Center), http://sgd.ngdc.noaa.gov/sgd/jsp/solarindex.jsp, 2008.
Oyeyemi, E. O., L. A. Mckinnell, and A. W. V. Poole, Near-real time foF2 predictions using neural networks, *J. Atmos. Solar-Terr. Phys.*, **68**, 1807–1818, 2006.
Popoola, A., S. Ahmad, and K. Ahmad, A Fuzzy-Wavelet method for analyzing non-stationary time series, in *Proceedings of the 5th Int. Conference on Recent Advances in Soft Computing (RASC 2004)*, pp. 231–236, December 16–18, Notthingham, UK, 2004.
Rideout, W. and A. Coster, Automated GPS processing for global total electron content, *GPS Solutions*, **10**, 219–228, 2006.
Sardon, E. and N. Zarraoa, Estimation of total electron content using GPS data: How stable are the differential satellite and receiver instrumental biases?, *Radio Sci.*, **32**(5), 1899–1910, 1997.
Takagi, T. and M. Sugeno, Fuzzy identification of systems and its application to modeling and control, *IEEE Trans. Syst. Man. Cybernet*, **15**, 116–132, 1985.
Wild, U., Ionosphere and geodetic satellite systems: Permanent GPS tracking data for modelling and monitoring, *Achtundvierzigster Band*, **48**, 1994.

O. Akyilmaz and N. Arslan (e-mail: narslan@yildiz.edu.tr)