VOLUME 5 ISSUE 2 JUNE 2020

IJEG

International Journal of Engineering and Geosciences

e-ISSN 2548-0960

The International Journal of Engineering and Geosciences (IJEG)

The International Journal of Engineering and Geosciences (IJEG) is a tri-annually published journal. The journal includes a wide scope of information on scientific and technical advances in the geomatics sciences. The International Journal of Engineering and Geosciences aims to publish pure and applied research in geomatics engineering and technologies. IJEG is a double peer-reviewed (blind) OPEN ACCESS JOURNAL that publishes professional level research articles and subject reviews exclusively in English. It allows authors to submit articles online and track his or her progress via its web interface. All manuscripts will undergo a refereeing process; acceptance for publication is based on at least two positive reviews. The journal publishes research and review papers, professional communication, and technical notes. IJEG does not charge for any article submissions or for processing.

CORRESPONDENCE ADDRESS
Journal Contact: engineeringandgeoscience@gmail.com

CONTENTS

Volume 5 - Issue 2

ARTICLES

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 060-065, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.573503

COMPARATIVE ANALYSIS OF THE PYSEBAL MODEL AND LYSIMETER FOR ESTIMATING ACTUAL EVAPOTRANSPIRATION OF SOYBEAN CROP IN ADANA, TURKEY

Alidou Sawadogo [1]*, Hessels Tim [2], Kemal Sulhi Gündoğdu[1], Ali Osman Demir [1], Mustafa Ünlü [3], Sander Jaap. Zwart [4]

[1] Uludag University, Faculty of Agriculture, Biosystems Engineering Department, Bursa-Turkey. (sawadogoalidou@yahoo.fr; kemalg@uludag.edu.tr; aodemir@uludag.edu.tr); **ORCID 0000-0002-7437-8415, ORCID 0000-0002-5591-4788, ORCID 0000-0003-3409-6680**

[2] UNESCO-IHE Institute for Water Education, Delft, The Netherland. (timhessels@hotmail.com); **ORCID 0000-0002-7264-7220**

[3] University of Çukurova, Faculty of Agriculture, Department of Agricultural Structures and Irrigation, Adana - Turkey. (munlu@cu.edu.tr); **ORCID 0000-0002-1889-516X**

[4] University of Twente, Faculty of Geo-Information Science and Earth Observation, Department of Natural Resources. (s.zwart@cgiar.org); **ORCID 0000-0002-5091-1801**

*Corresponding Author, Received: 02/06/2019, Accepted: 18/09/2019

ABSTRACT: Accurate estimation of evapotranspiration (ET) is an important factor in water management, especially in irrigated agriculture. Accurate irrigation scheduling requires accurate estimation of ET. The objective of this study was to estimate the actual evapotranspiration (ET_a) by the pySEBAL model and to compare it with the actual evapotranspiration measured by the lysimeter method of soybean crop in Adana, Turkey. Five Landsat 5 Thematic Mapper (TM) images and weather data were used for this study to estimate actual evapotranspiration by the pySEBAL model. The results showed a good relationship between ET_a estimated by the pySEBAL model and ET_a measured by the lysimeter method, with an R^2 of 0.73, an RMSE of 0.51 mm.day^{-1}, an MBE of 0.04 mm.day^{-1} and a Willmott's index of agreement (*d*) of 0.90. Based on this study, there is a good relationship between the actual evapotranspiration estimated by the pySEBAL model and the actual evapotranspiration measured by the lysimeter method. Consequently, ET_a of soybean crop can be estimated with high accuracy by the pySEBAL model in Adana, Turkey.

Keywords: *Actual Evapotranspiration, Lysimeter, PySEBAL Model, Remote Sensing.*

1. INTRODUCTION

Freshwater resources are becoming increasingly scarce in many parts of the world (Anonymous, 2007) and agriculture is the largest water user worldwide (FAO, 2011). Water use in agriculture is one of the main causes of the water scarcity, and according to the FAO (2009), the world's population is predicted to increase to 9.1 billion people by 2050. Face to this reality, there is a need to increase agricultural water productivity in a sustainability way to meet rising demands for water and food. Water productivity can be improved by increasing yield using less water. This could be done by proper irrigation management practices. In this process, more water could be save to meet the rising demand for water. To improve water productivity by proper irrigation management, a precise knowledge of evapotranspiration is required. ET can be estimated by many methods such as direct measurements using the lysimeters (Allen et al., 1998) and indirect measurements using remote sensing-based methods (Bastiaanssen, 1995).

The remote sensing model, such as the Surface Energy Balance Algorithm for Land (SEBAL) model developed by Bastiaanssen et al. (1998), is used to estimate the ET_a (Bastiaanssen et al., 2005). Other energy balance models such as the Surface Energy Balance System (SEBS) developped by Su (2002), the Simplified Surface Energy Balance Index (S-SEBI) developed by Roerink *et al.* (2000) and most recently, the Mapping Evapotranspiration at high Resolution with Internalized Calibration (METRIC) algorithm developed Allen et al. (2007) to estimate the crop water consumption. SEBAL has been validated in many countries under a wide variety of conditions (Bastiaanssen et al., 2005) and has been identified as the most promising approach currently available to estimate evapotranspiration. The Python module for Surface Energy Balance Algorithm for Land (pySEBAL) model was developed by the IHE-Delft Institute for Water Education. Lysimeters, both weighing and drainage types, are considered as basic method to calibrate evapotranspiration models (Mata et al., 1999; Centinari et al., 2009). Weighing lysimeters are considered as standard for ET measurements. However, due to their high prices, these units are few in number at any given location.

To estimate crop water use for irrigation management, remote sensing-based methods may be more appropriate than lysimeters. One of the main advantages of remote sensing-based methods is their ability to provide the spatial variability of the ET under a different range of scales. There is a general consensus that the SEBAL model provide reliable measure of ET_a. However, the model need to be validated for a given area. The accuracy of the evapotranspiration determined from SEBAL model can be assessed from lysimeters data. In this study, the pySEBAL model was applied to estimate actual evapotranspiration of soybean crop in Adana, Turkey. The estimated ET_a from the pySEBAL model was compared with the ET_a measured by the lysimeter method and the accuracy of the pySEBAL model to estimate ET_a of soybean crop in Adana, Turkey was discussed.

2. MATERIALS AND METHODS

2.1. Lysimeter Data Collection

In this study, secondary data of actual evapotranspiration measured by lysimeter is used (Table.1). Actual evapotranspiration was directly measured by a weighing lysimeter in the center of 0.12 ha field at Research Fields of the Agricultural Structures and Irrigation Department, Faculty of Agriculture, University of the Cukurova (37° 1' N, 35° 21' E, and 20 m above sea level) Adana, Turkey. Turkey is between the 36° and 42° north latitudes and 26° and 45° south meridians (Iscan and Ilgaz, 2017; Ernst et al 2019). Lysimeter was used to measure well-watered soybean (*Glycine max* L.) evapotranspiration growing in 2009 season. Arioglu soybean cultivar was used in this study. Soybean crop have been sown on 175 Day of year (DOY) and inter and intra row distances of 0.70 m and 0.10 m, respectively, were applied. Irrigation was applied using a drip irrigation system (Table.1) and crop growth inside and outside of the lysimeter was homogenous. The fertilization doses of 36 $kg.ha^{-1}$ pure nitrogen and 92 $kg.ha^{-1}$ phosphor, P_2O_5, recommended in the region for soybean were applied (Uncu and Arioglu, 2005). The soil texture in the experiment plot is heavy clay, as averagely, with 15.75 % of sand, 19.58 % of silt and 64.62 % of clay, pH is slightly basic, and it is poor in organic matter. The climate of study area is typical Mediterranean climate, with cool rainy winters, and hot dry summers. Temperature varies from 9.9 °C in January to 28.1 in August.

Table.1. Irrigation intervals (day of year), depth of irrigation water (mm) and ET_a (mm)

Day of Year	Irrigation Water (mm)	lysimeter ET_a (mm)
177	64.0	2.7
197	na[1]	2.9
210	43.4	3.7
213	na	2.5
223	32.3	5.0
229	na	4.1
231	23.6	3.8
239	24.0	4.4
245	na	3.5
247	26.4	4.6
253	16.0	3.3
261	16.5	3.8
277	na	1.1

[1] na: not applicable

2.2. Meteorological Data

Hourly air temperature, wind speed, relative humidity, and solar radiation, were used in the pySEBAL model (Table 2). Weather data were obtained from the Adana meteorology station and from http://www.soda-pro.com for solar radiation.

Table.2. Meteorological data

Day of Year	Wind speed ($m.s^{-1}$)	Air temperature (^{o} C)	Relative humidity (%)	Solar radiation ($W.m^{-2}$)
197	1.1	27.5	77.7	238
213	1.2	29.2	76.3	278
229	1.2	29.0	83.3	298
245	1.2	28.9	78.7	221
277	1.3	23.3	60.8	200

Table.3. Landsat scenes information

Number	Day of Year	Landsat Scene	Acquisition day	Overpass Time(AM)
1	197	LT51750342009197MOR00	2009-07-16	10:04:32 AM
2	213	LT51750342009213MOR00	2009-08-01	10:04:47 AM
3	229	LT51750342009229MOR00	2009-08-17	10:05:01 AM
4	245	LT51750342009245MOR00	2009-09-02	10:05:16 AM
5	277	LT51750342009277MOR00	2009-10-04	10:05:40 AM

2.3. Landsat Images

Due to the lack of Landsat free cloud images in the period of interest, five Landsat 5 TM clear-sky images were used in this study. Landsat images were downloaded from the Earth Explorer (https://earthexplorer.usgs.gov/) website. Table 3 shows the scene information used in this study.

2.4. Surface Energy Balance Algorithm for Land (SEBAL)

The SEBAL model is based on modelling the surface energy balance using remote sensing data. The pySEBAL model was developed by IHE-Delft Institute for Water Education in Python programming language (https://pypi.org/project/SEBAL/). A brief summary of the main algorithms of the pySEBAL model is given below. The model calculates the latent heat flux (λE, $W.m^{-2}$) as the residual of surface energy equation:

$$\lambda E = R_n - G - H \quad (1)$$

where R_n is the net radiation ($W.m^{-2}$), G is the soil heat flux ($W.m^{-2}$) and H is the sensible heat flux (Wm^{-2}).

Net radiation (R_n): R_n is computed using the following equation:

$$R_n = (1 - \alpha)\, Rs\downarrow + RL\downarrow - RL\uparrow - (1 - \varepsilon_0)\, RL\downarrow \quad (2)$$

where α is the surface albedo, $Rs\downarrow$ is the incoming shortwave radiation ($W.m^{-2}$), $RL\downarrow$ is the incoming longwave radiation ($W.m^{-2}$), $RL\uparrow$ is outgoing longwave radiation ($W.m^{-2}$), and ε_0 is surface emissivity.

Soil heat flux (G): The following equation is used to calculate G.

$$G = R_n(T_s\,(0.0038 + 0.0074\,\alpha)(1 - 0.978 \times NDVI^4)) \quad (3)$$

where T_s is the surface temperature (K) and NDVI is the normalized difference vegetation index.

Sensible heat flux (H): H is estimated using the heat transfer equation:

$$H = \frac{\rho \times c\rho \times dT}{rah} \quad (4)$$

where ρ is the air density ($kg.m^{-3}$) $c\rho$ is the specific heat of air at constant pressure ($J.kg^{-1}.K^{-1}$), dT is the vertical near surface temperature difference (K) and rah is the aerodynamic resistance to heat transport ($s.m^{-1}$). H is calculated in an iterative way due to the relationship between the aerodynamic resistance and sensible heat flux. Within the first iteration, a neutral air condition without convection is assumed. This first iteration will result in a sensible heat flux, and therefore, the air will become unstable, which will change the aerodynamic resistance. This will provide the input for the next iteration step. This iteration is performed multiple times to find the final sensible heat flux.

Evaporative Fraction (EF): The instantaneous EF is calculated using the following equation:

$$\mathbf{EF_{inst}} = \frac{\boldsymbol{\lambda E}}{\boldsymbol{\lambda E + H}} \quad (5)$$

The instantaneous evaporative fraction can be used to calculate the daily ET. The daily evaporation is calculated using the following equation:

$$ET_{24} = EF_{inst}\ x\ AF\ x\ \frac{R_{n24}}{\lambda\ x\ \rho_w}\ x\ 86400000 \quad (6)$$

where AF is the advection factor, λ is the latent heat of vaporization ($J.kg^{-1}$) and ρw is the density of water ($kg.m^{-3}$). The advection factor is used to account for any effects

of regional advection and is calculated by the following equation:

$$AF = 1 + 0.985(\exp((e_{sat\_24} - e_{act\_24}) \times 0.08) - 1)\, EF_{inst} \quad (7)$$

where e_{sat_24} is the 24-hour averaged saturated vapor pressure (kPa) and e_{act_24} is the 24-hour averaged actual vapor pressure (kPa).

2.5. Statistical Analysis

Statistical comparison between actual evapotranspiration obtained by the pySEBAL model and by the lysimeter method was done using a simple linear regression. Others statistical evaluations, such as the root mean square error (RMSE), the coefficient of determination (R^2), the Willmott's index of agreement (d) and the mean bias error (MBE) were used to determine the relationships between ET_a obtained from the pySEBAL model and from the lysimeter method.

3. RESULTS AND DISCUSSIONS

3.1. PySEBAL Model Based ET_a Maps

Fig. 1 and Fig. 2 shows maps of daily actual evapotranspiration around the area of interest. The white small square in Fig. 1 and Fig. 2 shows the pixel selected for ET_a data collection. The selected pixel shows the lysimeter field. The Day of Year (DOY) 197 and the DOY 245 of the Fig. 1 and Fig. 2, respectively, shows the variation of ET_a by the pySEBAL model throughout the 2009 growing season. The ET_a was 2.3 mm.day^{-1} and 3.1 mm.day^{-1} for the DOY 197 and the DOY 245, respectively. The increased values of ET_a from the DOY 197 to the DOY 245 are attributed to the crop growth and meteorological conditions.

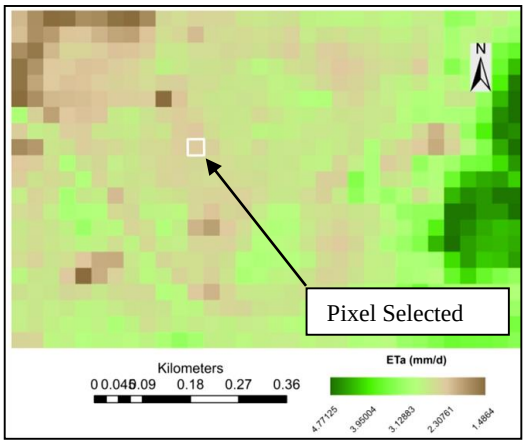


Fig.1. ETa map of the DOY 197

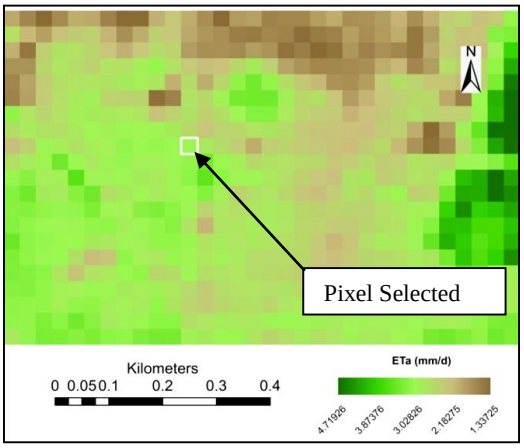


Fig.2. ETa map of the DOY 245

3.2. Comparison of ET_a by the PySEBAL Model with ET_a by Lysimeter Method

A simple linear relationship between ET_a from the pySEBAL and ET_a from the lysimeter method is shown in Fig. 3. The Statistical comparison between ET_a from the pySEBAL model and ET_a from the lysimeter method is shown in Table 4. The results revealed that the RMSE was 0.51 mm.day^{-1}, the R^2 was 0.73, the MBE was 0.04 and, the d was 0.9. The results of the simple linear regression showed a good relationship between ET_a by the pySEBAL model and ET_a by the lysimeter method and have been confirmed by many researchers. Similar results in R^2 (0.76) was reported by Ruhoff et al. (2012) by comparing ET_a estimated by the SEBAL model with ET_a measured by the eddy-covariance method for sugarcane croplands. Zamansani et al. (2018) also found a good coefficient of determination of 0.77 by comparing actual ET from the SEBAL model with actual ET from the lysimeter method for wheat crops. Based on the R^2 in this study, ET_a by the pySEBAL model and ET_a by the lysimeter method showed a good accuracy between the two data sets. For the RMSE found in this study, Bala et al. (2015) reported similar results in RMSE of 0.51 mm.d^{-1} by comparing ET_a estimated by the SEBAL model and ET_a measured by the lysimeter method. In addition, Hassanpour et al. (2011) by comparing ET_a estimated by the SEBAL model and ET_a measured by the lysimeter method found an RMSE of 0.29 mm.day^{-1}. Bhattarai et al. (2011) by comparing ET_a estimated by the SEBAL model and ET_a measured by the eddy covariance method on grass crop, reported a MBE of 0.05 mm.day^{-1}, and then confirm the MBE of 0.04 mm.day^{-1} found in this study. The Willmott's index of agreement of 0.9 and the lower values of RMSE and MBE indicates better performance of the pySEBAL model to estimate the ET_a.

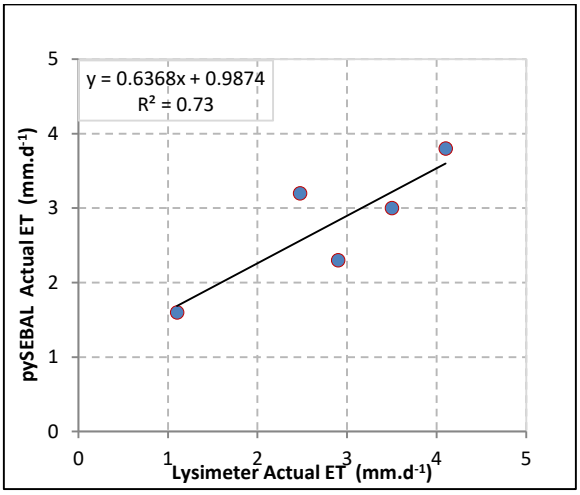


Fig. 3. The relationship between ETa from lysimeter and ETafrom pySEBAL.

Table.4. Statistic results of ET_a from pySEBAL and ET_a from lysimeter

Statistic parameters	Actual evapotranspiration
RMSE ($mm.day^{-1}$)	0.51
MBE ($mm.day^{-1}$)	0.04
R^2	0.73
d	0.90

RMSE = Root Mean Square Error
MBE = Mean Bias Error
R^2 = Coefficient of determination
d = Willmott index of agreement

The capability of the SEBAL model to estimate actual ET with R^2 close to one is shown by many researchers. Bala et al. (2015) found an R^2 of 0.91 by comparing ET_a estimated by the SEBAL model and ET_a measured by the lysimeter method. The estimation of ET_a by the SEBAL model as a residual energy of energy balance equation should therefore be as accurate as possible in R_n, H and G estimation. Therefore, the accuracy of SEBAL model can be affected by many errors related in R_n, H and G estimation. Bastiaanseen et al. (2005) shown that the typical accuracy of SEBAL is 85 % at field scale on a daily basis. One of the important source of error at field scale in ET_a estimation by the SEBAL model is the advection effect. Mkhwanazi and Chávez (2013) by comparing the SEBAL model under advective and non-advective conditions found and error ranged between 5 and 46%. Singh et al. (2008) observed an error of 28% for the SEBAL model due to advection effect. At regional scale, pySEBAL model include an advection factor to account for any effects of regional advection (Eq. [7]). Unfortunately, the minimum scale at which advection will have to be considered remain a research challenge. In this study, the small experimental area of 0.12 ha could be prone to decrease the accuracy of the pySEBAL model due to advection effect. The validation of remote sensing data with ground truth observation is the dissimilarity between the spatial scales of field and satellite data (Orhan et al., 2019). With an R^2 of 0.73, an RMSE of 0.51 $mm.day^{-1}$, an MBE of 0.04 $mm.day^{-1}$, and an *d* of 0.90, indicates a good accuracy of the pySEBAL model in daily ET_a estimation of soybean crop in Adana, Turkey.

4. CONCLUSION

We have evaluated the pySEBAL model in Adana, Turkey by comparing ET_a estimated by the pySEBAL model with ET_a measued by lysimeter method. In this study, Landsat 5 Thematic Mapper images were used to estimate the ET_a by pySEBAL model. The results showed a good relationship between ET_a estimated by the pySEBAL model and ET_a measured by the lysimeter method, with an R^2 of 0.73, an RMSE of 0.51 $mm.day^{-1}$, an MBE of 0.04 $mm.day^{-1}$ and a Willmott's index of agreement (*d*) of 0.90. Based on the results of this study, ET_a of soybean crops can be estimated with high accuracy by pySEBAL model in Adana, Turkey. Success of pySEBAL model ET_a estimation depends on the availability of Landsat cloud-free images. In this study, the availability of Landsat cloud-free images was limited. Thus, the application of pySEBAL model with more satellite images is recommended.

ACKNOWLEDGMENT

This study was possible thanks to the support of the Presidency for Turks Abroad and Related Communities (YTB).

REFERENCES

Allen, R. G., Pereira, L.S., Raes, D., and Smith, M. 1998. "Crop evapotranspiration: Guidelines for computing crop water requirements." FAO irrigation and drainage paper 56, Food and Agriculture Organization of the United Nations, Rome. p. 1-209.

Allen, R. G., Tasumi, M., and Trezza, R. 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)-Model. Journal of Irrigation and Drainage Engineering, 133, 380-394.

Anonymous. 2007. Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture. London: Earthscan, and Colombo: International Water Management Institute.

Bala, A., Rawat, K. S., Misra, A. K., and Srivastava, A. 2015. Assessment and validation of evapotranspiration using SEBAL algorithm and Lysimeter data of IARI agricultural farm, India. GEOCARTO INT. DOI: 10.1080/10106049.2015.1076062.

Bastiaanssen, W. G. M. 1995. Regionalization of surface flux densities and moisture indicators in composite terrain: a remote sensing approach under clear skies in Mediterranean climates. Ph.D. dissertation, CIP Data Koninklijke Bibliotheek, Den Haag, The Netherlands. 273 pp.

Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslag, A. A. M. 1998. A remote sensing surface energy

balance algorithm for land (SEBAL) . 1. Formulation. J HYDROL. 212(1-4), 198-212.

Bastiaanssen, W. G. M., Noordman, E. J. M., Pelgrum, H., Davids, G., Thoreson, B. P., and Allen, R. G.2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. J IRRIG DRAIN ENG. 131, 85-93.

Bhattarai, N., Dougherty, M., Marzen, J. L., and Kalin, L. 2011. Validation of evaporation estimates from a modified surface energy balance algorithm for land (SEBAL) model in the south-eastern United States. REMOTE SENS LETT. 3: 6, 511–519.

Centinari, M., Poni, S., Filippetti, I., Magnanini, E., and Intrieri, C. 2009. Evaluation of an open portable chamber system for measuring cover crop water use in a vineyard and comparison with a mini-lysimeter approach. AGR FOREST METEOROL. 149: 11, 1975-1982.

Ernst , F., Erdoğan, S., Bayram, Y. 2019. Human resource management using geographic information systems (gis): an example from Turkish land registry directorates. International Journal of Engineering and Geosciences (IJEG), Vol; 2; Issue; 02, pp. 41-51.

FAO. 2009. How to feed the world in 2050. Issue brief from the High-Level Expert Forum held in Rome, 12-13 October. FAO, Rome, Italy.

FAO. 2011. The state of the world's land and water resources for food and agriculture (SOLAW)-managing systems at risk. Food and Agriculture Organization of the United Nations, Rome and Earthscan, London.

Hassanpour, B., Mirzaie, F., Arshad, S., and Bakhtiari, B. 2011: Comparison of SEBAL based evapotranspiration with lysimeter data. ICID J. 21st International Congress on Irrigation and Drainage.

Iscan, F., Ilgaz, A., 2017. Analysis of Geographic/Urban Information System Web Presentations of Local Goverments In Turkey, International Journal of Engineering and Geosciences (IJEG), Vol; 2; Issue; 03, pp. 75-83.

Mata, M., Girona, J., Goldhammer, D., Fereres, E., Cohen, M., and Johnson, S. 1999. Water relations of lysimeter-grown peach trees are sensitive to deficit irrigation. CALIF AGR. 17-21.

Mkhwanazi, M. M., and Chávez, J. L. 2013. Mapping evapotranspiration with the remote sensing ET algorithms METRIC and SEBAL under advective and non-advective conditions: accuracy determination with weighing lysimeters. Hydrology Days. Colorado State University.

Orhan, O., Dadaser-Celik, F., Ekercin, S. 2019. Investigating Land Surface Temperature Changes Using Landsat-5 Data And Real-Time Infrared Thermometer Measurements At Konya Closed Basin In Turkey, International Journal of Engineering and Geosciences (IJEG), Vol; 4, Issue; 1, pp. 016-027.

Roerink, G.J., Su, Z., Menenti, M. 2000. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. Phys. Chem. Earth, Part B Hydrol. Oceans Atmos. 25 (2), 147-157. https://doi.org/10.1016/S1464-1909(99)00128-8.

Ruhoff, A., Paz, A. R., Collischonn, W., Aragao, L. E. O. C., Rocha, H. R., Yadvinder, S., and Malhi, Y. S. 2012. A MODIS-Based Energy Balance to Estimate Evapotranspiration for Clear-Sky Days in Brazilian Tropical Savannas. REMOTE SENS-BASEL. 4(3) 703-725. doi:10.3390/rs4030703.

Singh, R. K., Irmak, A., Irmak, S., and Martin, D. L. 2008: Application of SEBAL Model for mapping evapotranspiration and estimating surface energy fluxes in South-Central Nebraska. J IRRIG DRAIN E-ASCE. 134 (3): 273-285.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85-99.

Uncu, H. A., and Arioglu, H.2005. Ikinci urun soya tariminda farklı ekim zamanlarina gore bazi buyume duzenleyicilerin verim ve kalite uzerine etkileri, Turkiye VI. Tarla Bitkileri Kongresi, 5-9 Eylul, Antalya, 1: 375-380.

Zamansani, E., Khoorani, A., Sadeghi-e-lari, A., and Sadidi, J. 2018. Evaluation of evapotranspiration of wheat using SEBAL algorithm (Case study: Agricultural Research Station of Haji Abad). University of Tehran.

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 066-072, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.589489

INTEGRATION OF CUSTOM STREET VIEW AND LOW COST MOTION SENSOR

Tolga Bakirman [1]*, Mustafa Umit Gumusay [2]

[1]Istanbul Technical University, Center for Satellite Communications and Remote Sensing, Istanbul, Turkey (tolga@cscrs.itu.edu.tr); **ORCID 0000-0001-7828-9666, ORCID 0000-0001-6464-919X**
[2]Yildiz Technical University, Civil Engineering Faculty, Department of Geomatic Engineering, Istanbul, Turkey (gumusay@yildiz.edu.tr); **ORCID 0000-0001-6464-919X**

*Corresponding Author, Received: 09/07/2019, Accepted: 29/08/2019

ABSTRACT: Virtual reality is an artificial computer-generated environment generally referred as virtual reality environment which can be navigated and interacted with by a user. Street View, which was released by Google in 2007, is an ideal tool to discover places and locations. This service doesn't only provide spatial information, but also a virtual reality environment for the user. Since this service is only available in certain locations, Google enables users to create a street view with custom panoramic images with the help of Google Maps Application Programming Interface (API) for JavaScript. In this study, it is aimed to integrate body motions with a custom created street view service for Yildiz Technical University Davutpasa Campus which has a historical environment and huge places to discover. Microsoft Kinect for Xbox 360 motion sensor along with Flexible Action and Articulated Skeleton Toolkit (FAAST) interface has been employed for this purpose. This integration provides a low-cost alternative for virtual reality experience. The proposed system can be implemented for virtual museums, heritage sites or planetariums consisting of panoramic images.

Keywords: *Street view, Kinect, Virtual reality, Body motion control, Panorama*

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 066-072, June, 2020,

1. INTRODUCTION

Virtual reality is a virtual environment that can be navigated and interacted e.g. moving around and exploring the scene or having the ability to select and manipulate objects (Gutierrez et al., 2008). As photogrammetric technologies advance, researchers are able to represent reality more accurately. Researchers and scientists are able to generate affordable and satisfying walk through representations of complex facilities by using panoramic images (Chapman & Deacon, 1998; Le Yaouanc et al., 2010) or 3D data (Yemenicioglu et al., 2016).

Street View is a technology that consists of street-level 360 degree panoramic images and provides mobile and desktop clients with a virtual reality environment in which users can virtually explore streets and cities (Anguelov et al., 2010). Panoramic image is defined as a picture of an area, providing an unlimited view in all directions (Amiri Parian & Gruen, 2010). The omnidirectional vision of an area gives an overview understanding of the environment (Fangi & Nardinocchi, 2013), therefore street view service is used in various applications. (Rundle et al., 2011) used Google street view to audit neighbourhood environment in their study. (Hanson et al., 2013) used Google street view to calculate the severity of pedestrian crashes. (Kelly et al., 2012) and (Curtis et al., 2013) used Google street view to observe the built environment.

Human computer interaction is an indispensable need for seamless communication (Isikdag, 2020). Mouse, keyboard, joystick etc. are the most commonly used tools for navigation in desktop virtual reality applications. But the use of the human body motions can provide a better understanding and interpretation of the virtual environment (Roupé et al., 2014). Human action recognition is a challenging subject in the computer vision community, which aims to understand human gestures from video and image sequences (Tran et al., 2012; Zhou et al., 2009). A new way to overcome this challenging task has risen with the release of depth cameras that allow acquiring dense and three-dimensional scans of a scene in real-time (Schwarz et al., 2012). However, these types of devices (e.g. time of flight cameras) couldn't become widespread due to their high prices until the release of Microsoft Kinect for Xbox 360 in November 2010.

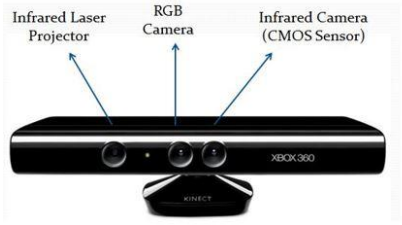
Microsoft Kinect is a depth sensing hardware which was designed to change the way people play games. It enables users to play video games with body motions. The way of playing games without controllers have brought a new perspective and users adapted Kinect to other applications. Depth maps acquired with Microsoft Kinect are widely used in computer vision applications. (Bakirman et al., 2017) employed Kinect for human face modelling. (Yue et al., 2014) and (Izadi et al., 2011) used depth images captured with Kinect to reconstruct 3D environment. (Xia et al., 2011) and (Shotton et al., 2011) proposed different human detection methods using depth information derived from Kinect. (Raheja et al., 2011) used Kinect depth images to track fingertips and centres of palm.

In this study, it is aimed to create a virtual reality environment for Yildiz Technical University Davutpasa Campus via developing a custom Google Street View service using Google Maps JavaScript API and integrate this service with Microsoft Kinect and FAAST software (Suma et al., 2013) to navigate it with human body motions.

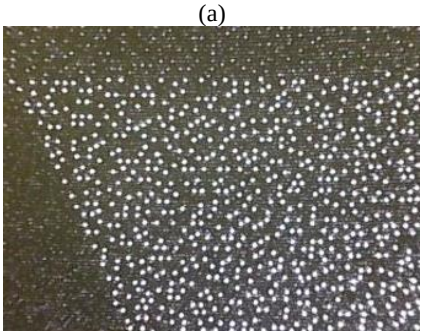
2. MATERIALS AND METHODS

The study area is Yildiz Technical University Davutpasa Campus which has a historical environment located in Istanbul, Turkey. The location was used as a military base during Ottoman Empire and called Davutpasa Barracks which is believed to have built in 1832. In 1999, Davutpasa Barracks was turned into a campus and became a part of Yildiz Technical University. The campus area is 1.75 square kilometres so it is a huge area to explore. Therefore, it is aimed to create a virtual reality environment to explore and learn about the campus.

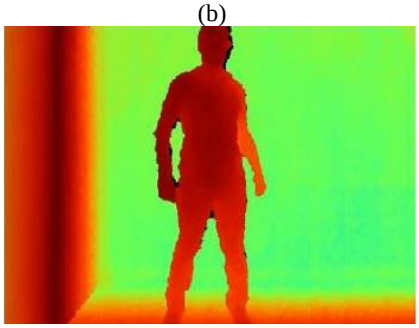
In this study, Microsoft Kinect for Xbox 360 is used for sensing human body motions. Kinect has two types of drivers in order to function on PC. The first driver is Kinect for Windows SDK released by Microsoft and the second driver is released by an organization called OpenNI which consists of three members including PrimeSense (along with ASUS and Willow Garage) who has developed the base technology behind Kinect. In this study, Kinect for Windows SDK v1.8 is used.



(a)



(b)



(c)

Figure.1 (a) Microsoft Kinect, (b) Projected Infrared Pattern (Roborealm, 2016), (c) Depth Image.

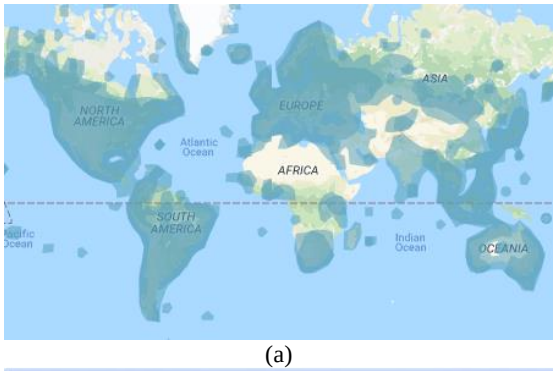
Microsoft Kinect consists of as infrared camera, RGB camera and infrared laser projector (Fig 1a). It measures the distance from the sensor into the environment by

using the structured light principle (Freedman et al., 2013). A pattern which is known by the device is scattered into the scene by infrared laser projector (Fig 1b). The scattered pattern is captured by the infrared camera which is a monochrome complementary metal oxide semiconductor (CMOS) sensor. Since relative geometry between the infrared projector and the infrared camera is known, the depth map can be produced by using 3D triangulation (Fig 1c). Dark red and light green represent small to high distance from the sensor respectively.

FAAST (Flexible Action and Articulated Skeleton Toolkit) is a toolkit that lets users control video games and virtual reality environments by human motion using Kinect for Windows SDK or OpenNI developed by University of South California, Institute for Creative Technologies (Suma et al., 2013).

Street View is a technology that presents panoramic images in the street level around the world via Google Earth software or Google Maps. In 2007, Google released Street View for 5 American cities. Coverage area has been rapidly increased in the following years (Fig. 2). With the release of Google Maps JavaScript API v3, Google provided users with Street view service. Thus, third party users can present custom street view services in personal websites with Google interface. Google also enables users to linking custom Street View services with Google's existing street view panoramic images.

Street view consists of 360 degree spherical panoramas. Spherical panoramas are obtained by using spherical video cameras. Panoramic images that are used in Street View must be conformed to the equirectangular (plate carrée) projection in which meridians, parallels and two poles are straight lines and these images have 2:1 aspect ratio. In this study, 360 degree video streams were captured using Ladybug 2 spherical video camera which is developed by Point Grey Research Inc. This camera has 6 fisheye lenses with Sony ICX204 sensors (Point-Grey, 2014).



(a)

(b)

Figure.2 (a) Street view coverage (Google, 2019), (b) An example equirectangular panorama.

Panoramic image is created from raw images that are captured from these 6 cameras. Image stitching process can be overviewed in Fig 3. Six images are synchronically captured. Images can be compressed as JPEG to reduce time while transferring files to PC (Akcay et al., 2017). In this case, images will be uncompressed on PC to get raw images again. Raw images are converted to RGB colour code with selected interpolation technique (Fig 4a). In this study, we have used Rigorous colour processing technique which provides the best quality colour results. RGB images are rectified and mapped to polygon meshes whose geometric vertices arranged in a three dimensional coordinate system (Fig 4b). Since all images are captured in outdoors, a 20 meter virtual sphere has been utilized. We have also used blending width value of 100 and applied brightness correction for darker areas.

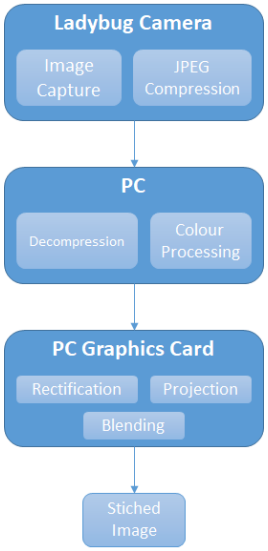


Figure.3 Image stitching process overview (Point-Grey, 2014)

3. RESULTS AND DISCUSSION

In this study, 561 images were captured from 11 spherical video streams on February 3rd 2009 in the campus. Weather conditions were mostly cloudy, so images were not as bright as expected.

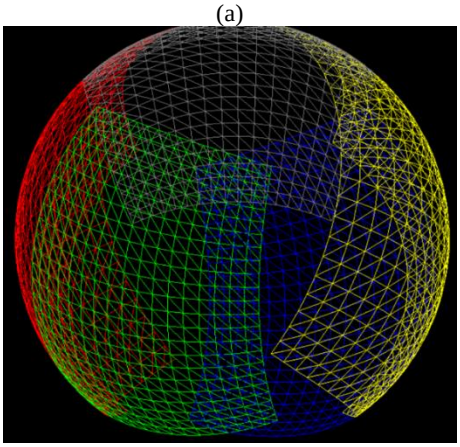
Google Maps JavaScript API requires images to be on the equirectangular projection which has 2:1 orientation. So, spherical images with 5400x2700 resolution are used.

Street view was created with selected 447 images by using Google Maps JavaScript API. Custom street view service was created with workflow shown in Fig 5.

Images were given IDs based on their paths. API's Street View Service doesn't work on local PCs due to security reasons. Hence, all images were uploaded to a server. Street view provider is set up using HTML which includes street view options like zoom levels, starting panorama, etc. Subsequently, street view object and street view link object are created using API library which is followed by modelling a function to get custom panoramas. This function determines panorama image size and custom panorama URLs. With the help of the created function, all panoramic images were defined by their image IDs and locations using a switch-case loop. Finally, links which provide shifting from one panorama

to another are created for each panorama (case). An HTML file was created for each of different starting locations. The final look of the created street view page can be seen in Fig. 6.

(a)



(b)

Figure.4 (a) RGB images (Camera 0 to 5), (b) Polygon meshes (Point-Grey, 2014)

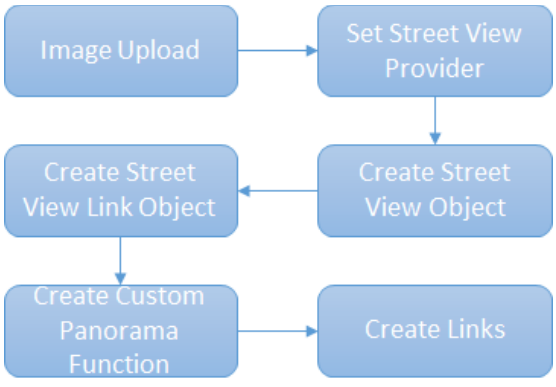


Figure.5 Custom Street View development workflow

FAAST interface has been employed to integrate custom street view application with human motions. FASST can support up to four skeletons and twenty-four joint points are detected for each skeleton. These joints are listed in Table 1.

Figure.6 Custom Street View application

Table 1 Body joint list (Suma et al., 2013)

Joint ID	Body Part	Joint ID	Body Part
0	Head	12	R. Elbow
1	Neck	13	R. Wrist
2	Torso	14	R. Hand
3	Waist	15	R. Fingertip
4	L. Collar	16	L. Hip
5	L.Shoulder	17	L. Knee
6	L. Elbow	18	L. Ankle
7	L. Wrist	19	L. Foot
8	L. Hand	20	R. Hip
9	L.Fingertip	21	R. Knee
10	R. Collar	22	R. Ankle
11	R.Shoulder	23	R. Foot

With the use of FAAST, a specified human motion can trigger a keyboard command. In this study, six moves are determined to assign them as keyboard commands. Body motions with their responsive keyboard commands and functions are listed in Table 2.

Each input has a different type and amount of descriptors. For example, turn right, turn left, lean backwards and lean forwards moves have five descriptors. The first descriptor defines the type of move, for example, turn, lean or jump. So, if turn right move is defined, the first descriptor would be 'Turn'. The second descriptor defines what direction the body will turn to. In this case, the second descriptor would be 'Right'. The third descriptor determines if the move would be the upper limit (at most) or the lower limit (at least) move. In this scenario, the third descriptor would be 'At least'. Fourth and fifth descriptor define move's measure and unit. Twenty-five degrees would be enough measure for turn right move.

Thus, turn right move occurs when the user's body turns at least twenty-five degrees to the right. Turn left, lean backwards and lean forwards moves are defined as explained above. The moves and descriptors are listed in Table 3.

Table 2 FAAST input motions and output commands

Body Motion (Input)	Keyboard Command (Output)	Function
Turn Right	Right Arrow	Turns street view to right
Turn Left	Left Arrow	Turns street view to left
Lean Backwards	Page Up	Turns street view to up
Lean Forwards	Page Down	Turns street view to down
Right Foot Forwards	Up Arrow	Switches to next panorama
Right Foot Backwards	Down Arrow	Switches to the previous panorama

Table 3 Input descriptors of 'Turn Right', 'Turn Left', 'Lean Backwards', 'Lean Forwards'

	Turn Right	Turn Left	Lean Backwards	Lean Forward
1st Descriptor	Turn	Turn	Lean	Lean
2nd Descriptor	Right	Left	Backward	Forward
3rd Descriptor	At least	At least	At least	At least
4th Descriptor	25	25	10	10
5th Descriptor	Degree	Degree	Degree	Degree
1st Descriptor	Turn	Turn	Lean	Lean

Right foot forward and right foot backwards moves have a different set of descriptors because these moves consist of the relation of two body parts. The first of six descriptors defines the first body part. So, if right foot forwards move is defined, the first descriptor would be 'Right Foot'. The second descriptor is the relationship type with the second body part such as 'to the right of', 'above', etc. The third descriptor defines the second body part which will be related to the first body part. Fourth, fifth and sixth descriptors are same as the first set of moves' third, fourth and fifth descriptor respectively. As a result right foot forwards move occur when user's right foot is at least 25 centimetres in front of user's torso. Right foot backwards move is defined in the same manner and descriptors are listed in Table 4.

Each input has an output command which consists of four descriptors. The first descriptor of output determines if the command button will be 'pressed once' or 'kept hold'. The second descriptor is the keyboard command. The third descriptor specifies when the keyboard command will end. The last descriptor defines the measure of the third descriptor. All output commands and their descriptors are listed in Table 5.

Table 4 Input descriptors of 'Right Foot Forwards' and 'Right Foot Backwards'

	R. Foot Forwards	R. Foot Backwards
1st Desc.	Right Foot	Right Foot
2nd Desc.	In front of	Behind
3rd Desc.	Torso	Torso
4th Desc.	At least	At least
5th Desc.	25	25
6th Desc.	Centimeter	Centimeter

Table 5 Output descriptors for all moves.

	1st Desc.	2nd Desc.	3rd Desc.	4th Desc.
Turn Right	Hold	Right Arrow	Until Complete	1
Turn Left	Hold	Left Arrow	Until Complete	1
Lean Backwards	Hold	Page Up	Until Complete	0
Lean Forwards	Hold	Page Down	Until Complete	0
Right Foot Forwards	Hold	Up Arrow	Until Complete	0
Right Foot Backward	Hold	Down Arrow	Until Complete	0

Since input and output descriptors are assigned into the software, Street View can be controlled with body movements by starting FAAST emulator. For example, required move to turn street view angle to the right is shown in Fig. 7. Fig. 7a shows the current state of street view. Turn right move occurs as can be seen in Fig. 7b and street view rotates to right as long as the move goes on (Fig. 7c).

4. CONCLUSION

Video game technologies have rapidly improved around the world so these technologies can be integrated into engineering applications. An application of this integration was presented by this study.

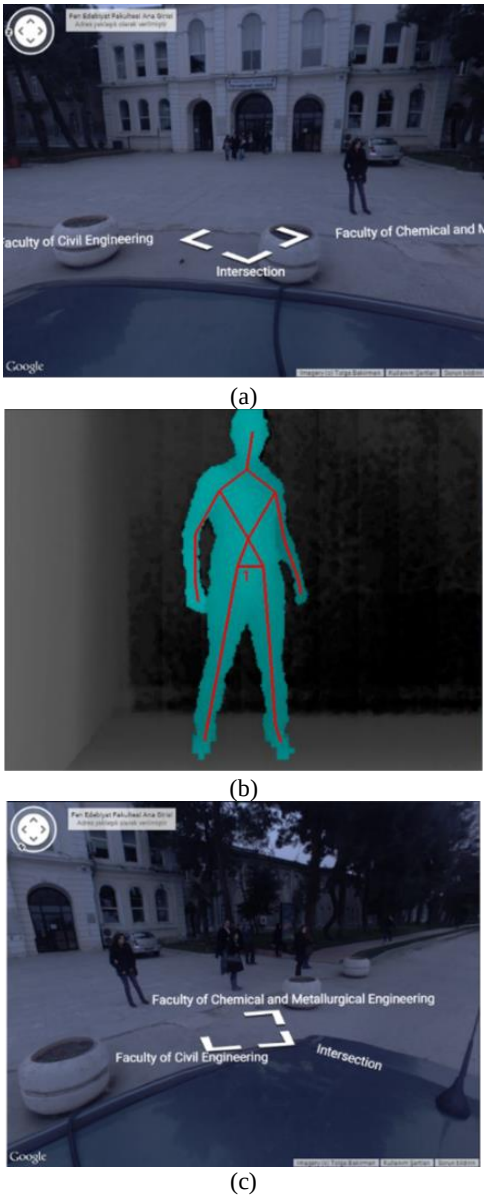


Figure.7 (a) Street view before turn right move, (b) Turn right move on FAAST emulator, (c) Street view after turn tight move

In this study, a virtual reality environment for Yildiz Technical University Davutpasa Campus is created via developing a custom Street view service using panoramic images via Google Maps JavaScript API v3 and integrating with human body motions. Human body motions are used to navigate in a virtual reality environment with Microsoft Kinect, which creates a greater experience. Thus, an alternative low-cost way to control Google street view service to achieve a better virtual reality environment and provide information about the study area are proposed. We also plan to implement the proposed framework for an indoor application in the future. This system can also be implemented for other applications e.g. creating virtual museums, heritage sites or planetariums which will also contribute to preservation and documentation of cultural heritage.

ACKNOWLEDGEMENTS

We would like to thank Mapbis for their contributions on image capturing and processing.

REFERENCES

Akcay, O., Erenoglu, R. C., & Avsar, E. O. (2017). The Effect of JPEG Compression in Close Range Photogrammetry. International Journal of Engineering and Geosciences, 2(1), 35-40. doi: 10.26833/ijeg.287308

Amiri Parian, J., & Gruen, A. (2010). Sensor modeling, self-calibration and accuracy testing of panoramic cameras and laser scanners. ISPRS Journal of Photogrammetry and Remote Sensing, 65(1), 60-76. doi: https://doi.org/10.1016/j.isprsjprs.2009.08.005

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., . . . Weaver, J. (2010). Google Street View: Capturing the World at Street Level. Computer, 43(6), 32-38. doi: 10.1109/MC.2010.170

Bakirman, T., Gumusay, M. U., Reis, H. C., Selbesoglu, M. O., Yosmaoglu, S., Yaras, M. C., . . . Bayram, B. (2017). Comparison of low cost 3D structured light scanners for face modeling. Applied Optics, 56(4), 985-992. doi: 10.1364/AO.56.000985

Chapman, D., & Deacon, A. (1998). Panoramic imaging and virtual reality — filling the gaps between the lines. ISPRS Journal of Photogrammetry and Remote Sensing, 53(6), 311-319. doi: https://doi.org/10.1016/S0924-2716(98)00016-1

Curtis, J. W., Curtis, A., Mapes, J., Szell, A. B., & Cinderich, A. (2013). Using google street view for systematic observation of the built environment: analysis of spatio-temporal instability of imagery dates. International Journal of Health Geographics, 12(1), 53. doi: 10.1186/1476-072X-12-53

Fangi, G., & Nardinocchi, C. (2013). Photogrammetric Processing of Spherical Panoramas. The Photogrammetric Record, 28(143), 293-311. doi: 10.1111/phor.12031

Freedman, B., Shpunt, A., Machline, M., & Arieli, Y. (2013). Depth mapping using projected patterns: Google Patents.

Google. (2019). Where we've been. Retrieved 11.06.2019, from https://www.google.com/streetview/explore/

Gutierrez, M., Vexo, F., & Thalmann, D. (2008). Stepping into virtual reality: Springer Science & Business Media.

Hanson, C. S., Noland, R. B., & Brown, C. (2013). The severity of pedestrian crashes: an analysis using Google Street View imagery. Journal of Transport Geography, 33, 42-53. doi: https://doi.org/10.1016/j.jtrangeo.2013.09.002

Isikdag, U. (2020) An IoT Architecture for Facilitating Integration of GeoInformation. International Journal of Engineering and Geosciences, 5(1), 15-25. doi: 10.26833/ijeg.587023

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., . . . Fitzgibbon, A. (2011). KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. Paper presented at the Proceedings of the 24th annual ACM symposium on User interface software and technology, Santa Barbara, California, USA.

Kelly, C. M., Wilson, J. S., Baker, E. A., Miller, D. K., & Schootman, M. (2012). Using Google Street View to Audit the Built Environment: Inter-rater Reliability Results. Annals of Behavioral Medicine, 45(suppl_1), S108-S112. doi: 10.1007/s12160-012-9419-9

Le Yaouanc, J.-M., Saux, É., & Claramunt, C. (2010). A semantic and language-based representation of an environmental scene. GeoInformatica, 14(3), 333-352. doi: 10.1007/s10707-010-0103-6

Point-Grey. (2014). Overview of the Ladybug Image Stitching Process. Retrieved 06.02.2019, from https://www.flir.eu/globalassets/support/iis/application-notes/tan2008010_overview_ladybug_image_stitching.pdf

Raheja, J. L., Chaudhary, A., & Singal, K. (2011). Tracking of Fingertips and Centers of Palm Using KINECT. Paper presented at the Proceedings of the 2011 Third International Conference on Computational Intelligence, Modelling & Simulation.

Roborealm. (2016). Kinect Targeting. Retrieved 01.11.2018, from http://www.roborealm.com/tutorial/FIRST/slide010.php

Roupé, M., Bosch-Sijtsema, P., & Johansson, M. (2014). Interactive navigation interface for Virtual Reality using the human body. Computers, Environment and Urban Systems, 43, 42-50. doi: https://doi.org/10.1016/j.compenvurbsys.2013.10.003

Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. American journal of preventive medicine, 40(1), 94-100. doi: 10.1016/j.amepre.2010.09.034

Schwarz, L. A., Mkhitaryan, A., Mateus, D., & Navab, N. (2012). Human skeleton tracking from depth data using geodesic distances and optical flow. Image and Vision Computing, 30(3), 217-226. doi: https://doi.org/10.1016/j.imavis.2011.12.001

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., . . . Blake, A. (2011, 20-25 June 2011). Real-time human pose recognition in parts from single depth images. Paper presented at the CVPR 2011.

Suma, E. A., Krum, D. M., Lange, B., Koenig, S., Rizzo, A., & Bolas, M. (2013). Adapting user interfaces for gestural interaction with the flexible action and articulated skeleton toolkit. Computers & Graphics, 37(3), 193-201. doi: https://doi.org/10.1016/j.cag.2012.11.004

Tran, K. N., Kakadiaris, I. A., & Shah, S. K. (2012). Part-based motion descriptor image for human action recognition. Pattern Recognition, 45(7), 2562-2572. doi: https://doi.org/10.1016/j.patcog.2011.12.028

Xia, L., Chen, C., & Aggarwal, J. K. (2011, 20-25 June 2011). Human detection using depth information by Kinect. Paper presented at the CVPR 2011 WORKSHOPS.

Yemenicioglu, C., Kaya, S., & Seker, D. Z. (2016). Accuracy of 3D (Three-Dimensional) Terrain Models in Simulations. International Journal of Engineering and Geosciences, 1(1), 34-38, doi: 10.26833/ijeg.285223

Yue, H., Chen, W., Wu, X., & Liu, J. (2014). Fast 3D modeling in complex environments using a single Kinect sensor. Optics and Lasers in Engineering, 53, 104-111. doi: https://doi.org/10.1016/j.optlaseng.2013.08.009

Zhou, H., Wang, L., & Suter, D. (2009). Human action recognition by feature-reduced Gaussian process classification. Pattern Recognition Letters, 30(12), 1059-1066. doi: https://doi.org/10.1016/j.patrec.2009.03.013

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 073-093, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.580027

PERFORMANCE ANALYSIS OF AMBIGUITY RESOLUTION ON PPP AND RELATIVE POSITIONING TECHNIQUES: CONSIDERATION OF SATELLITE GEOMETRY

Sermet Ogutcu*

Necmettin Erbakan University, Engineering Faculty, Department of Geomatic Engineering, Konya, Turkey
(sermetogutcu@konya.edu.tr); **ORCID 0000-0002-2680-1856**

*Corresponding Author, Received: 19/06/2019, Accepted: 28/08/2019

ABSTRACT: Ambiguity resolution plays an important role in surveying using Precise Point Positioning (PPP) and relative positioning techniques that require high accuracy. In this study, ambiguity resolution performance of PPP and relative positioning under the unobstructed (with 7° cut-off angle) and constrained environment (with 25° cut-off angle, such as nearby buildings and street-canyons) using final/ultra-rapid orbit and clock products are investigated for different observation time. Seventeen globally distributed stations and six baselines of lengths from 270 km to 2100 km are chosen for conducting PPP and relative positioning, respectively. A 31-day period in January 2018 is chosen for processing using 24-, 12-, 6-, 4-, 2- and 1-h observations. The results indicate that sub-mm to cm levels of improvement in horizontal and vertical coordinate components are generally observed with ambiguity resolution for PPP and relative positioning techniques compared to the float counterparts. Moreover, accuracy degradation of ambiguity resolution compared to float solution is observed generally in the vertical component using the 25° elevation cut-off angle for both techniques. As the observation time increases, the accuracy improvements from ambiguity resolution decrease for each technique. In addition, fixing to the wrong integer ambiguities are generally seen with a short observation time and a 25° elevation cut-off angle for both techniques due to the poor satellite geometry. As far as baseline length in relative technique is concerned, the testing results show that there is no direct relation between baseline length and the accuracy improvement from ambiguity resolution compared to the float solution. The results also reveal that the coordinates obtained from ambiguity resolution does not significantly change in the relative technique using final or ultra-rapid orbit/clock products, whereas the changes in PPP are significant for most of the stations.

Keywords: *Ambiguity Resolution, GPS, Precise Point Positioning, Relative Positioning, Satellite Geometry*

1. INTRODUCTION

Global Navigation Satellite System (GNSS) has been commonly used in several scientific and engineering applications including positioning, navigation, and time transfer for several decades. Relative positioning and Precise Point Positioning (PPP) are two main techniques for geodetic surveying. Relative positioning techniques are based on the principle that at least one of the receivers serves as a base station with known coordinates and the other receivers serve as rover stations, in which the coordinates are determined relative to the base station (Blewitt and Young 1989; Schwarz et al. 2009).

PPP (Zumberge et al. 1997) has become a powerful tool for obtaining precise positioning using precise satellite orbit and clock corrections provided by, for example, International GNSS service. It plays a key role in much scientific research, such as tectonic, geophysical studies, structural health monitoring and, (Shi et al. 2010; Geng et al. 2017; Alcay et al. 2019) atmosphere and earthquake monitoring (Bar-Server et al. 1998; Li et al. 2013).

PPP and relative methods have advantages and drawbacks. While relative technique requires as minimum of two GNSS receivers, PPP requires a single GNSS receiver. In this way, field operational cost can be significantly reduced using PPP. Even though millimeter level accuracies can be achieved when processing daily solutions in static mode using both techniques, the long convergence time restricts PPP users to short observation sessions or real-time applications (Pehlivan et al. 2019; Yigit et al. 2014). Moreover, the accuracy of coordinates obtained from traditional PPP is still weaker than that obtained from the relative techniques for short observation time, since traditional PPP is mostly implemented with an ambiguity-float solution. The main obstacle that prevents PPP ambiguity resolution (AR) is the non-integer uncalibrated hardware delay (UHD) originating both in receivers and satellites, which is eliminated by double differencing in relative positioning (Geng et al. 2009). Implementing AR in PPP requires determining UHD in advance using a network. In this way, AR in PPP is only possible when service providers run the network consisting of reference stations to compute hardware biases, in addition to the satellite orbits and clocks that are consistent for PPP AR (Teunissen and Khodabandeh 2015; Choy et al. 2017). Most of the online PPP service providers such as GAPS, CRSR-PPP (transition process to AR started on August 16th, 2018), MagicGNSS, and several types of PPP software such as Bernese, RTKLIB, and GpsTools provide the ambiguity-float solution.

There are numerous research studies addressing AR in PPP. Goudarzi et al. (2018) investigated the performance of PPP with and without AR to recover station velocities using daily observations. They found that the estimated horizontal and vertical velocities from the two solutions were in good agreement. Gandolfi et al. (2017) assessed the effect of GPS-only PPP-AR on coordinate precision using one year of data from 14 European GNSS stations. They found that precision only for the easting component obtained from fixed solutions significantly changed compared to the float solutions. Xiao et al. (2018) investigated the GPS-only AR in PPP using 10 days of data with 3-h sessions for 20 stations. They found that root-mean-squared-error (RMSE) of the horizontal components obtained from the fixed solutions is significantly lower compared to the float solutions. Li et al. (2018) evaluated the performance of kinematic GPS-only PPP-AR under different cut-off angles (7-30) and different session lengths (from 10 min to 2 -hours). They found that as the degree of elevation mask increased, the accuracy of GPS-only PPP-AR solution decreased dramatically. Xin et al. (2018) investigated BeiDou-only PPP-AR with two local networks (Hong Kong, CMONOC) and one global network (MGEX) using 10 days of data. They found that after AR, the improvements in two-dimensional and vertical positioning of 1h static PPP were 1.8 cm/4.5 cm, 8.6 cm/4.5 cm and 1.7 cm/3.1 cm for the Hong Kong, CMONOC, and MGEX networks respectively. Geng et al. (2009) conducted GPS-only PPP- AR using six days of data with 27 IGS stations. The results show that hourly position accuracy is improved from 3.8 cm, 1.5 cm, and 2.8 cm to 0.5 cm, 0.5 cm and 1.4 cm in east, north, and up respectively.

There are also studies related to AR in relative positioning. Chen et al. (2017) investigated the repeatability of short baseline components for float and fixed solutions using 24 h data. They found 2.8 mm and 5 mm improvements for horizontal and vertical components, respectively. Brach and Zasada, 2014 conducted relative positioning using fixed and float solutions under forest conditions. The experimental results show that a relatively small number of fixed solutions were obtained under forest conditions compared to open sky conditions. The results also confirmed that by extending the float observation time, it is possible to achieve accuracy similar to the fixed solution in the forest conditions. Bezcioglu et al. (2019) conducted kinematic PPP-AR in Antarctic for the first time. They found that PPP-AR produced more accurate results than PPP-float, with a faster convergence.

The above-mentioned studies were mainly conducted using the fixed elevation cut-off angles to obtain an unobstructed observation environment (around 7°-10°) and final orbit/clock products. Coordinate improvement and reliability of fixing ambiguities after AR for PPP and relative positioning under poor satellite geometry (such as nearby buildings and street-canyons) and using ultra-rapid orbit and clock products taking into consideration of the observing sessions (for near-real-time users) are not well addressed in the current literature, and worth investigating. High success rates are required for AR to be reliable. Unsuccessful AR, when passed unnoticed, often leads to unacceptable errors in the positioning results. In this paper, AR performance of PPP and relative positioning under the constrained environment and using final/ultra-orbit and clock products are investigated.

This paper is organized as follows. Section 2 describes the PPP and relative positioning functional models, Section 3 presents the data processing. The results are summarized and discussed in Section 4 and the work is concluded in Section 5.

2. FUNCTIONAL MODELS OF PPP AND RELATIVE POSITIONING

In general, ionosphere-free code and carrier phase observations are used for PPP and relative positioning in order to eliminate the first-order ionosphere effect. The ionosphere-free code and carrier phase observables can be written as:

$$P^s_{IF,r} = (f1^2 * P1^s - f2^2 * P2^s)/(f1^2 - f2^2) \quad (1)$$

$$\emptyset^s_{IF,r} = (f1^2 * \emptyset1^s - f2^2 * \emptyset2^s)/(f1^2 - f2^2) \quad (2)$$

where f1 and f2 are two carrier-phase frequencies in Hertz and P1, P2, ∅1, and ∅1 are the measured pseudorange and carrier phase in meters on two frequencies. The observation equations can be written for PPP as follows:

$$P^s_{IF,r} = \rho + c * (dt_r - dt^s) + d_{trop} + HD_{P,r,IF} - HD^{P,s,IF} + \epsilon_{P,IF} \quad (3)$$

$$\emptyset^s_{IF,r} = \rho + c * (dt_r - dt^s) + d_{trop} + \lambda_{IF} * (N^s_{IF} + UHD_{IF,r} - UHD^{IF,s}) + \epsilon_{\emptyset,if} \quad (4)$$

where the superscript s represents the satellite, the subscript r represents the receiver, $P^s_{IF,r}$ and $\emptyset^s_{IF,r}$ are the ionosphere-free combination of code and phase observations, ρ is the geometric range in meters, c is the speed of light in meters per second, dt_r is the receiver clock offset in seconds, dt^s is the satellite clock offset in seconds, d_{trop} is the tropospheric delay in meters, λ_{IF} is the ionosphere-free wavelength, N_{IF} is the ionosphere-free phase initial ambiguity in cycle, and $UHD_{IF,r}$ and $UHD^{IF,s}$ are the ionosphere-free uncalibrated hardware delays in the cycle originating both in receivers and satellites. $HD_{P,r,IF}$ and $HD^{P,s,IF}$ are the ionosphere-free receiver and satellite code hardware delays and $\epsilon_{P,IF}$ and $\epsilon_{\emptyset,IF}$ represent the ionosphere-free code and phase measurement noise. Ionosphere-free wavelength, ambiguity, code hardware delay and UHD can be written as:

$$\lambda_{IF} = \frac{f1^2}{f1^2 - f2^2} * \lambda_1 - \frac{f2^2}{f1^2 - f2^2} * \lambda_2 \quad (5)$$

$$N_{IF} = \frac{f1^2 * N_1}{f1^2 - f2^2} - \frac{f2^2 * N_2}{f1^2 - f2^2} \quad (6)$$

$$HD_{P,r,IF} = (f1^2 * HD_{r,1} - f2^2 * HD_{r,2})/(f1^2 - f2^2) \quad (7)$$

$$HD^{P,s,IF} = (f1^2 * HD^{s,1} - f2^2 * HD^{s,2})/(f1^2 - f2^2) \quad (8)$$

$$UHD_{IF,r} = (f1^2 * UHD_{r,1} - f2^2 * UHD_{r,2})/(f1^2 - f2^2) \quad (9)$$

$$UHD^{IF,s} = (f1^2 * UHD^{s,1} - f2^2 * UHD^{s,2})/(f1^2 - f2^2) \quad (10)$$

As is seen in Equation 4, ionosphere-free ambiguity can only be estimated as a float (real-value) unknown unless using UHD information from the network. For PPP users, the receiver dependent UHD is generally not a concern and can be safely ignored (Li et al. 2016).

The functional model in relative positioning uses double-differenced (DD) observables, and functional equations can be written as follows (Wielgosz, 2011);

$$\emptyset^{kl}_{1,ij} = \rho^{kl}_{ij} - T^{kl}_{ij} + I^{kl}_{ij} - \lambda_1 * N^{kl}_{1,ij} + \epsilon^{kl}_{\emptyset,1,ij} \quad (11)$$

$$\emptyset^{kl}_{2,ij} = \rho^{kl}_{ij} - T^{kl}_{ij} + \left(\frac{f1^2}{f2^2}\right) * I^{kl}_{ij} - \lambda_2 * N^{kl}_{2,ij} + \epsilon^{kl}_{\emptyset,2,ij} \quad (12)$$

$$P^{kl}_{1,ij} = \rho^{kl}_{ij} - T^{kl}_{ij} - I^{kl}_{ij} + \epsilon^{kl}_{P,1,ij} \quad (13)$$

$$P^{kl}_{2,ij} = \rho^{kl}_{ij} - T^{kl}_{ij} - \left(\frac{f1^2}{f2^2}\right) * I^{kl}_{ij} + \epsilon^{kl}_{P,2,ij} \quad (14)$$

where i and j are the receiver indices, k and l the satellite indices, $\emptyset^{kl}_{n,ij}$ and $P^{kl}_{n,ij}$ are the DD phase and code observations on frequency n (n=1, 2), $\rho^{kl}_{,ij}$ is the DD geometric distance between receiver and satellite, T^{kl}_{ij} is the DD tropospheric delay, I^{kl}_{ij} is DD ionospheric delay and $N^{kl}_{n,ij}$ is the DD carrier phase ambiguities on two frequencies, $\epsilon^{kl}_{\emptyset,n,ij}$ and $\epsilon^{kl}_{P,n,ij}$ are the DD phase and code measurements noise. The unknown parameters are receiver three-dimensional coordinates, DD ionospheric delay and DD ambiguities. Receiver and satellite clock biases are removed using the DD combination, whereas in PPP, receiver clock bias should be estimated as an unknown and satellite clock bias should be obtained from an external source such as IGS clock products. Another advantage of DD combination is that the ambiguity term preserves its integer nature.

Obtaining the fixed solution is generally done in three sequential steps. In the first step, the unknown parameters are solved without taking into account the integer ambiguity. This result is often referred to as the float solution. In the second step, the float solution is used to estimate the integer ambiguity. In the last step, the computed integer ambiguities are used to improve the unknown parameters in the float solution (Joosten and Tiberius, 2000). The unknown parameters are recomputed with the ambiguities constrained to the integer values. This recomputed process is generally referred to in GNSS processing software as the iteration.

There are several techniques to compute the integer value of ambiguities. Wide-lane and narrow-lane combinations (Teunissen and Montenbruck, 2017) are commonly used in the GNSS community, and one of the subjects of this study. The equations for phase and code observations can be written as follows;

$$\emptyset_{LW} = (f_1 * \emptyset_1 - f_2 * \emptyset_2)/(f_1 - f_2) \quad (15)$$
$$\emptyset_{LN} = (f_1 * \emptyset_1 + f_2 * \emptyset_2)/(f_1 + f_2) \quad (16)$$
$$P_{LW} = (f_1 * P_1 - f_2 * P_2)/(f_1 - f_2) \quad (17)$$
$$P_{LN} = (f_1 * P_1 + f_2 * P_2)/(f_1 + f_2) \quad (18)$$

where $\emptyset_{LW}/\emptyset_{LN}$ and P_{LW}/P_{LN} are wide-lane/narrow-lane phase and code combinations respectively, with corresponding wavelengths of ~0.86 m and ~0.107 m for wide-lane and narrow-lane combinations. Ionosphere-free ambiguity is usually expressed as wide- and narrow-lane ambiguity fixing. The wide- and narrow-lane ambiguities cannot be estimated and fixed simultaneously. The main reason is that the equation corresponding to the narrow-lane ambiguity contains the wide-lane ambiguity. Usually, ambiguities are resolved by first computing wide-lane and then narrow-lane ambiguities. An ionosphere-free ambiguity can be fixed only if both its wide- and narrow-lane ambiguities are fixed (Ge et al. 2008).

Ambiguity resolution in PPP was conducted using JPL orbit and clock products which provide wide-lane and dual-frequency phase bias information from the global network of GPS receivers that were used to generate the orbit and clock solutions for the GPS satellites (Bertiger et al., 2010).

3. DATA PROCESSING

Seventeen globally distributed stations and six baselines of lengths varying from 270 km to 2100 km were chosen for conducting PPP and relative positioning, respectively. Figure 1 and Table 1 show the stations and the baselines used for PPP and the relative positioning processing respectively. First stations in Table 1 were used as fiducial stations. To form varying lengths of baselines for relative positioning, most of the used stations are different for PPP and relative positioning.

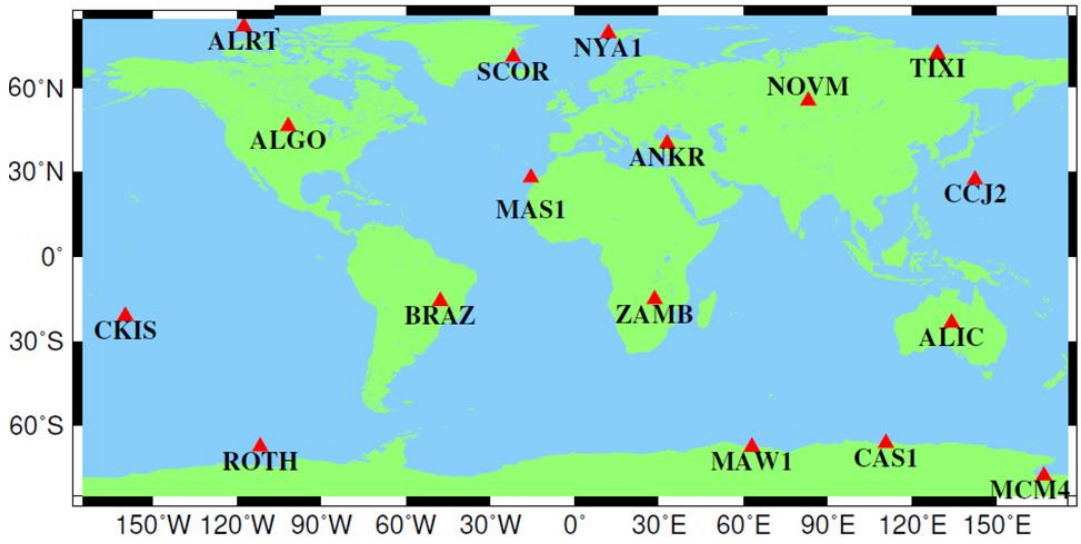


Figure 1. Stations used for PPP processing

Table 1. Baselines

Baselines	Length (km)	Mean latitude	Mean longitude
MERS-MEDI	2100	40.3	23
ALRT-NYA1	1285	80.7	-25
VIS0-MDVJ	1100	56.8	28
GANP-LAMA	536	51.4	20
PALM-ROTH	362	-66.2	-66
MARS-IENG	270	44.1	89

The 31-day period of January 2018 was chosen for PPP and relative processing. Daily RINEX files of the stations were subdivided into mutually non-overlapping sessions as in 24-12-6-4-2-1 h subsets. Data integrity of the RINEX files was checked using in-house software. Data integrities of less than 99% were removed from the processing. Eight daily RINEX files were discarded from the dataset.

Satellite elevation cut-off angles of 7° and 25° were chosen to represent the open sky and obstructed conditions such as street-canyons. Final and ultra-rapid satellite orbit/clock products (observed part) were used to investigate the ambiguity resolution impact on the estimated coordinates for post-processing and near-real-time users (Tusat et al. 2018). Each data set was processed

by GAMIT/GLOBK and GIPSY/OASIS scientific software for relative and PPP techniques respectively. Table 2 summarizes the processing parameters for each type of software.

Table 2. Processing parameters

Processing parameters	Relative processing strategy	PPP processing strategy
GNSS system	GPS	GPS
Observations	Phase and code data on two frequencies	Phase and code data on two frequencies
Data weight of linear combination phase and code	1 cm/ 1 m	1 cm/ 1 m
Weighting with elevation	Applied	Applied
Sampling interval	30 s	30 s
Cut-off angle	7°/25°	7°/25°
Satellite orbit and clock	IGS final/ultra-rapid	JPL final/ultra-rapid (fiducial frame)
Relativistic corrections	Removed using double differences	Periodic clock corrections and gravitational bending (shapiro delay) were applied
Receiver clock error	Removed using double differences	Estimated as white noise
Reference frame realization	Tight constraint using the fiducial station's true coordinates (IGS14) from IGS weekly combined solutions	Fix to satellites ECEF coordinates of JPL fiducial frame (IGS14)
Cycle-slip	Corrected by wide-lane combination using phase and code data	Corrected by Melborne-Wubbena combination.
Receiver clock jump	Corrected	Corrected
Ionosphere	The first order is eliminated by IF combination. Second order is removed using IGS IONEX file	The first order is eliminated by IF combination. Second order is removed using JPL IONEX file.
A Priori troposphere	GPT2 model (Lagler et al. 2013) were applied using tropospheric gradient	GPT2 model (Lagler et al. 2013) were applied using tropospheric gradient.
Wet tropospheric delay	Estimated as random-walk model ($5x10^{-8}$ m^2/s)	Estimated as random-walk model ($5x10^{-8}$ m^2/s)
Tropospheric gradients	Estimated as random-walk model ($5x10^{-8}$ m^2/s)	Estimated as random-walk model ($5x10^{-8}$ m^2/s)
Phase ambiguity	Ambiguity resolution using wide- and narrow-lane and additionally float estimation	Ambiguity resolution using wide- and narrow-lane and additionally float estimation.
Adjustment model	Stochastic Kalman filter	Stochastic Kalman filter/smoother implemented as square root information filter with smoother
Satellite/receiver antenna phase offset	PCO and PCV values from up-to-date igs.atx were used	PCO and PCV values from up-to-date igs.atx were used
Tidal effects	Solid tides, ocean tide loading and polar tides (IERS 2010; Petit and Luzum, 2010)	Solid tides, ocean tide loading and polar tides (IERS 2010; Petit and Luzum, 2010)
Wind-up effect	Corrected (Wu et al. 1992)	Corrected (Wu et al. 1992)
Satellite DCB	Corrected by up to date CODE DCB product	Corrected by up to date CODE DCB product.
Eclipse strategy	Eclipsing satellites were not used until satellites reach nominal attitude	Eclipsing satellites were not used until satellites reach nominal attitude.

4. ANALYSIS AND RESULTS

The reference coordinates of the stations were taken from IGS weekly combined solutions with an accuracy of within a few millimeters. If any coordinate component's error in regard to the true solution was greater than 30 cm, it was assumed to be an outlier and removed from the RMSE computation. This is an arbitrary assumption based on the range of errors and is a suitable criterion with which to determine a threshold without making other outlier assumptions, which sometimes are not suitable (Gandolfi et al. 2017). AR impacts on the estimated coordinates were investigated as the coordinates' improvement with regards to the float solutions in horizontal and vertical components. The positive improvement denotes that RMSE of fix solution is smaller compared to float solution. The negative improvement denotes that RMSE of float solution is smaller compared to fix solution. In order to provide a better visualization. Tables 3 and 4 show the RMSEs of fixed and float solutions with final products using 7°/25° cut-off angles (σ_n, σ_e, σ_u) for PPP. Tables 5 and 6 show the same statistics for relative positioning as in the PPP results, but RMSEs of ultra-products were not given due to the nearly identical results compared to the final product. Figure 2-5 show the PPP horizontal and vertical RMSE improvements of fixed solution with respect to float solutions for final and ultra-products using 7°/25° cut-off angles. The same statistical values were given in Figure 6-7 for relative positioning techniques except for ultra-product due to the nearly identical results with respect to final product.

Table 3. RMS of stations used for PPP processing for 24-12-6 h (values are in mm)

SITES			24 h			12 h			6 h		
			σ _n	**σ _e**	**σ _u**	**σ _n**	**σ _e**	**σ _u**	**σ _n**	**σ _e**	**σ _u**
ALGO	7°	fix	4	5	10	4	5	11	4	6	12
		float	3	7	11	3	6	11	4	8	13
	25°	fix	2	7	17	3	7	18	4	7	21
		float	2	10	16	3	8	18	4	9	20
ALIC	7°	fix	3	2	4	3	2	5	4	3	7
		float	3	2	4	3	3	5	4	6	9
	25°	fix	3	3	21	4	3	23	5	5	27
		float	3	4	19	4	4	21	5	7	26
ALRT	7°	fix	4	5	10	2	1	8	2	3	14
		float	3	7	11	3	2	9	5	4	11
	25°	fix	2	2	14	2	2	19	4	4	24
		float	3	3	11	4	3	17	6	6	22
ANKR	7°	fix	2	2	4	2	2	5	2	2	5
		float	2	6	5	2	5	6	3	5	7
	25°	fix	2	2	6	2	2	8	2	3	11
		float	1	2	5	2	3	6	3	4	11
BRAZ	7°	fix	2	2	4	3	3	6	4	4	10
		float	2	4	5	3	6	7	4	7	11
	25°	fix	4	2	17	5	4	24	5	6	24
		float	4	5	25	5	8	27	7	10	30
CAS1	7°	fix	2	1	7	2	1	7	2	2	8
		float	2	1	7	2	2	8	3	3	9
	25°	fix	5	1	35	5	1	35	6	3	37
		float	5	3	33	5	4	35	7	6	36
CCJ2	7°	fix	3	1	12	3	2	12	4	2	13
		float	3	3	11	4	3	12	4	5	15
	25°	fix	2	2	22	3	2	24	3	2	25
		float	3	3	23	3	3	26	3	5	28
CKIS	7°	fix	2	3	6	3	3	7	3	3	10
		float	2	3	7	2	5	9	3	7	13
	25°	fix	3	2	9	4	3	11	4	4	18
		float	3	3	12	3	5	14	4	7	22
MAS1	7°	fix	2	2	5	3	2	7	3	3	8
		float	2	2	6	3	3	8	4	5	11
	25°	fix	2	2	11	3	4	12	3	4	14
		float	2	3	10	3	5	13	4	8	17
MAW1	7°	fix	1	2	6	1	2	6	1	2	7
		float	1	3	6	1	3	6	2	3	7
	25°	fix	2	2	9	3	2	10	4	3	12
		float	1	4	9	2	5	10	4	5	11
MCM4	7°	fix	2	2	12	2	2	13	2	2	13
		float	2	3	13	2	3	14	3	4	15
	25°	fix	4	3	16	5	3	18	5	4	24
		float	5	6	17	4	5	21	5	6	25
NOVM	7°	fix	2	1	26	2	1	26	3	2	27
		float	2	3	26	2	4	26	4	10	31
	25°	fix	3	4	9	3	4	93	5	4	92
		float	3	5	9	3	6	91	5	8	91

Table 3 (Continued)

NYA1	7°	fix	2	3	4	2	3	5	2	3	7
		float	2	3	6	2	3	6	3	4	9
	25°	fix	1	2	40	2	3	40	2	3	42
		float	2	3	37	2	3	38	3	4	40
ROTH	7°	fix	2	2	4	3	2	5	3	2	7
		float	2	3	5	2	3	5	3	4	6
	25°	fix	2	2	36	3	2	36	4	3	37
		float	2	3	36	3	4	36	5	5	35
SCOR	7°	fix	4	4	15	4	4	16	5	5	16
		float	4	5	15	6	1	9	5	6	18
	25°	fix	4	3	10	4	3	12	5	4	17
		float	4	6	9	6	3	5	5	6	16
TIXI	7°	fix	1	1	3	1	1	4	3	2	11
		float	2	2	3	2	2	6	4	4	12
	25°	fix	1	2	10	2	2	14	2	3	19
		float	2	2	10	2	3	14	4	5	24
ZAMB	7°	fix	2	2	4	2	2	5	3	3	8
		float	2	3	5	2	4	6	3	6	10
	25°	fix	3	2	23	3	3	24	4	12	26
		float	3	4	24	3	6	26	4	12	31

Table 4. RMS of stations used for PPP processing for 4-2-1 h (values are in mm)

SITES			**4 h**			**2 h**			**1 h**		
			***σ* _n**	***σ* _e**	***σ* _u**	***σ* _n**	***σ* _e**	***σ* _u**	***σ* _n**	***σ* _e**	***σ* _u**
ALGO	7°	fix	4	5	12	5	7	16	8	19	24
		float	5	8	13	7	13	17	9	23	21
	25°	fix	4	7	22	10	9	27	20	42	60
		float	6	8	20	10	18	27	24	46	61
ALIC	7°	fix	3	3	9	5	16	22	13	35	50
		float	4	9	11	7	20	23	13	40	47
	25°	fix	4	5	29	8	18	42	23	65	103
		float	5	12	34	9	27	53	22	69	110
ALRT	7°	fix	9	9	16	15	13	24	20	14	25
		float	9	9	14	16	13	23	21	16	26
	25°	fix	13	9	33	15	10	35	17	13	41
		float	11	11	30	15	10	34	18	16	41
ANKR	7°	fix	3	9	9	3	7	8	7	21	23
		float	3	10	9	5	12	13	9	26	22
	25°	fix	3	10	16	5	9	20	16	50	74
		float	4	11	13	7	22	28	19	55	70
BRAZ	7°	fix	4	8	11	5	17	20	10	38	42
		float	4	12	14	6	22	21	11	44	41
	25°	fix	5	9	29	10	30	59	21	76	102
		float	7	16	35	10	36	57	22	86	105
CAS1	7°	fix	3	2	9	3	3	10	7	10	17
		float	3	5	11	6	9	13	13	16	21
	25°	fix	6	4	41	10	9	43	29	49	81
		float	7	9	40	15	17	45	31	51	84
CCJ2	7°	fix	4	2	14	4	4	19	9	21	43
		float	4	6	16	6	13	27	12	29	50
	25°	fix	3	3	27	4	8	35	16	46	73
		float	4	7	30	6	17	42	16	56	82

CKIS	7°	fix float	4 4	4 9	13 15	5 7	11 18	23 25	11 12	40 43	55 52

Table 4 (Continued)

	25°	fix float	6 5	5 10	20 24	8 9	14 24	35 44	19 19	57 64	84 92
MAS1	7°	fix float	4 4	3 7	10 13	5 6	12 17	20 22	8 18	32 54	40 81
	25°	fix float	5 5	5 11	18 21	7 9	17 23	35 38	17 18	51 54	72 81
MAW1	7°	fix float	2 3	2 4	8 8	3 5	3 8	10 11	9 23	9 37	20 50
	25°	fix float	5 5	3 6	14 14	7 9	5 11	21 23	22 23	35 37	50 50
MCM4	7°	fix float	4 5	3 5	15 17	5 8	8 10	18 22	10 14	12 15	28 31
	25°	fix float	6 6	5 7	29 29	10 9	8 10	34 36	19 19	20 22	57 59
NOVM	7°	fix float	4 8	2 15	28 33	8 12	16 28	34 42	17 44	29 67	54 107
	25°	fix float	8 11	5 10	93 90	12 19	27 34	93 84	39 39	72 74	118 106
NYA1	7°	fix float	2 4	3 4	7 9	3 6	3 7	11 14	6 10	6 11	18 20
	25°	fix float	4 6	4 5	45 41	6 8	5 8	46 48	12 14	12 15	60 62
ROTH	7°	fix float	3 3	3 5	8 9	3 6	3 8	10 11	4 9	5 14	14 17
	25°	fix float	5 5	4 7	39 37	7 9	4 11	41 40	14 17	22 28	53 54
SCOR	7°	fix float	5 5	5 8	16 18	5 7	6 10	19 20	12 14	13 17	27 28
	25°	fix float	6 6	5 9	21 18	9 9	7 12	30 26	23 23	29 31	53 45
TIXI	7°	fix float	3 4	2 6	12 15	5 8	5 10	18 20	8 15	10 18	29 37
	25°	fix float	4 8	5 8	23 30	8 15	9 15	42 43	28 30	41 47	82 89
ZAMB	7°	fix float	3 3	7 9	11 13	6 7	12 18	28 23	12 13	31 37	44 45
	25°	fix float	4 5	12 15	30 33	11 11	19 26	43 48	19 23	54 58	85 87

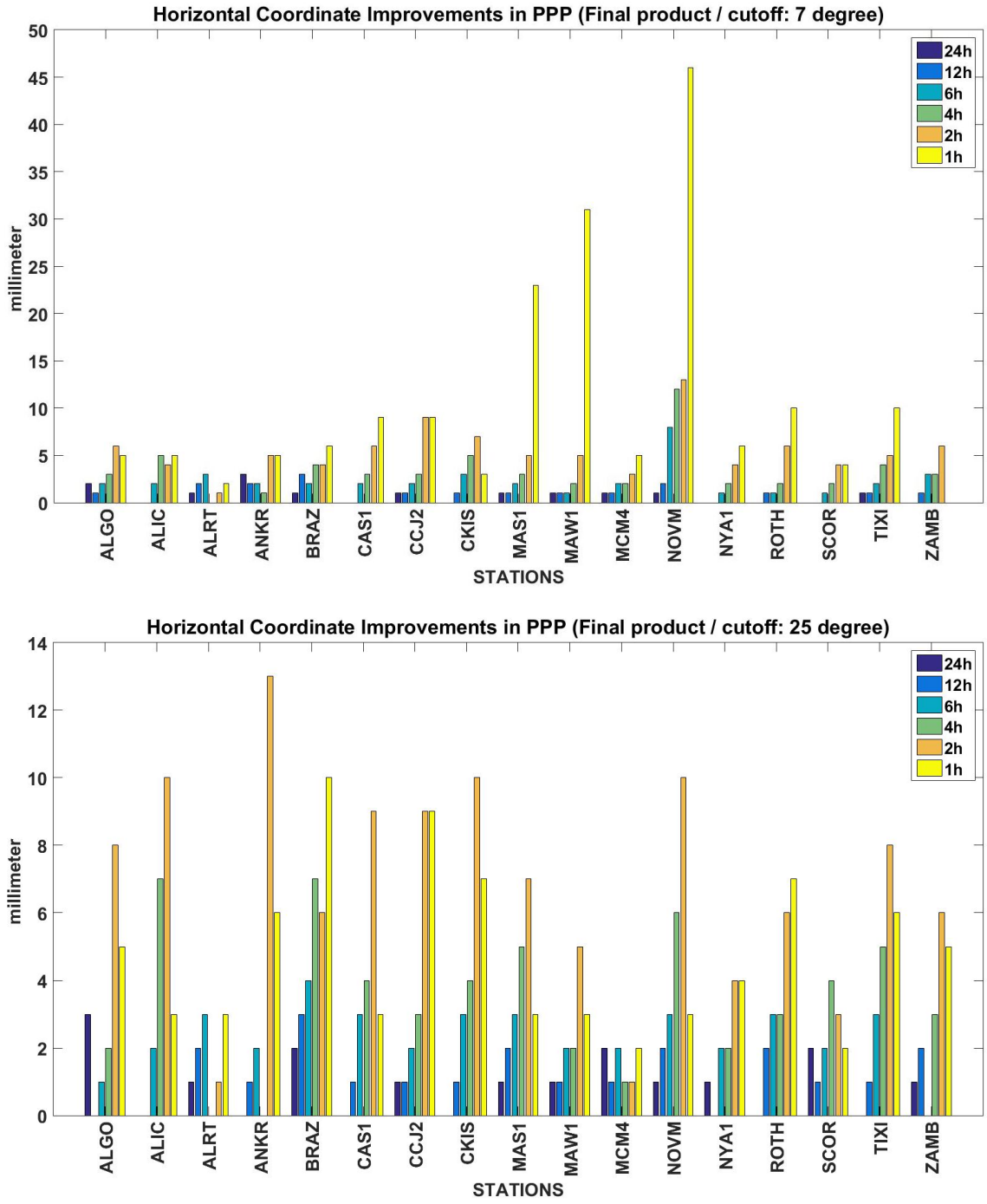


Figure 2. Horizontal coordinate improvements for PPP using final products with 7°/25° cut-off angle

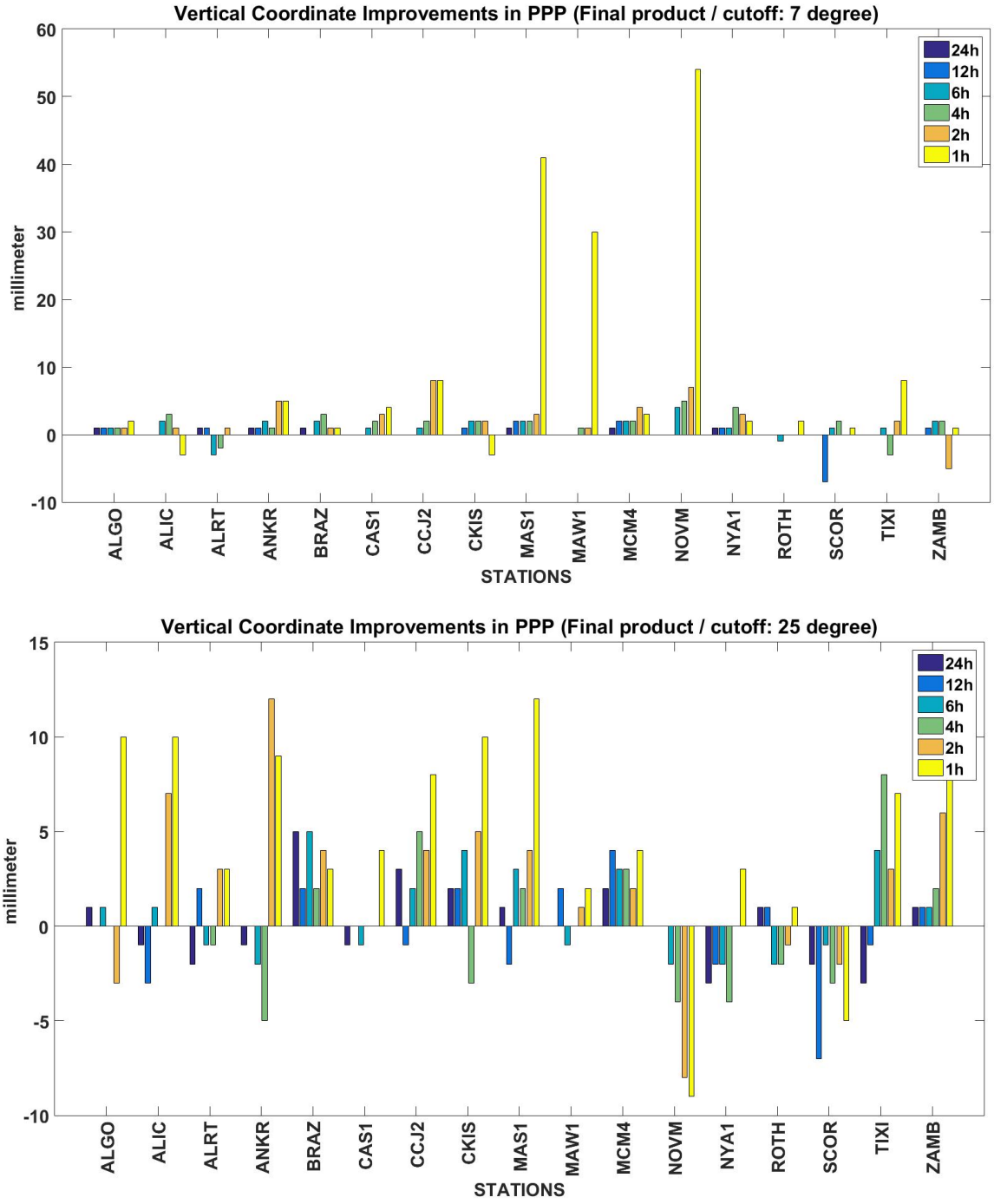


Figure 3. Vertical coordinate improvements for PPP using finals product with 7°/25° cut-off angle

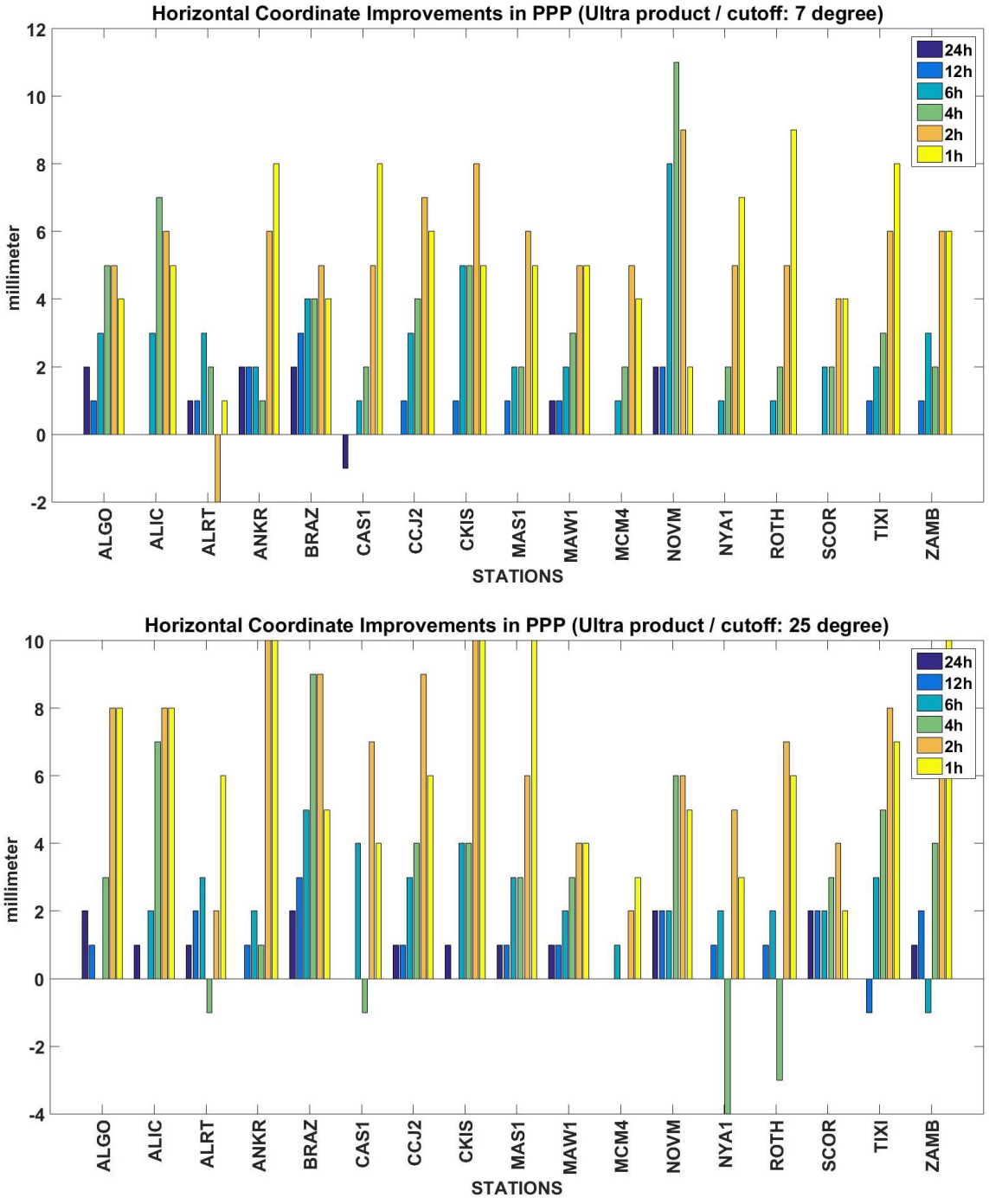


Figure 4. Horizontal coordinate improvements for PPP using ultra products with 7°/25° cut-off angle

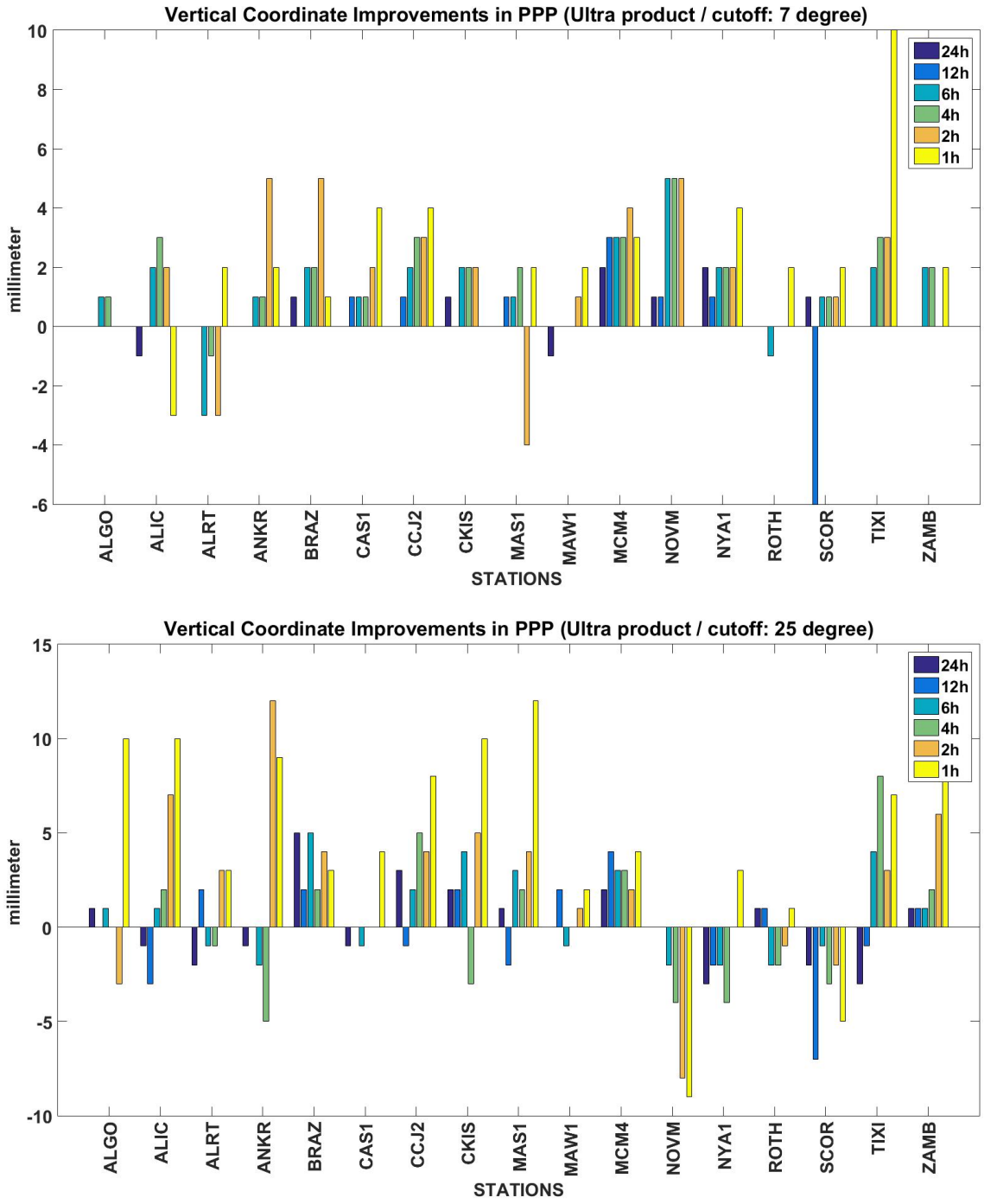


Figure 5. Vertical coordinate improvements for PPP using ultra product with 7°/25° cut-off angle

As can be seen from the PPP results, AR impacts on the estimated coordinates depended on the observation session, satellite geometry, and orbit/satellite product. As was expected, AR impact increased as the observation session length decreased. The results also showed that accuracy degradation of AR compared to the float solutions was generally observed with 25° elevation cut-off angle due to the unstable AR under the poor satellite geometry. The results also revealed that final and ultra-rapid products affected the improvement of PPP-AR on the coordinates for each observation session but no trend was found with regards to the magnitude and direction of the AR impact on the coordinates.

Table 5. RMS of stations used for relative processing for 24-12-6 h (values are in mm)

SITES			24 h			12 h			6 h		
			σ _n	**σ _e**	**σ _u**	**σ _n**	**σ _e**	**σ _u**	**σ _n**	**σ _e**	**σ _u**
IENG	7°	fix	3	3	13	3	3	14	4	4	15
		float	4	6	10	4	7	11	9	18	13
	25°	fix	7	1	159	7	3	158	8	5	156
		float	4	2	158	5	3	155	5	8	149
LAMA	7°	fix	1	1	5	2	1	6	2	2	8
		float	1	2	5	1	3	7	3	4	9
	25°	fix	2	1	23	2	2	24	4	2	24
		float	2	2	24	2	2	25	4	4	25
MDVJ	7°	fix	1	3	9	1	3	10	2	3	11
		float	1	4	9	1	4	10	4	7	10
	25°	fix	5	4	34	5	4	32	4	4	29
		float	4	3	32	4	4	31	4	6	23
MEDI	7°	fix	4	7	22	4	7	22	4	9	23
		float	3	12	23	3	11	22	6	11	22
	25°	fix	5	11	30	6	12	30	7	12	33
		float	6	18	27	7	19	29	8	18	35
NYA1	7°	fix	3	2	5	3	2	6	3	3	8
		float	3	3	5	3	3	6	3	4	8
	25°	fix	3	2	14	4	3	20	5	4	24
		float	3	2	18	4	3	23	6	6	27
ROTH	7°	fix	0	0	3	1	1	4	2	1	6
		float	0	2	3	1	2	4	2	3	6
	25°	fix	2	1	15	3	1	16	3	2	18
		float	3	2	15	3	2	18	3	5	19

Table 6. Statistics of stations used for relative processing for 4-2-1h (values are in mm)

SITES			4 h			2 h			1 h		
			σ _n	**σ _e**	**σ _u**	**σ _n**	**σ _e**	**σ _u**	**σ _n**	**σ _e**	**σ _u**
IENG	7°	fix	6	7	18	18	43	43	40	82	94
		float	11	22	21	23	51	52	43	83	97
	25°	fix	8	6	157	16	45	152	57	113	157
		float	8	11	149	20	50	141	58	116	152
LAMA	7°	fix	2	2	9	3	4	10	11	28	26
		float	3	5	10	6	13	15	14	33	29
	25°	fix	4	2	27	8	14	35	27	64	67
		float	6	6	27	11	21	35	29	68	71
MDVJ	7°	fix	2	3	13	3	6	15	11	26	33
		float	5	7	12	6	13	15	15	32	36
	25°	fix	5	5	33	6	12	35	51	88	99
		float	5	7	28	9	19	36	52	90	102
MEDI	7°	fix	6	10	26	10	24	36	31	81	85
		float	6	16	30	13	31	42	32	83	87
	25°	fix	8	15	39	16	41	69	53	116	142
		float	10	21	46	19	50	79	54	121	141
NYA1	7°	fix	9	9	16	10	6	20	14	11	27
		float	10	10	17	11	8	21	18	15	30
	25°	fix	7	5	35	10	16	46	35	26	98
		float	6	6	36	12	17	50	38	27	101
ROTH	7°	fix	2	1	7	2	1	6	2	1	7
		float	3	4	7	2	3	6	3	4	7
	25°	fix	4	2	19	3	2	18	4	2	19
		float	4	7	20	3	5	19	4	7	20

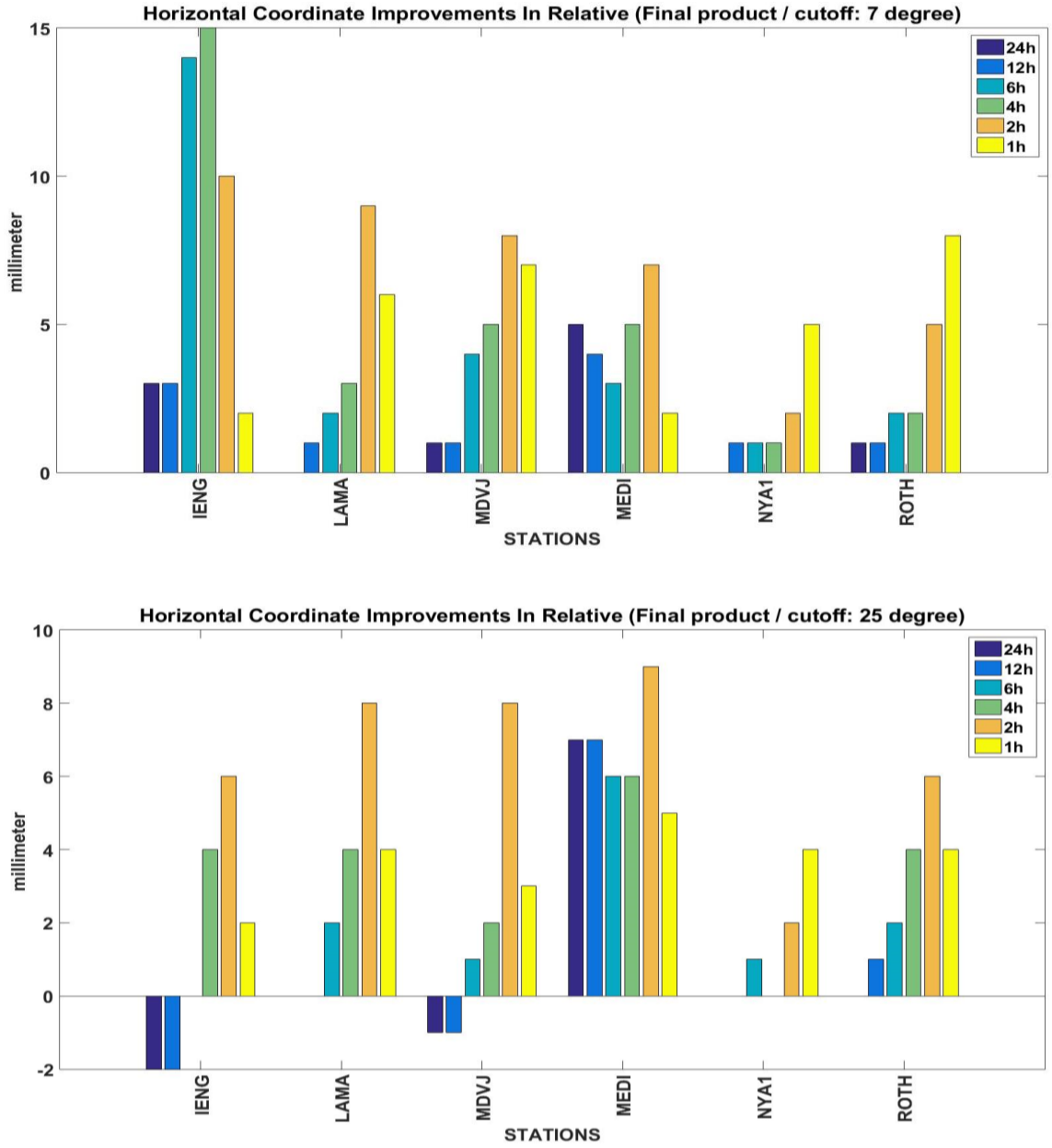


Figure 6. Horizontal coordinates improvements for relative using final product with 7°/25° cut-off angle

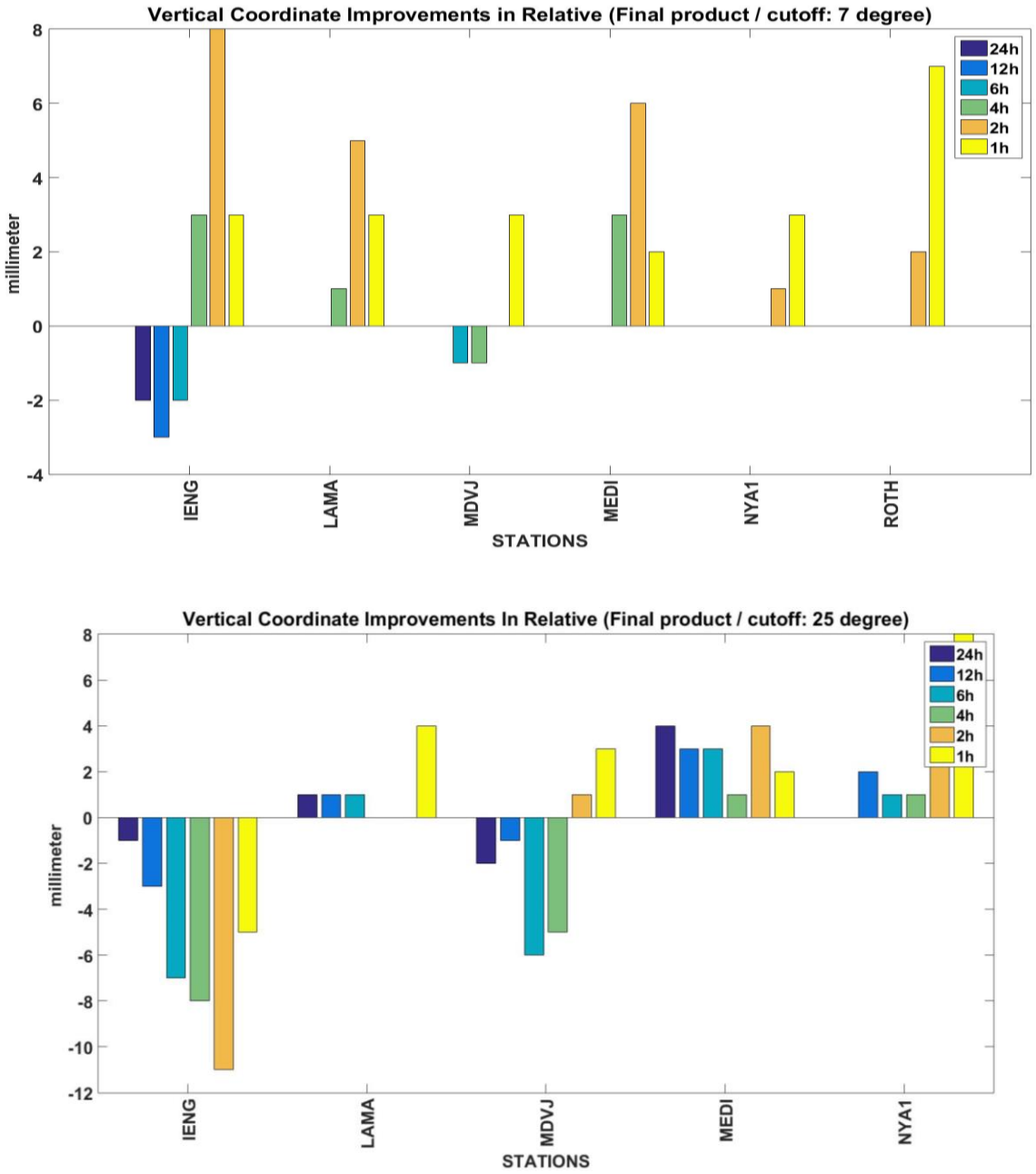


Figure 7. Vertical coordinates improvements for relative using final product with 7°/25° cut-off angle

Relative processing results show that the AR impact on the estimated coordinates did not depend on the baseline length. The accuracy degradation of AR compared to float solutions was generally observed with 25° elevation cut-off angle as in the PPP results. Due to the nearly identical results (sub-mm difference) between the final and ultra-rapid products, AR improvement of ultra-rapid products was not given for the relative processing results. The maximum difference was observed for the longest baseline: a 0.5 mm change in the horizontal and a 0.9 mm change in the vertical. The main reason behind this is that satellite clock bias is removed using double differences, whereas in PPP, it directly lumps to the station coordinates. Orbit differences between the products were relatively small compared to the satellite clock bias. Table 7 shows the average RMSE values of the radial, along-track and cross-track components of all GPS satellites between the final and ultra-rapid products for 31-day period in January 2018.

Table 7. Average RMSE values of radial (R), along-track (A) and cross-track (C) components of all GPS satellites during the observation period (units: mm)

FINAL – ULTRA-RAPID		
R	**A**	**C**
7	12	11

The relationship between observation session and AR impact on the coordinates in relative processing was similar to that in PPP, but the magnitudes of improvement were much higher in PPP. Maximum improvement and degradation in coordinates of AR in horizontal and vertical components were 54/12 mm, -9/-4 mm and 15/10 mm, and -2/-11 mm for PPP and relative positioning, respectively. The AR improvement of each technique in the horizontal component was much bigger than the improvement in the vertical component for most of the stations.

Fixing to the wrong integer ambiguities was investigated for each technique and split into two categories. The wrong fix in the first category is assumed if any fixed solution is within the outliers, but its float solution is not an outlier. The wrong fix in the second category is assumed if any two-dimensional error of the fix solution is 20 mm higher than its error in the float solution. Since the impact of AR is highly correlated with the two-dimensional, this assumption was applied. A 20 mm threshold was taken as an arbitrary value to exclude small biases between the fix and float solutions. Tables 8, -9, -10 and -11 show the total number of wrong fix and their AR success rate for PPP and relative positioning respectively using the final product. Due to the fact that most of the wrong fixes are accumulated in 1h, the other sessions were omitted to save space in this manuscript. Some float solutions are also within the outliers except its fixed solutions. These float solutions within outliers are also shown in Table 8-10. The number of float solution outliers is significantly lower than the number of fixed solution outliers for PPP, but in relative positioning, the numbers of fixed and float outliers are similar except for IENG and MEDI stations. In these stations, the number of fixed solution outliers are higher than theirs float solution outliers.

The fixing rate is defined as the ratio of the ambiguities used over all the ambiguities, and is expressed as follows:

$$\boldsymbol{F} = \frac{\boldsymbol{N}_{used_{ambiguities}}}{\boldsymbol{N}_{all_{ambiguities}}} * \mathbf{100}\% \quad (19)$$

where F denotes the ambiguity fixing rate, $\boldsymbol{N}_{used_{ambiguities}}$ refers to the number of fixed ambiguities that are used for the coordinate estimations after the validation test, and $\boldsymbol{N}_{all_{ambiguities}}$ is the number of all ambiguities that have been found.

Table 8. Wrong fix within the first category for PPP

SITES		1 h	
		Number of wrong fix and float outlier	Mean fixing rate (wide- and narrow- lane)
ALGO	7°	-	-
	25°	Fix: 17 Float: 2	38 %
ALIC	7°	1	65 %
	25°	Fix: 29 Float: 12	80 %
BRAZ	7°	Fix: 1 Float: 0	75 %
	25°	Fix.27 Float: 0	69 %
CCJ2	7°	Fix:2 Float: 0	98 %
	25°	Fix:21 Float: 0	86 %
CAS1	7°	-	-
	25°	Fix: 4 Float: 0	80 %
CKIS	7°	Fix: 3 Float: 0	75 %
	25°	Fix: 18 Float: 4	81 %
MAS1	7°	Fix:4 Float:13	88 %
	25°	Fix:10 Float:3	51 %
MAW1	7°	Fix: 0 Float: 7	-
	25°	Fix: 3 Float: 0	75 %
NOVM	7°	-	-
	25°	Fix: 26 Float: 19	72 %
ROTH	7°	-	-
	25°	Fix: 4 Float: 0	79 %
SCOR	7°	-	-
	25°	Fix: 3 Float: 0	68 %
TIXI	7°	-	-
	25°	Fix: 7 Float: 4	70 %
ZAMB	7°	-	-
	25°	Fix: 7 Float: 12	78 %

Table 9. Wrong fix within the second category for PPP

SITES		1 h	
		Number of wrong fix	Mean fixing rate (wide- and narrow- lane)
ALGO	7°	8	40 %
	25°	84	26 %
ALIC	7°	63	29 %
	25°	130	30 %
ANKR	7°	11	86 %
	25°	78	74 %
BRAZ	7°	39	85 %
	25°	113	72 %
CAS1	7°	-	-
	25°	33	35 %
CCJ2	7°	23	66 %
	25°	78	68 %
CKIS	7°	103	65 %
	25°	115	51 %
MAS1	7°	47	80 %
	25°	99	69 %
MAW1	7°	11	59 %
	25°	54	64 %
NOVM	7°	15	78 %
	25°	191	69 %
NYA1	7°	-	-
	25°	1	49 %
ROTH	7°	-	-
	25°	31	48 %
SCOR	7°	2	76 %
	25°	78	51 %
TIXI	7°	-	-
	25°	111	65 %
ZAMB	7°	58	83 %
	25°	66	67 %

Table 10. Wrong fix within the first category for relative positioning

SITES		1 h		
		Number of wrong fix	Mean fixing rate	
			WL	NL
IENG	7°	-	-	-
	25°	Fix: 16 Float: 5	99 %	31 %
LAMA	7°	-	-	-
	25°	Fix: 2 Float: 3	87 %	32 %
MDVJ	7°	-	-	-
	25°	Fix: 8 Float: 5	100%	36 %
MEDI	7°	Fix: 0 Float:2	-	-
	25°	Fix: 21 Float: 14	100%	33 %
NYA1	7°	-	-	-
	25°	Fix: 1 Float:2	86 %	34 %
ROTH	7°	-	-	-
	25°	Fix: 0 Float: 1		

Wide-lane and narrow-lane fixing rates of outliers were given separately for relative positioning, but due to the software restrictions they were given a single mean value (wide and narrow-lane ambiguities) for PPP. The mean wide-lane and narrow-lane fixing rates of 1h non-outlier solutions using 7° and 25° cut-off angles were 95%/59% and 99%/58% for relative positioning, respectively. It is observed that the narrow-lane ambiguity fixing rate was significantly lower than the wide-lane ambiguity fixing due to its narrower wavelength. The mean wide and narrow-lane ambiguities fixing rates of 1h non-outlier solutions using 7° and 25° cut-off angles were 79 % and 70 % for PPP respectively. Wrong fix results show that fixing to the wrong integer ambiguities occurred more frequently for the 25° cut-off angle compared to the 7° cut-off angle for each technique. This situation can be explained by the fact that arch length of the satellites is significantly smaller for the 25° cut-off angle compared to the 7° cut-off angle, and this leads to unreliable ambiguity resolution for each satellite.

Table 11. Wrong fix within the second category for relative positioning

SITES		1 h		
		Number of wrong fix	Mean fixing rate	
			WL	NL
IENG	7°	45	98 %	34 %
	25°	210	99 %	18 %
LAMA	7°	12	98 %	24 %
	25°	118	95 %	46 %
MDVJ	7°	23	100 %	50 %
	25°	208	100 %	39 %
MEDI	7°	114	98 %	22 %
	25°	209	100 %	34 %
NYA1	7°	5	89 %	28 %
	25°	87	86 %	77 %
ROTH	7°	4	100 %	57 %
	25°	57	100 %	55 %

5. CONCLUSION

Ambiguity resolution impacts on estimated coordinates for static PPP and relative positioning techniques are presented in this paper. Seventeen globally distributed stations and six baselines with varying lengths were chosen for conducting PPP and relative positioning respectively for a 31-day period in January 2018. Daily RINEX files of the stations were subdivided into mutually non-overlapping sessions as 24-12-6-4-2-1 h subsets. Ambiguity resolution impacts on the estimated coordinates were tested using 7° and 25° cut-off angles and final/ultra-rapid orbit and clock products. The results can be summarized as follows:

Ambiguity resolution impacts on the estimated coordinates increases as observation time decreases for PPP and relative positioning techniques respectively.

Using orbit/satellite products with differences in latency affects AR impacts on the estimated coordinates for PPP, especially for short observation sessions. However, it does not significantly affect the estimated coordinates for relative positioning due removal of the satellite clock bias in double difference. The biggest difference found between the products was 46 mm (2D) and 54 mm (up) in NOVM station for PPP processing.

The results reveal that accuracy degradation of AR compared to the float solutions on the estimated coordinates are generally seen for up component when using the 25° cut-off angle for PPP and relative positioning. Maximum positive and negative ambiguity resolution improvements in horizontal and vertical components were 54 mm (1h, 7°) / 12 mm (1h, 25°), -4 mm (4h, 25°) / -9 mm (1h, 25°) and 15 mm (4h, 7°) / 10 mm (2h, 25°), -2 mm (12h, 25°) / -11 mm (2h, 25°) for PPP and relative positioning respectively.

Fixing to the wrong integer ambiguities was investigated in two categories; first category was applied as when any fixed solution is within the outliers (errors greater than 30 cm), but its float solution is not an outlier, and the second category was applied as when the fixed solution which its two-dimensional error was 20 mm higher than the error of the float solution. In accordance with the results, fixing to the wrong integer ambiguities

are accumulated generally in 1h solutions for each technique. The highest amount of fix outliers in the second category were found 26% and 29% of all data sets for PPP and relative positioning techniques with 25° cut-off angle.

The results also reveal that fixing to the wrong integer ambiguities occurred more frequently for the 25° cut-off angle compared to the 7° cut-off angle for each technique due to the satellite arch length being inadequate to resolve ambiguities reliably. In accordance with the outlier results of the first category, fix outliers of PPP are 36% and 75% of all fix and float outliers using 7 and 25-degree cutoff angle respectively. Fix outliers of relative positioning are 0% and 60% of all fix and float outliers using 7 and 25-degree cutoff angle respectively.

If PPP or relative technique need to be applied with a short observation period (for example, 1h) under the obstructed conditions, the float solution can be chosen for a higher likelihood of avoiding the wrong fix solution. In terms of using AR under the poor satellite geometry with short observation time, partial ambiguity fixing or changing the AR validation might be conducted to avoid the wrong fix solution.

No direct relationship was observed between the baseline length and the coordinates' improvement of AR compared to the float solution.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to NASA Jet Propulsion Laboratory (JPL) and Massachusetts Institute of Technology (MIT) for providing GIPSY/OASIS and GAMIT/GLOBK scientific software.

REFERENCES

Alcay, S., Ogutcu, S., Kalayci, I and Yigit, C.O. (2019). Displacement monitoring performance of relative positioning and Precise Point Positioning (PPP) methods using simulation apparatus, Advances in Space Research 63, 5, 1697–1707.

Bar‐Sever, Y. E., Kroger, P. M., and Borjesson, J. A. (1998). Estimating horizontal gradients of tropospheric path delay with a single GPS receiver. Journal of Geophysical Research: Solid Earth, 103(B3), 5019-5035.

Blewitt G., Young LE., and Meehan TH. (1989). Sub-centimeter baselines within seconds using ROGUE receivers: introducing the rapid static survey (RSS) method. In: International association of geodesy conference, Edinburgh, August, 1989.

Bezcioglu M., Yigit, O. C and El-Mowafy, A. (2019). Kinematic PPP-AR in Antarctic: Comparing Methods for Precise Positioning. Sea Technology.

Bertiger, W., Desai, S. D., Haines, B., Harvey, N., Moore, A. W., Owen, S and Weiss, J. P. (2010). Single receiver phase ambiguity resolution with GPS data. Journal of Geodesy, 84(5), 327-337.

Brach, M and Zasada, M. (2014). The effect of mounting height on GNSS receiver positioning accuracy in forest conditions. Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering, 35(2), 245-253.

Chen, H., Xiao, Y., Jiang, W., Zhou, X and Liu, H. (2017). An improved method for multi-GNSS baseline processing using single difference. Advances in Space Research.

Choy, S., Bisnath, S and Rizos, C. (2017). Uncovering common misconceptions in GNSS Precise Point Positioning and its future prospect. GPS solutions, 21(1), 13-22.

Gandolfi, S., Tavasci, L and Poluzzi, L. (2017). Study on GPS–PPP precision for short observation sessions. GPS Solutions, 21(3), 887-896.

Ge, M., Gendt, G., Rothacher, M. A., Shi, C and Liu, J. (2008). Resolution of GPS carrier-phase ambiguities in precise point positioning (PPP) with daily observations. Journal of Geodesy, 82(7), 389-399.

Geng, J., Teferle, F. N., Shi, C., Meng, X., Dodson, A. H and Liu, J. (2009). Ambiguity resolution in precise point positioning with hourly data. GPS solutions, 13(4), 263-270.

Geng, J., Jiang, P and Liu, J. (2017). Integrating GPS with GLONASS for high‐rate seismogeodesy. Geophysical research letters, 44(7), 3139-3146.

Goudarzi, M. A and Banville, S. (2018). Application of PPP with ambiguity resolution in earth surface deformation studies: a case study in eastern Canada. Survey Review, 50(363), 531-544.

Joosten, P. (2000). Fixing the ambiguities-are you sure they're right?. GPS world, 11(5), 46-51.

Lagler, K., Schindelegger, M., Böhm, J., Krásná, H and Nilsson, T. (2013). GPT2: Empirical slant delay model for radio space geodetic techniques. Geophysical research letters, 40(6), 1069-1073.

Li, X., Ge, M., Zhang, X., Zhang, Y., Guo, B., Wang, R ... and Wickert, J. (2013). Real‐time high‐rate co‐seismic displacement from ambiguity‐fixed precise point positioning: Application to earthquake early warning. Geophysical Research Letters, 40(2), 295-300.

Li, P., Zhang, X., Ren, X., Zuo, X and Pan, Y. (2016). Generating GPS satellite fractional cycle bias for ambiguity-fixed precise point positioning. GPS solutions, 20(4), 771-782.

Li, X., Chen, X., Ge, M and Schuh, H. (2018). Improving multi-GNSS ultra-rapid orbit determination for real-time precise point positioning. Journal of Geodesy, 1-20.

Li, X., Li, X., Liu, G., Feng, G., Guo, F., Yuan, Y and Zhang, K. (2018). Spatial–temporal characteristic of BDS phase delays and PPP ambiguity resolution with

GEO/IGSO/MEO satellites. GPS Solutions, 22(4), 123.

Pan, L., Cai, C., Santerre, R and Zhu, J. (2014). Combined GPS/GLONASS precise point positioning with fixed GPS ambiguities. Sensors, 14(9), 17530-17547.

Pehlivan, H., Bezcioğlu, M and Yılmaz, M. (2019). PERFORMANCE OF NETWORK RTK CORRECTION TECHNIQUES (FKP, MAC and VRS) UNDER LIMITED SKY VIEW CONDITION. International Journal of Engineering and Geosciences, 4(3), 106-114.

Petit, G.; Luzum, B. IERS Conventions (2010). Bureau International des Poids et Mesures Sevres (France), 2010. Available online: https://www.iers.org/IERS/EN/Publications/TechnicalNotes/tn36
.html (accessed on 1 July 2018).

Teunissen, P. J. G and Khodabandeh, A. (2015). Review and principles of PPP-RTK methods. Journal of Geodesy, 89(3), 217-240.

Teunissen, P and Montenbruck, O. (Eds.). (2017). Springer handbook of global navigation satellite systems. Springer.

Tusat, E and Ozyuksel, F. (2018). Comparison of GPS satellite coordinates computed from broadcast and IGS final ephemerides. International Journal of Engineering and Geosciences, 3(1), 12-19.

Schwarz, C. R., Snay, R. A and Soler, T. (2009). Accuracy assessment of the National Geodetic Survey's OPUS-RS utility. GPS solutions, 13(2), 119-132.

Shi, C., Lou, Y. D., Zhang, H. P., Zhao, Q., Geng, J., Wang, R. ... and Liu, J. (2010). Seismic deformation of the Mw 8.0 Wenchuan earthquake from high-rate GPS observations. Advances in Space Research, 46(2), 228-235.

Wielgosz, P. (2011). Quality assessment of GPS rapid static positioning with weighted ionospheric parameters in generalized least squares. GPS solutions, 15(2), 89-99.

Wu, J. T., Wu, S. C., Hajj, G. A., Bertiger, W. I and Lichten, S. M. (1992, August). Effects of antenna orientation on GPS carrier phase. In Astrodynamics 1991 (pp. 1647-1660).

Xiao, G., Li, P., Sui, L., Heck, B and Schuh, H. (2019). Estimating and assessing Galileo satellite fractional cycle bias for PPP ambiguity resolution. GPS Solutions, 23(1), 3.

Yigit C.O, Gikas V., Alcay S and Ceylan A. (2014) Performance evaluation of short to long term GPS, GLONASS and GPS/GLONASS post-processed PPP, Survey Review, 46(3), 155-166.

Zumberge, J. F., Heflin, M. B., Jefferson, D. C., Watkins, M. M and Webb, F. H. (1997). Precise point positioning for the efficient and robust analysis of GPS data from large networks. Journal of geophysical research: solid earth, 102(B3), 5005-5017.

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 094-099, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.614856

A NOVEL APPROACH FOR IONOSPHERIC TOTAL ELECTRON CONTENT EARTHQUAKE PRECURSOR AND EPICENTER DETECTION FOR LOW-LATITUDE

Santanu Kalita[1]*, Bornali Chetia[2]

[1]Department of Computer Application, Mahapurusha Srimanta Sankaradeva Viswavidyalaya, Assam, India
(santanu.kalita@rediffmail.com); **ORCID 0000-0003-2715-9225**
[2] Department of Physics, Royal Global University, Assam, India
(bornalichetia.physics@gmail.com); **ORCID 0000-0002-9997-0300**

*Corresponding Author, Received: 03/09/2019, Accepted: 21/09/2019

ABSTRACT: The earthquake precursory phenomena detection using ionospheric perturbation characteristics is a new technique used by the scientist now days. This paper focuses a new technique for detecting any modification in the time series profile shape caused by an impending earthquake to identify precursors as well as an image processing technique for epicenter detection. For this purpose IGS Global Navigation Satellite System (GNSS) Total Electron Content Data (TEC) are utilized from different stations across the world. From the experiment it is observed that the method may detect earthquake precursors a few hours or days prior to the main event due to ionospheric perturbations induced by initiation of earthquake process.

Keywords: *GNSS, Total Electron Content, ionosphere*

1. INTRODUCTION

The ionospheric variability has been used as a precursor marker for an impending earthquake by several researchers around the globe (ChuoY.J.Y, 2001; J. Y. Liu 2000; Yuri Ruzhin, 2007). The ionospheric irregularity and seismic event relations are detected through various parameters like fof2, ionospheric density measurement, very high frequency, Total Electron Content (TEC) etc. Out of these parameters TEC data are widely used for earthquake precursor studies. TEC is the measure of the total amount of electrons along a particular line of sight, i.e., from the satellite to the receiver and the unit of total electron content is TECU where 1TECU=10^{16} electrons/m2. The Global Positioning System (GPS) is the popular system which is used in estimation of TEC which provides the estimate for the total number of electrons on the receiver-satellite link. The TEC of the ionosphere varies with space and time and the behavior depended on other geomagnetic activities like geomagnetic storm, sun solar activity and the location of receiving station. During the earthquake preparatory process, due to the movement of tectonic plates positive ions are produced which travels through troposphere to ionosphere and joins electrons which leads to high TEC values. Several studies have reported the GPS derived TEC variations before the occurrence of large earthquakes (E. Calais et al, 1995; G K. Davies et al, 1965; M. Hayakawa, 1999; V. N. Oraevsky et al; S. Kalita, 2015). Some of important cases include 17 August 1999 of magnitude 7.6M Izmit earthquake, 12 May 2008 8.0M magnitude Wenchuan earthquake, 12 January 2010 M=17 Haiti earthquake and 11 March 2011 Mw 9 Tohoku earthquake which describes anomalous TEC variation before the earthquake events (Dogan, U et al, 2011; Zhang, 2008; Liu, 2011, Dimitar, 2011). In India, attempts have been made by S Kalita (S Kalita et al., 2015), Dutta (Dutta et al., 2007) and Priyadarshi, (Priyadarshi et al. 2011) to find the relation between TEC variations and seismic activities.

2. DATA AND ANALYSIS

It has been observed that by analyzing ionospheric TEC, one can detect earthquake precursors in a few hours or days prior to the main event due to ionospheric perturbations induced by initiation of earthquake process. For this purpose, in this work, the nearest GNSS stations of the earthquake events described in Table 1 are considered for analyzing and processing. The International GNSS Service (IGS) is located in different parts of the world and various data analysis centers provides high-quality GPS data and data products in near real time for scientific and engineering works. The data from Crustal Dynamics Data Information System (CDDIS), National Aeronautics and Space Administration (NASA) were used for TEC estimation in raw format and then it is pre-processed and converted to RENIX format and processed to obtain TEC values using GPS-TEC program (C. Noll, 2010).

Date of earthquake event	Epicentre	Magnitude	Kp	Nearest Station
21/09/2009	27.32° N, 91.42° E	5 Mw	13	Lasha, Chaina
10/08/2010	14.013°N 92.923°E	7.5 Mw	6	Lasha, Chaina
13/04/2009	33.165°N , 96.629°E	7.1 Mw	5	Andaman and Nicobar Islands

Table 1: Earthquake epicentres and magnitudes of the cases analyzed with pattern matching and image processing technique

For image processing technique for earthquake epicenter detection the Ionosphere Research Laboratory (IONOLAB) global TEC map data has been utilized. The TEC estimation method of IONOLAB group is the one of the most important contributions for ionospheric mapping and electron density reconstruction (H. Tuna et al, 2014) regional model-based computerized ionospheric tomography using GPS measurements.
It is well known that ionospheric characteristics depends on different seasonal and variations of geomagnetic conditions. To study the seasonal variations of TEC, the seasonal median for different seasons are considered. From the experiment it is observed that due to the meridional wind flow during daytime from pole to the equator decrease the natural composition and O/N_2 ratio at the low latitude and results higher during equinox followed by the winter and lowest TEC in the summer. In the next step, for filtering unwanted geomagnetic influence on TEC variations of geomagnetic parameter Kp is considered for analysis for both high and low solar activity period. Based on Kp index typical quiet day and disturb day TEC data filtered out for both solar activity period. From the experiment it is also observed that during high solar activity period changes in the profile shape are clearly manually detectable but such a modification partially not visible during low solar activity period prior to an impending earthquake.

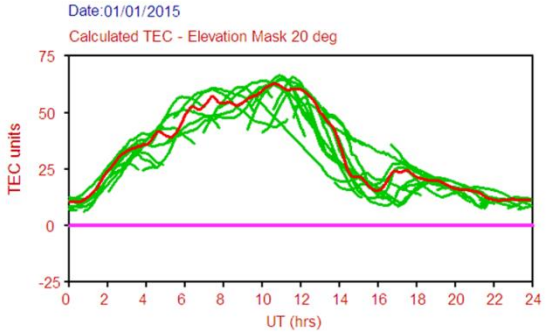


Figure 1: A typical TEC profile from IISC Bangalore station

It is noted that pre earthquake changes in TEC value at low latitude station occurs not only during the noon

periods but also at post noon hours. Therefore it is necessary to analyze the seismic features of TEC for the entire day. To check the pattern of TEC profile pattern matching technique based on PIP is considered for earthquake time deviation in the entire TEC data for a day relation to quiet day. For this a template is formed for by selecting TEC profiles of quiet days for each month to filter out the geomagnetically disturbed factors. Earlier research reports that the anomaly enhancement of TEC peak occurs in not only during noon periods but also at post noon and after sunset hour prior to an impending earthquake (M Devi et al, 2012), which is an important mark along with the shape of profile in the new proposed method. The proposed pattern matching method for earthquake prediction is based on perceptually important point (PIP) identification. The whole method can be divided into two phrases, i.e. PIP identification and similarity measure. For the experiment, first compute the vertical total electron content ΔVTEC for each day following formula:

$$\mathbf{\Delta VTEC} = \left(\frac{\mathbf{\Delta\rho}}{\mathbf{40.3}}\right)\left(\frac{\mathbf{{f_1}^2 {f_2}^2}}{\mathbf{{f_1}^2 - {f_2}^2}}\right)$$

where $\mathbf{\Delta\rho}$ is the difference between time delays measured by the L1 and L2, f_1 is the frequency of L1 wave and f_2 is the frequency of L2 wave.

Now for each month construct a template by taking moving average of 60 points. Now, compare the daily TEC data with the template for identification of earthquake precursor as follows:

At the first step compute Perceptually Important Points (PIP) for both the profiles. The first two PIPs will be the first and last points of the series say p1 = (x1, y1) and p2 = (x2, y2) respectively. Measure the sum of the Euclidian distances (ED) of the test point say p3 = (x3, y3) to its adjacent PIPs p1 =(x1, y1) and p2 =(x2, y2), i.e. $\mathbf{ED} = \sqrt{(\mathbf{x_2 - x_3})^2 + (\mathbf{y_2 - y_3})^2} + \sqrt{(\mathbf{x_1 - x_3})^2 + (\mathbf{y_1 - y_3})^2}$

The second measure is the perpendicular distance (PD) between the test point p3 and the line connecting the two adjacent PIPs as shown in

$$\mathbf{Slope}(\mathbf{p_1}, \mathbf{p_2}) = \mathbf{s} = \frac{\mathbf{y_2 - y_1}}{\mathbf{x_2 - y_1}}$$

$$\mathbf{x_c} = \frac{\mathbf{x_3 + (sy_3) + (s^2 x_2) - (sy_2)}}{\mathbf{1 + s^2}} - \mathbf{(x_3)^2}$$

$$\mathbf{y_c} = \mathbf{(sx_c) - (sx_2) + y_2}$$

$$\boldsymbol{PD}(\mathbf{p_3}, \mathbf{p_c}) = \sqrt{(\mathbf{x_c - x_3})^2 + (\mathbf{y_c - y_3})^2}$$

The final measure, is the vertical distance (VD) between the test point p3 and the line connecting the two adjacent PIPs, i.e

$$\boldsymbol{VD}(\boldsymbol{p_3}, \boldsymbol{p_c}) = |\boldsymbol{y_c} - \boldsymbol{y_3}| = |\left(\boldsymbol{y_1} + (\boldsymbol{y_2} - \boldsymbol{y_1})\frac{\boldsymbol{x_c} - \boldsymbol{x_1}}{\boldsymbol{x_2} - \boldsymbol{x_1}}\right) - \boldsymbol{y_3}|$$

where $x_c = x_3$. After identifying the PIPs of the subsequence, a similarity measuring mechanism is essential for pattern matching. As different sequences may have different ‘‘amplitudes’’, after identifying the PIPs in the data sequence, it is necessary to re-scale the points so that the comparison between sequences in different ‘‘amplitudes’’ range (e.g. 0–1) can be facilitated. This is typically addressed as normalizing all the sequence values to a given range. Then, the amplitude distance (AD) between P and Q can be computed by using point-to-point direct comparison, i.e.,

$$\boldsymbol{AD}(\boldsymbol{SP}, \boldsymbol{Q}) = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} (\boldsymbol{sp_k} - \boldsymbol{q_k})^2}$$

Here, SP and sp_k denote the PIPs found in P. Now, consider the horizontal distortion of the pattern against the pattern templates. The temporal distance (TD) between P and Q is defined as

$$\mathbf{TD}(\mathbf{SP}, \mathbf{Q}) = \sqrt{\frac{\mathbf{1}}{\mathbf{n-1}}\sum_{\mathbf{k=2}}^{\mathbf{n}} (\mathbf{sp^t_k} - \mathbf{q^t_k})}$$

where sp^t_k and q^t_k denote the time coordinate of the sequence points sp_k and q_k, respectively. To take both horizontal and vertical distortion into consideration in our similarity measure, the distance (or similarity) measure could be modified as

$$\mathbf{D}(\mathbf{SP}, \mathbf{Q}) = \mathbf{w_1}.\mathbf{AD}(\mathbf{SP}, \mathbf{Q}) + (\mathbf{1} - \mathbf{w_1}).\mathbf{TD}(\mathbf{SP}, \mathbf{Q})$$

In our next method we shall utilize an image segmentation technique for epicenter detection. Before the formal processing of the image, we perform some necessary preprocessing on the image to meet the requirements of the subsequent steps and achieve faster and better segmentation accordingly. The basic k-mean algorithm for image segmentation is discussed below:

Let us consider an image with resolution of x×y and the image has to be cluster into k number of cluster. Let p(x, y)be an input pixels to be cluster and c_k be the cluster centers. The algorithm for k-means clustering is following as:

1. Initialize number of cluster k and centre.
2. For each pixel of an image, calculate the Euclidean distanced, between the center and each pixel of an image using the relation given below.

$$\mathbf{d} = ||\mathbf{p}(\mathbf{x}, \mathbf{y}) - \mathbf{c_k}||$$

3. Assign all the pixels to the nearest centre based on distanced.
4. After all pixels have been assigned, recalculate new position of the centre using the relation given below

$$\mathbf{c_k} = \frac{\mathbf{1}}{\mathbf{k}}\sum_{\mathbf{y}\in\mathbf{c_k}}\sum_{\mathbf{x}\in\mathbf{c_k}} \mathbf{p}(\mathbf{x}, \mathbf{y})$$

5. Repeat the process until it satisfies the tolerance or error value.
6. Reshape the cluster pixels into image

3. RESULTS AND DISCUSSIONS

With the help of the above algorithms, now we shall proceed towards testing its application as a precursor and epicenter detection in few of earthquake cases. The earthquake events are considered so that the observing stations lies within the preparatory zone. In the first case, let us consider the earthquake event occurred in Bhutan (epicenter position: 27.320 N, 91.420 E) held on 21 September 2009 of magnitude M= 6.1 occurred at 8.53UT. For this purpose, the nearest IGS station to the epicenter Lasha, Chaina (29.6548° N, 91.1406° E) is considered for TEC data analysis. Figure 2 represents the TEC profile from 17 September to 21 September 2009, i.e. prior to and during earthquake. From the profile, it can be clearly observed that the TEC profile shape pattern is fully different in 18 September i.e. three days before the earthquake event.

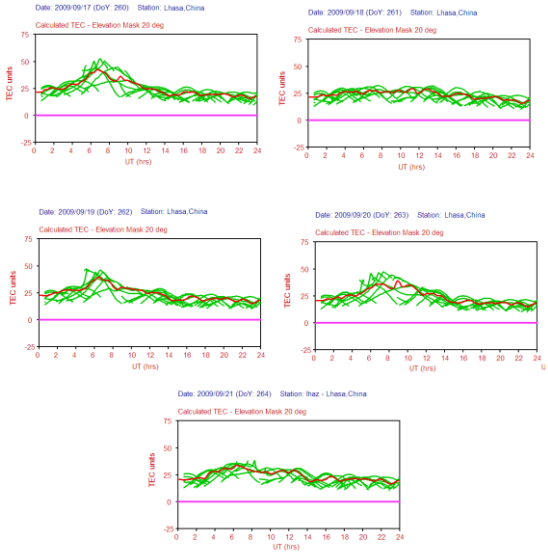
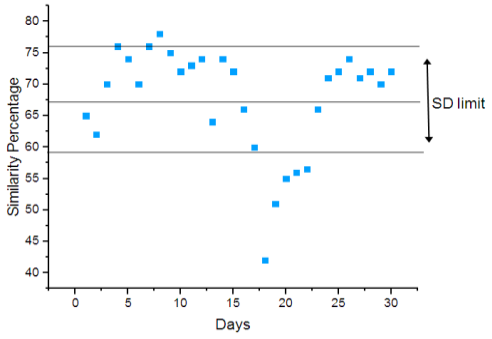
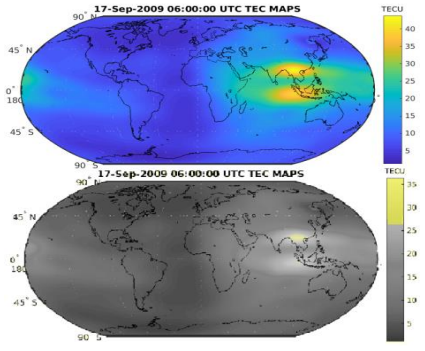


Figure 2: TEC profile shape from 17-21 September 2009 prior to and during Bhutan earthquake Station Lasha, Chaina

Now, for analyzing the event using the proposed method for similarity measure, one template is framed by using at least five quietest days of that month. The test is performed with a calculation of 'similarity percentage' for percentage matching of two time series and it is expected that the factor should be low before the earthquake event. The result is presented in figure 3. The standard deviation of the 'similarity percentage' is also calculated for obtaining its maximum possible fluctuations for this month.



(a)

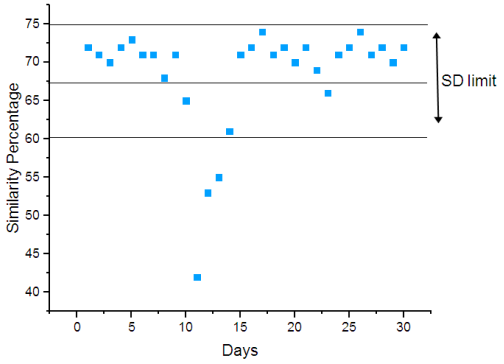


(b)

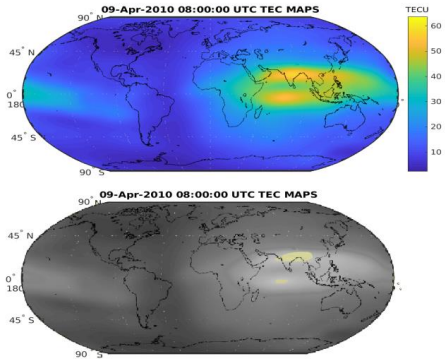
Figure 3: (a) 'Similarity percentage' for the month of September 2009, (b) Clustering result for 17 September 2009

From the above result (Figure 3a) one can clearly observe a clear drop of this factor prior to the earthquake. The second method using k-mean clustering results (Figure 3b) for epicenter detection produces a new cluster covering Lat 20-30^{o} Lon 88-120^{0}. The cluster analysis is performed for each day prior to the earthquake and a better result is produced on September 17, 2009 five days prior to the event. The earthquake epicenter position 27.32^{0} N, 91.42^{0} E is fall inside the above grid.

To test the reliability of the above algorithms, the very strong earthquake of Yushu, China (epicentre at 33.165°N 96.629°E) occurred on 13 April 2010 of magnitude M=7.1 is considered for analysis, Lasha, Chaina lies well within the earthquake preparatory zone and is TEC data from this station considered for analysis. The Q-day template has been made for the month again and the 'Similarity percentage' is calculated and plotted in figure 4(a). Here too, one can easily observes a clear drop of similarity factor prior to the earthquake. The low values after the certainty factor after the earthquake event are interesting and may be associated with aftershocks. The clustering results (Figure 4b) also focus that there may be an earthquake event between Lat 20-36^{o} Lon 86-100^{o}. Thus, both the two results of utilizing the algorithms for earthquake precursor and epicenter identification in TEC data are encouraging.



(a)

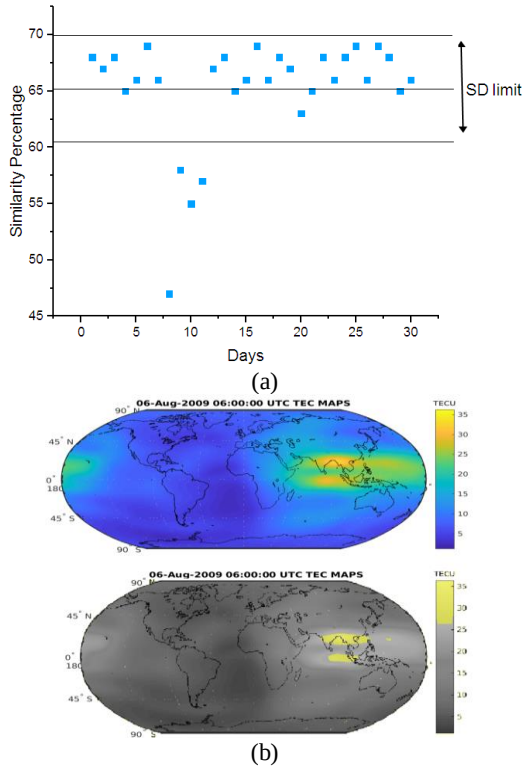


(b)

Figure 4: (a) 'Similarity percentage' for the month of April 2010, (b) Clustering result for 09 April 2010

The above technique is now applied for identifying the earthquakes of strong earthquake of 10 August, 2009 of Andaman Islands. The magnitude of the earthquake was M=7.5 and was one of the strongest earthquake of Indian Ocean. For the analysis the data is collected from

IGS Andaman and Nicobar Islands GPS station is considered and Q-day template has been made for the month accordingly. Here too, one can easily see a clear drop of similarity factor prior to the earthquake (Figure 5a). In this case, the clustering results produce two clusters and one is associated with 10 August, 2009 earthquake event and later one may be due to some other factor related to some geomagnetic events (Figure 5b). It is important to note that in all cases the profile shape is not influenced by increased electron density.



(a)

(b)

Figure 5: (a) 'Similarity percentage' for the month of August 2009, (b) Clustering result for 10 August 2009

Depueva reports that the analysis of TEC variation before an impending earthquake have shown that precursors at low-latitude stations are fully detectable as a kind of Appleton type EXB charge movement when the earthquake epicenter lies near to the equator (Depueva et.al., 1995). Again, the profile behavior of electron content as precursor imposed by an equatorial earthquake and low latitude earthquake may not be similar, since anomaly crest shift may lead to both increase and decrease of TEC, which depends on magnitude, closeness of the epicenter to the observing station. Past studies suggests that the TEC peak value analysis for prediction of an impending earthquake which is not reliable most of the time and produces false prediction results since the peak value is also associated with other factors (Devi et al, 2012). The Dynamic time warping method has also been applied to identify TEC anomalies related to earthquakes with low latitude TEC data (Devi et al., 2012), but this method suffers from the limitation in that it cannot resolve fine structure. On this background, an alternative method i.e pattern matching technique based on PIP for analysis of TEC data and an image processing technique for earthquake detection is applied successfully and earthquake precursor identification and detection of epicenter position in TEC using these methods are encouraging, though there are scopes for improvement (like without taking the moving average of the data), which will be included in our future studies.

4. CONCLUSION

In this paper an attempt is made for detecting earthquake cursor using TEC data using a pattern recognition algorithm, where variations in the entire 24-hour profile of TEC are taken and an image processing method for epicenter detection. It is seen that the 'similarity factor' in the a pattern recognition technique has provided a tool for marking a cursor of moderate to strong earthquakes, when the observing station lies around the earthquake preparatory zone. The results are an improvement on prediction capability, compared to earlier Dynamic Time Series algorithm, since the earlier method has some limitations that it cannot resolve finer temporal and spatial information because of the enormous increase in size of the necessary matrix. The image processing technique based on k means constructs a cluster near to the epicenter. Our next attempt will be to make the technique for epicenter detection more efficient through use of multiple features of TEC such as azimuth and elevation angle position of the satellite and using other ionospheric parameter.

REFERENCES

Chen, M.Q. (2011) Observations and simulations of seismoionospheric GPS total electron content anomalies before the 12 January 2010 M7 Haiti earthquake. J. Geophys. Res. 116 (A04302). 1e9.

ChuoY.J.Y, .I.Chen, J.Y.Liu, S.A. Pulinets (2001) Ionospheric foF2 variations prior to strong earthquakes in Taiwan. Advances in Space Research. https://doi.org/10.1016/S0273-1177(01)00209-5. Volume 27 Issues 6–7 Pages 1305-1310

C. Noll, The Crustal Dynamics Data Information System (2010) A resource to support scientific analysis using space geodesy, Advances in Space Research,Volume 45, Issue 12, Pages 1421-1440, ISSN 0273-1177, DOI: 10.1016/j.asr.2010.01.018.

Depueva, A. H. And Ruzhin, Yu. Ya (1995) Seismoionospheric fountain effect as analogue of active space experiment. Advances in Space Research. 15. pp. 151–154.

Dimitar, O., Pullinets, S., Alexey, R.A., Konstantin, T., Dimitri, D., Menas, K., Patrick, T. (2011) Atmosphere-ionosphere response to the M9 Tohoku earthquake revealed by multi instrument space-borne and ground observations: preliminary results. Earthq. Sci. 24. 1-7.

Dogan, U., Ergintav, S., Skone, S., Arslan, N., Oz, D. (2011) Monitoring of the ionosphere TEC variations during the 17th August 1999 Izmit earthquake using GPS data. Earth Planets Space. 63. 1183-1192.

Dutta, H.N., Dabas, R.S., Das, M.R., Sharma, K., Singh, B., (2007) Ionospheric perturbations over Delhi caused

by the 26 December 2004 Sumatra earthquake. Int. J. Remote Sens. 28 (13-14). 3141-3151

E. Calais and J. B. Minster (1995) GPS Detection of Iono- spheric Perturbations Following the January 17, 1994, Northridge Earthquake. Geophysical Research Letters. Vol. 22. No. 9, pp. 1045-1048. doi:10.1029/95GL00168

G K. Davies and D. M. Baker (1965) Ionospheric Effects Ob- served Round the Time of Alaska Earthquake of March 28, 1964. Journal of Geophysical Research. Vol. 70. No. 9. pp. 2251-2263. doi:10.1029/JZ070i009p02251

J. Y. Liu, Y. I. Chen, S. A. Pulinets, Y. B. Tsai, Y. J. Chuo (2000) Seismo-ionospheric signatures prior to M>= 6.0 Taiwan earthquakes. Geophysical Research Letters. Vol. 27. No. 19. October 1. Pages 3113-3116

M Devi, AJD Sarma, S Kalita, AK Barbara, A Depueva, (2012) Adaptive techniques for extraction of pre-seismic parameters of Total Electron Content (TEC) at anomaly crest station. Geomatics, Natural Hazards and Risk. 3 (3). 193-206

M. Hayakawa (1999) Atmospheric and Ionospheric Electro- magnetic Phenomena Associated with Earthquakes. Terra Scientific Publishing Co, Tokyo

M. Hayakawa, O. A. Molchanov, T. Ondoh and E. Kawai (1996) The Precursory Signature Effect of the Kobe Earthquake on VLF Sub Ionospheric Signals. Journal of Communications Research Laboratory Tokyo. Vol. 43. 1996 pp. 169-180.

Priyadarshi, S., Kumar, S., Singh, A.K. (2011) Changes in total electron content associated with earthquakes (M>5) observed from GPS station Varanasi, India. Geomatics, Nat. Hazard Risk. 2 (2). 123e139.

S. Kalita (2015) Monitoring the TEC variation using pattern matching method during earthquakes as determined from ground based TEC measurement and satellite data. Int. J. Sci. Res. Publications, vol. 5. no. 6. pp. 4.

Tuna H., O. Arikan, F. Arikan, T.L. Gulyaeva, and U. Sezen (2014) Online User-Friendly Slant Total Electron Content Computation from IRI-Plas: IRI-Plas-STEC. Space Weather. 12(1). 64-75. doi:10.1002/2013SW000998.

V. N. Oraevsky, Y. Y. Ruzhin and I. I. Shagimuratove (2000) Anomalies of Ionospheric TEC above Turkey before Two Strong Earthquakes. Proceedings of 15th Wroclaw EMC Symposium. Brugge. pp. 508-512.

Yuri Ruzhin, Costas Nomicos (2007) Radio VHF precursors of earthquakes. Nat Hazards. 006-9021-1. 40:573–583
DOI 10.1007/s11069

Zhang, Zhengh & Liu, J & Sun, J & Wen, L & Tapponnier, P & Xing, X & Xu, Qun & Hu, Guyue & Zeng, Lingsen & Ding, Lin & Ji, Chen & Hudnut, Kenneth (2008) Co-seismic ruptures of the 12 May, 2008, Mw 8.0 Wenchuan earthquake, Sichuan: EW crustal shortening on oblique, parallel thrusts along the eastern edge of Tibet. AGU Fall Meeting Abstracts. 1. 0054

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 100-108, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.623592

A RULE-BASED APPROACH FOR GENERATING URBAN FOOTPRINT MAPS: FROM ROAD NETWORK TO URBAN FOOTPRINT

Müslüm Hacar [1]*

[1]Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul, Turkey (mhacar@yildiz.edu.tr); **ORCID 0000-0002-8737-8262**

*Corresponding Author, Received: 23/09/2019, Accepted: 12/11/2019

ABSTRACT: Decision and policymakers need urban footprint data for monitoring human impact on the urban ecosystem for politics and services. Deriving urban footprint is a challenging work since it has rapidly changing borders. The existing methods for deriving urban footprint map based on raster images have several steps such as determination of indicators and parameters of image classification. These steps limit the process by an operator since they require human decisions. This paper proposes a new rule-based approach for obtaining urban footprint based on Delaunay triangulation among selected centroids of roads and dead-end streets. The selection criterion is determined as maximum road length by using standard deviation operator. To produce urban footprints, this method needs no other data or information apart from road network geometry. This means that the proposed method uses only intrinsic indicators and measures. The experimental study was conducted with OpenStreetMap road data of Washington DC, Madrid, Stockholm, and Wellington. The comparisons with authority data prove that the proposed method is sufficient in many parts of urban and suburban lands.

Keywords: *OpenStreetMap, Urban footprint, TIN, Road network analysis, Land use*

1. INTRODUCTION

Natural and artificial entities on Earth surface differ from each other with regards to living area of human. While natural entities are composed of any habitat such as forest, mountain, and sea, an artificial construction maybe a tribal temple in a village or a skyscraper, highway, train station in a megacity. However, the differences between natural and artificial entities are not as distinct since there are several semi-natural and semi-artificial area in urbanised/sub-urbanised lands such as a park, open space, zoo, dam or farming land. For these purposes,land use/cover processes classify the Earth surface into its urban, suburban, green open place, forest, dam and/or sea patterns. A land-use map, thematically proposed to represent the urbanised area, is called "urban footprint map". Urban footprint represents the border of human impact on the land surface. Determination of urban footprints in a specified land surface is a challenging issue since human activity varies from villages to metropole cities. Some measurements used for producing spatially significant urban footprint map are common with urban sprawl measurements. Many methods in the literature use remote sensing imagery as the major data source for analysing and predicting urban growth, with several classifications and indicators (Musa et al. 2017; Karakuş et al. 2017; Canaz Sevgen, 2019). Tsai (2005) developed a set of quantitative variables (i.e. metropolitan size, density, degree of equal distribution and clustering) to characterise urban forms at the municipal level. Angel et al. (2007) mentioned the metrics used for measuring attributes of sprawl. They define urban extents as built-up area (derived from satellite imagery), urbanised area (built-up area + urbanised open space), urbanised open space, urban footprint (built-up area + urbanised open space + peripheral open space) and peripheral open space. Bhatta et al. (2010) described spatial metrics used to quantify the urban sprawl with some examples such as area-density-edge metrics, shape metrics, isolation-proximity metrics, contrast metrics, contagion-interspersion metrics, connectivity metrics, and diversity metrics. They claimed that the measurement of sprawl from remote sensing data is still in its research domain.

Jiang et al. (2007) measured urban sprawl from the spatial configuration, urban growth efficiency and external impacts. They developed a geo-spatial indices system for measuring sprawl, a total of 13 indicators. For this study, they used any different data sources including land use maps, former land use planning, land price and floor-area-ratio samples, digitised map of the highways and city centres, population and statistical data. Triantakonstantis and Stathakis (2015) used spatial indicators to calculate urban morphological properties such as shape, aggregation, compactness and dispersion. They applied the indicators to the urban areas in order to measure urban sprawl in 24 European countries.

Esch et al. (2013) presented a fully automated processing system for the delineation of human settlements worldwide based on the synthetic aperture radar (TanDEM-X). They assessed the high potential of the TanDEM-X data and the proposed urban footprint processor to provide highly accurate geo-data for improved global mapping of human settlements.

Entropy method is one of the widely used techniques to measure the extent of urban sprawl with the integration of remote sensing and geographic information system (GIS) (Kumar et al. 2007; Bhatta et al. 2010). Kumar et al. (2007) used buffer zones with Shannon entropy to determine the spatial concentration or dispersion of built-up land. They integrated the observations with the road network to check the influence of infrastructure on haphazard urban growth.

Road networks were also used to generate urban maps. Owen and Wong (2013) used road networks, shape, terrain geomorphology, texture and dominant settlement materials (vegetation, soil, asphalt) to distinguish informal neighbourhoods from formal ones in developing countries. They used dangling nodes to determine the ratio of connected nodes. A high value in the ratio implies that roads are better connected, and multiple routes between endpoints exist. Liu and Jiang (2011) stated that the dangling lines inside a block define the field as low residential density. They claim that the longer the dangling roads, the more sprawling the block.

Triangulated Irregular Network (TIN) was used in many scientific works as part of cartography; contour generation, generalisation, surface models and spatial analysis (Gökgöz, 2005; Yang et al. 2005; Kang et al. 2015). Semboloni (2000) proposed a growth model based on Delaunay triangles and road network. The model based on cellular automata operated within a lattice by using the essential elements of cells and roads which differ in size and form and the dynamic system functions by changing the state of the cells and generating new cells and roads.

This paper proposes a method for obtaining urban footprint based on the characteristics of road networks. The following section, firstly, presents study area and data. Then, the rule-based method is explained step by step. In section 3, the proposed method is implemented in four capital cities around the world (Washington, Stockholm, Madrid and Wellington), and the results are given. Section 4 evaluates the results of the study by comparing with authority data. Finally, the last section discusses the advantages and disadvantages of the method.

2. MATERIALS AND METHOD

2.1 Study Area and Data

This study was conducted using OpenStreetMap (OSM) road data of Washington DC, Madrid, Stockholm and Wellington (Fig. 1). The study area was determined as capital cities in different places of the world to reveal the efficiency of the proposed method in different metropole cities. They have the urban and suburban areas, rural lands, coasts, open spaces and forests.

Volunteers from all around the world contribute OSM road data. Some studies were conducted to determine the behaviour of the contributors and the quality of OSM data (Neis and Zipf, 2012; Koukoletsos et al. 2012; Corcoran et al. 2013; Zhao et al. 2015; Hacar et al. 2018). They show that the accuracy of the road data is sufficient for most of the GIS applications.

The authority of land use data is also used to evaluate the accuracy of the study. OSM road data and authority data were projected into the specific coordinate systems for each study area to apply the same measurement units (Table 1).

Some statistics in each study area requires to be compared and explained before the experiment. While

Fig. 1 can give the observable information that infers the most urbanised patterns per city area in both Washington DC and Stockholm, Table 2 gives the same information quantitatively. In Table 2, the road density (length per area) is the highest in Washington DC and Stockholm since length per area is maximum, so it is expected that the cities have less open places. The density is relatively middle or high in Madrid, so it may have more open places or forests. However, Wellington is expected to have many open place or forests since its density is very low. Besides, the measure of length per road may give some information about the road characteristics. For instance, if a city has a lower value, this means the city may have more residential area since the urban lands have shorted roads than rural lands. In this assumption, none of the cities has a major difference from others. Moreover, Table 2 also gives the source information of reference authority data. Each authority data (DCGIS, 2019; EEA, 2019; WCC, 2019) and the city boundaries (GADM, 2019) have open access facilities.

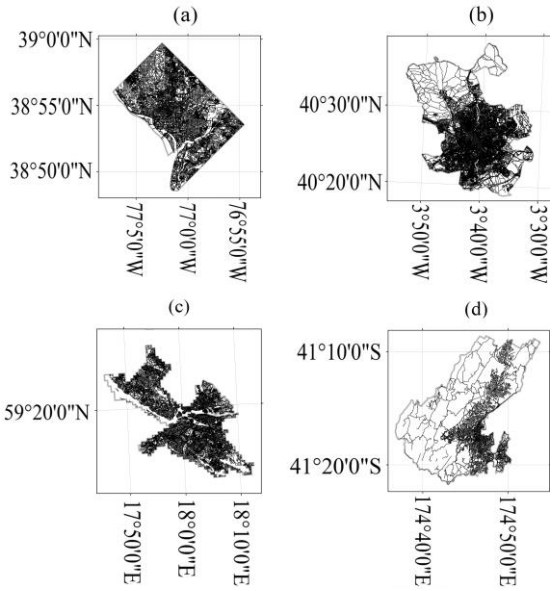


Figure 1. OSM road data in Washington DC (a), Madrid (b), Stockholm (c) and Wellington (d).

Table 1. Description of coordinate systems.

City	Coordinate System	Projection
Washington DC	NAD 1983 StatePlane Maryland FIPS 1900	Lambert Conformal Conic
Madrid	1870 Madrid Spain	Lambert Conformal Conic
Stockholm	SWEREF99 TM	Transverse Mercator
Wellington	NZGD 2000 Wellington Circuit	Transverse Mercator

2.2 The Proposed Method: from Road Network to Urban Footprint

Transportation is essential in any dynamic human land (Polat et al. 2017). Urban and suburban areas tend to have relatively short road lines (streets) since residential roads relatively have shorter length. The proposed method consists of several geometric indicators of road lines such as lengths of road and triangle edge and area of polygons. It starts with selecting road lines shorter than a threshold. The threshold aims to select the residential roads from all roads to be used in the rest of the process. Standard deviation (σ) of road lengths was pragmatically used as the threshold since it is represented by the same unit with lengths and emphasize the root square of the variations of lengths. The method uses the TIN model generated from centroids and dead-ends of selected roads to depict the urban land. The fusion of the adjacent triangles generates the urban polygons (Fig. 2).

Table 2. Statistics of the data.

City	OSM road data			City boundary		Source of authority data
	Number of roads	Total length (km)	Length per road (km)	Area (km^2)	Length (km) per area(km^2)	
Washington DC	28446	3707	0.13	166	22.33	DCGIS Open Data
Madrid	78076	10300	0.13	604	17.05	European Environment Agency
Stockholm	44247	4886	0.11	215	22.73	European Environment Agency
Wellington	10277	1693	0.16	290	5.84	Wellington City Council

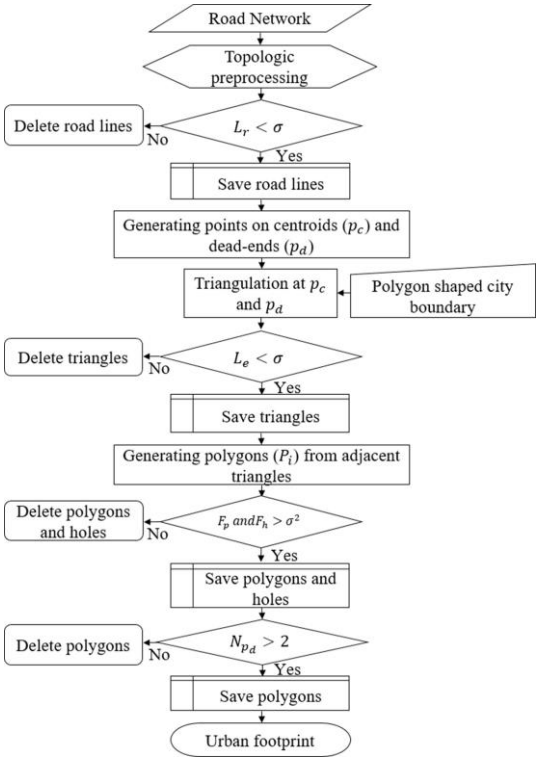


Figure 2. The workflow of RNUF.

The proposed method is referred hereafter as "RNUF" (initial letters of Road Network to Urban Footprint). A series of geometric rules with threshold values makes RNUF an easy-to-use method. Accordingly, the steps 1 - 8 were explained with the following definitions below:

Step 1:

RNUF starts with eliminating the topological inconsistencies of the input road network (pre-processing). After σ threshold calculation with the length of raw OSM roads, RNUF conducts a pre-process that eliminate the duplication of roads and generate topologically structured road networks by merging continuous roads and splitting them at intersections.

Step 2:

RNUF selects and retains the road lines less than σ value in length (Eq. (1)), others (longer) are eliminated.

$$L_r < \sigma \tag{1}$$

L_r represents the lengths of the road lines. In Fig. 3(a), red road lines represent the lines that are longer than σ.

Step 3:

Dangling points are used as parameters for determining urban or suburban areas (Owen and Wong, 2013; Liu and Jiang, 2011). In this study, dangling points are generated on the road lines that have dead-ends. Also, centroids are generated from the rest of the road lines using the formula below.

$$X_g = \sum_{i=1}^{w} \frac{S_i x_i}{\sum S} \qquad Y_g = \sum_{i=1}^{w} \frac{S_i y_i}{\sum S} \tag{2}$$

where X_g and Y_g are the centroid coordinates of a line, S_i is the length of segment *i* of the line, x_i and y_i are the midpoint coordinates of the segment *i*, *w* is the total number of segments, and *S* is the total length of the line.

Step 4:

Delaunay triangulation is generated using dangling points and centroids. In this stage, optionally, a city boundary file can be used to get a better result by constraining the TIN.

Step 5:

TIN model is updated by deleting the triangles that have at least one edge longer than σ (Fig. 3(b)). This criterion helps to eliminate big triangles represent non-urban areas among far neighbourhoods and the triangles connecting them to the far surrounding districts. Below Eq. (3), L_e represents the lengths of the triangle edges.

$$L_e < \sigma \tag{3}$$

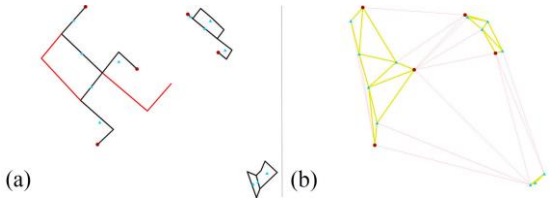


Figure 3. Dangling points (red) and centroids (blue), (a) valid road lines (black) and not valid road lines (red), (b) valid triangles (yellow) and not valid triangles (light red).

Step 6:

All adjacent triangles are merged, and their common edges are fused. Unique triangles and merged triangles represent the premature footprints (Fig. 4(a)). Many small polygons and holes are generated during this step.

Step 7:

The polygons and holes have a smaller area than the threshold σ^2 are deleted.

$$F_p < \sigma^2 \quad and \quad F_h < \sigma^2 \tag{4}$$

where F_p is the polygon area size and F_h is the hole size inside any polygon. In Fig. 4, the south-east polygon is eliminated due to the area size.

Step 8:

With a variety of observations on road networks and suburban lands, it was seen that even small neighbourhoods have at least three streets that have dead-ends. This criterion is the last rule of RNUF:

$$N_{p_d} \geq 3 \tag{5}$$

where N_{p_d} is the number of dangling points in a polygon. The polygons have at least three dangling points are retained to generate urban footprint (Fig. 4(b)), others are eliminated. The rule-based approach retains the polygons representing the urban and suburban footprint.

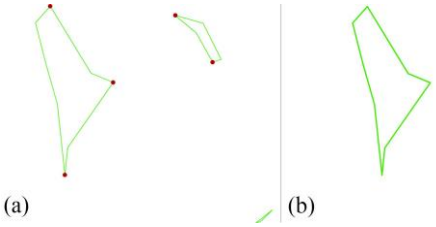


Figure 4. (a) Three polygons derived from triangles (green), (b) polygon representing urban footprint.

3. EXPERIMENT AND RESULTS

The proposed method has been tested with OSM data from the capital cities; Washington DC, Madrid, Stockholm, Wellington. Since the road networks in the cities have their characteristics in accordance with both geometric and topological properties, σ had different values (Table 3). Besides, road data had topological inconsistencies such as continuous lines at junctions since its raw structure consists of node and way geometries with their relations. RNUF automatically dealt with such situations by splitting lines at connections.

Table 3. σ and σ^2 values of study areas.

City	σ (m)	σ^2 (m^2)
Washington DC	196.5	38612.3
Madrid	242.6	58854,8
Stockholm	164.0	26896
Wellington	428.3	183440.9

As mentioned in *section 2.2*, OSM road data are topologically structured by RNUF. Therefore, the statistics for the study areas were computed by using topologically structured roads. The results for each study area are shown in Table 4. RNUF selected the roads more than 95% considering the σ values. These values were also used for the selection of significant triangles. If this criterion was not applied, the small settlements (Fig. 5) would have many triangles connecting them to the far surrounding settlements. As shown in Fig. 5, red zones represent the significant triangles that have all edges shorter than related σ. After all the steps were conducted, RNUF generated urban footprints of study areas, as shown in Fig. 6.

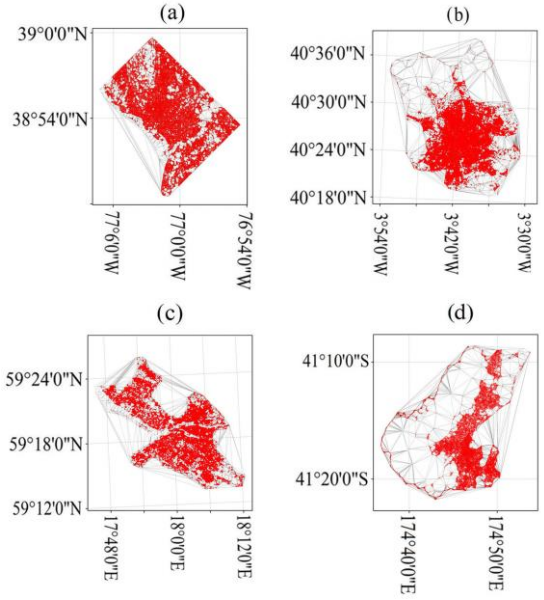


Figure 5. Triangle edges longer (grey) and shorter (red) than σ in the study areas: (a) Washington DC, (b) Madrid, (c) Stockholm, and (d) Wellington.

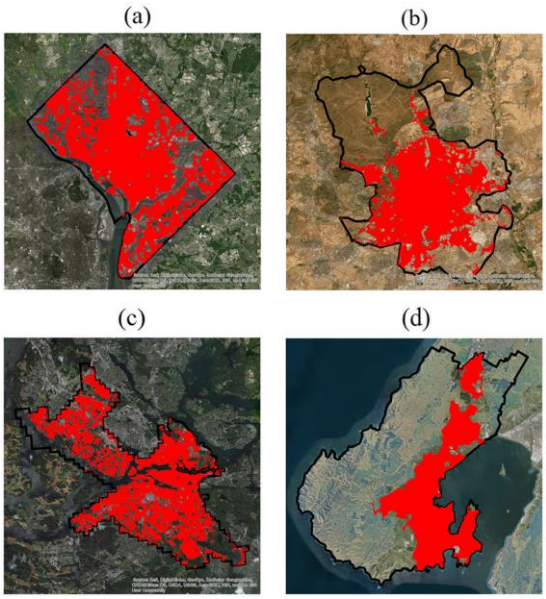


Figure 6. Urban footprints (red): (a) Washington DC, (b) Madrid, (c) Stockholm, and (d) Wellington (images from ESRI (2019)).

Table 4. Results of RNUF.

City	Number of roads	Number of selected roads	Selection percentages (%)	Number of dangling points	Area of urban footprint (km^2)	Dangling points per km^2	Dangling points per roads
Washington DC	60933	58427	96	5901	108.22	55	0.0009
Madrid	176197	170638	97	13996	249.65	56	0.0003
Stockholm	89915	85044	95	9236	126.60	73	0.0008
Wellington	18283	17786	97	4883	77.72	63	0.0034

4. EVALUATION OF THE RESULTS

The urban footprint map (*UF*) generated by RNUF were evaluated by using the authority data (*AD*) as the reference to determine the accuracy in each study area (Fig. 7). Determination of the accuracy by using the overlapped areas between *AD* and *UF* ($F_{overlap}$) is essential (Eq. (6)), but not sufficient if *UF* covers a very large area containing and exceeding whole reference data. For instance, the accuracy is the highest (94%) in Wellington, however, *UF* covered an area, approximately two times larger than the area of *AD* (Table 5 and 6). This means that half of the *UF* did not overlap with *AD*. This case shows that the completeness of *UF* is also an essential measure as its accuracy. To determine the completeness, Eq. (7) might be used. F-measure can be used to quantify the balance between accuracy and completeness (Samal et al. 2004; Song et al. 2011; Akbulut et al. 2018; Hacar and Gökgöz, 2019).

$$Accuracy = \frac{F_{overlap}}{F_{AD}} \tag{6}$$

$$Completeness = \frac{2 \times F_{overlap}}{F_{UF} + F_{AD}} \tag{7}$$

$$F - measure = \frac{2 \times Accuracy \times Completeness}{Accuracy + Completeness} \tag{8}$$

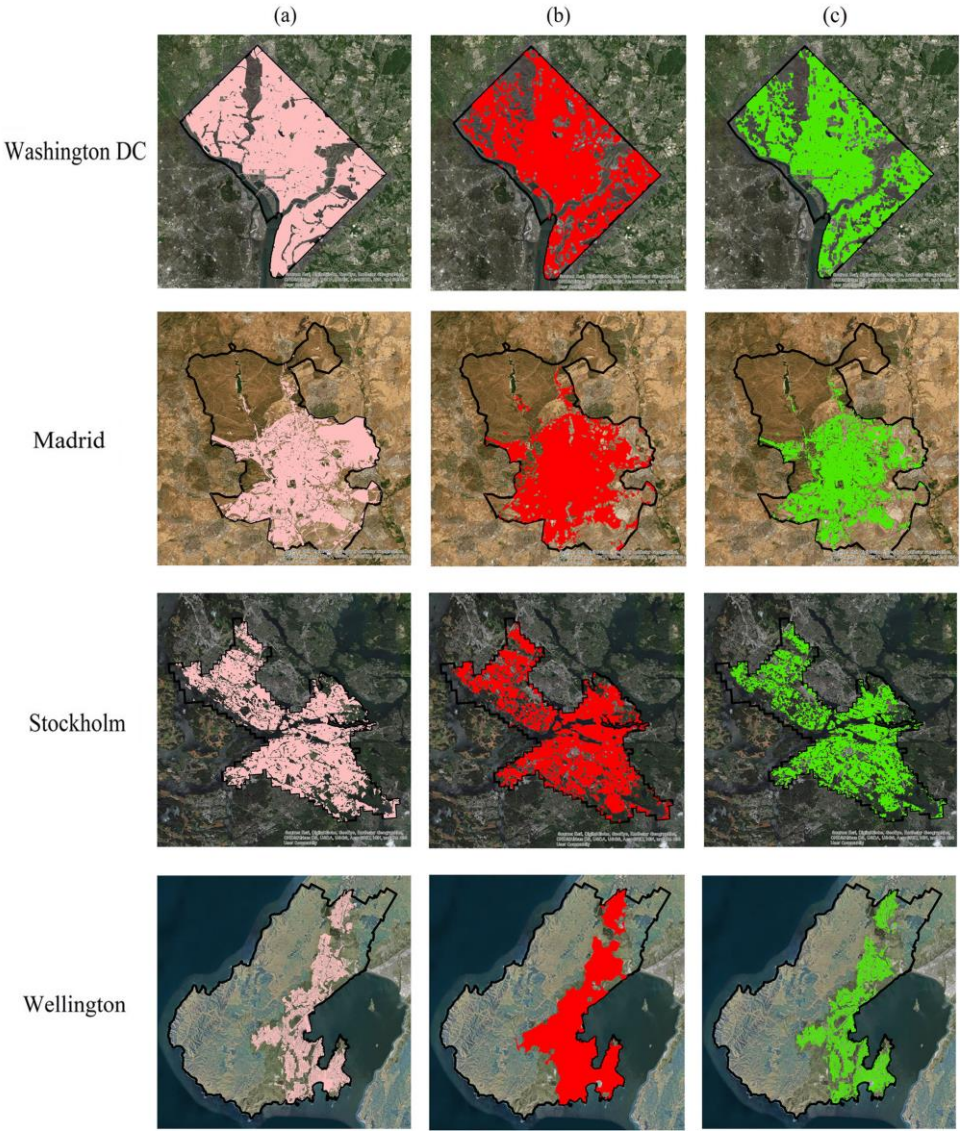


Figure 7. (a) Authority urban data (pink), (b) urban footprint (red), and (c) their overlapping area (green) (images from ESRI (2019)).

Table 5. Comparisons of RNUF with city boundary and authority data.

City	F_{CA}[1] (km^2)	F_{AD}[2] (km^2)	F_{UF}[3] (km^2)	$F_{overlap}$[4] (km^2)
Washington DC	166.05	124.98	108.22	97.69
Madrid	604.19	266.78	249.65	194.53
Stockholm	215.24	119.32	126.60	95.80
Wellington	289.95	45.33	77.72	42.48

[1]City area; [2]area of urban settlements in authority data; [3]area of urban footprint; [4]overlapping areas between *AD* and *UF*

Table 6. Statistical results of RNUF.

City	Accuracy (%)	Completeness (%)	F-measure (%)
Washington DC	78	84	80.9
Madrid	73	75	74.0
Stockholm	80	78	79.0
Wellington	94	69	79.6

F_{UF} values are close to F_{AD} in study areas except for Wellington (Table 5) since Wellington had the highest tolerances for both linear and polygonal thresholds, σ and σ^2 values, respectively. Besides, as seen in Fig. 8, in Madrid and Wellington, the F_{AD} and F_{UF} is very smaller than the areas of their city boundaries (F_{CA}). As a result, they have more rural land proportions than in other cities. In contrast, it can also be inferred that Washington and Stockholm have more urbanised land proportions. These two cities also had similar accuracy, completeness, and F-measure in Washington and Stockholm (Table 6) (Fig. 8). Furthermore, the accuracy and completeness had close values in Madrid, but in Wellington, they were very different. The completeness (69%) in Wellington is very low since half of the *UF* did not overlap with *AD* (Table 5).

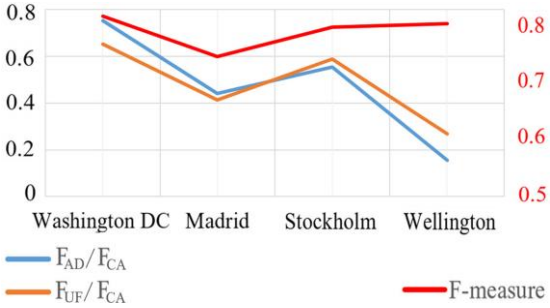


Figure 8. The rates of F_{AD} and F_{UF} in F_{CA} (black axis), and F-measures (red axis) in each city.

Dangling points are generally located in the cities of developing countries, sub-urban lands, and where height differences are high in metropole cities of developed countries. Since the study data is from the metropole cities, the density of the dangling points might show how the urban settlements are formed. Wellington had the most density of the dangling points (dangling points per roads) (Table 4). In other words, it has the most number of dead-ends in its urban lands. Therefore, it has more sub-urban lands and/or the height differences are very high. The dead-end statistics and the accuracy measures show that there is a significant relationship between the landforms/proportions and confidence of RNUF.

5. CONCLUSION

Producing urban footprint maps by RNUF is an easy and efficient way for small and middle scale maps represent such an area of a district, city or state. It is not recommended for large scale maps since its accuracy is dependent on accuracy, date and scale of input network data. If input data is not up-to-date, the footprint may not represent new small settlements. However, this situation may only affect the width of the related local neighbourhood in large-scaled studies, not the global result in middle and small-scaled studies. In this study, the experiment was depended on the completeness of OSM road data. One can conduct a completeness study for the source data to obtain current urban footprints since the higher completeness the data has, more up-to-date footprints can be generated. However, this paper does not have such aims that choose the more consistent data sources. This study remarks the possible usage of the road networks for extracting urban lands.

Conventional ways of generating urban footprint maps require many data (especially raster data) and indicators; instead, RNUF works independently of the aerial views of landscapes and surface models. It works with only the road network as the input data, and the used parameters are calculated from the network by easy-to-use GIS methods. RNUF has defined parameters and data-dependent thresholds like the minimum number of dangling nodes and standard deviation value, respectively. Users take advantages of these intrinsic values without trying to search complex indicators of urban sprawl like population, density, vegetation ratio or geomorphology. They may also integrate these indicators as additional rules with their thresholds into RNUF.

RNUF conducted the four instances satisfactorily with almost over 80% F-measure in each case. It can be said that RNUF can be used as an alternative for the methods that use remote sensing data. The experiments also show that landforms/proportions may decrease the confidence of RNUF.

Some future works are required to determine the thresholds in accordance with the exact scale of the resulting map and also the confidence of RNUF. An additional study may also be conducted for the last step of RNUF to measure how many dead-end streets a neighbourhood covers with regards to the urban patterns.

REFERENCES

Akbulut, Z. Özdemir, S. Acar, H. Dihkan, M. and Karslı, F. (2018). Automatic extraction of building boundaries from high resolution images with active contour segmentation. International Journal of Engineering and Geosciences , 3(1) , 37-42.

https://doi.org/10.26833/ijeg.373152

Angel, S., Parent, J. and Civco, D. (2007). Urban sprawl metrics: an analysis of global urban expansion using GIS. Proceedings of ASPRS 2007 Annual Conference, Tampa, Florida May 7–11. URL: http://clear.uconn.edu/publications/research/tech_papers/Angel_et_al_ASPRS2007.pdf

Bhatta, B. Saraswati, S. and Bandyopadhyay, D. (2010). Urban sprawl measurement from remote sensing data. Appl Geogr. 30(4), 731-740. https://doi.org/10.1016/j.apgeog.2010.02.002

Canaz Sevgen, S . (2019). Airborne lidar data classification in complex urban area using random forest: a case study of Bergama, Turkey. International Journal of Engineering and Geosciences , 4 (1) , 45-51. https://doi.org/10.26833/ijeg.440828

Corcoran, P. Mooney, P. and Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. Spat Stat. 3, 21-32. https://doi.org/10.1016/j.spasta.2013.01.002

DCGIS. (2019). Existing Land Use. DC Office of Planning - Long Range Planning. [accessed on 20 September 2019]. http://opendata.dc.gov/datasets/245179183eee41e08852ff9d5dbd3bcb_4

EEA. (2019). Urban atlas. EU Open Data Portal. [accessed on 20 September 2019]. URL: https://data.europa.eu/euodp/data/dataset/data_urban-atlas

Esch, T. Marconcini, M. Felbier, A. Roth, A. Heldens, W. Huber, M. Schwinger, M. Taubenböck, H. Müller, A. and Dech, S. (2013). Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. IEEE Geosci Remote S. 10(6), 1617-1621. https://doi.org/10.1109/LGRS.2013.2272953

ESRI. (2019). Imagery Basemap. Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

GADM. (2019). GADM data. Database of Global Administrative Areas. [accessed on 20 September 2019]. URL: https://gadm.org/data.html

Gökgöz, T. (2005). Generalisation of contours using deviation angles and error bands. Cartogr J. 42(2), 145-156. https://doi.org/10.1179/000870405X61441

Hacar, M. and Gökgöz, T. (2019). A new, score-based multi-stage matching approach for road network conflation in different road patterns. ISPRS Int J Geo-Inf, 8(2), 81. https://doi.org/10.3390/ijgi8020081

Hacar, M. Kılıç, B. and Şahbaz, K. (2018). Analysing OpenStreetMap road data and characterising the behavior of contributors in Ankara, Turkey. ISPRS Int J Geo-Inf. 7(10), 400. https://doi.org/10.3390/ijgi7100400

Jiang, F. Liu, S. Yuan, H. and Zhang, Q. (2007). Measuring urban sprawl in Beijing with geo-spatial indices. J Geogr Sci. 17(4), 469-478. https://doi.org/10.1007/s11442-007-0469-z

Kang, M. Wang, M. and Du, Q. 2015. A method of DTM construction based on quadrangular irregular networks and related error analysis. PloS one, 10(5), e0127592. https://doi.org/10.1371/journal.pone.0127592

Karakuş, P. Karabork, H. and Kaya, S. (2017). A comparison of the classification accuracies in determining the land cover of Kadirli region of Turkey by using the pixel based and object based classification algorithms. International Journal of Engineering and Geosciences , 2(2) , 52-60. https://doi.org/10.26833/ijeg.298951

Koukoletsos, T. Haklay, M. and Ellul, C. (2012). Assessing data completeness of VGI through an automated matching procedure for linear data. Trans GIS. 16(4), 477-498. https://doi.org/10.1111/j.1467-9671.2012.01304.x

Kumar, J.A.V. Pathan, S.K. and Bhanderi, R.J. (2007). Spatio-temporal analysis for monitoring urban growth–a case study of Indore city. J Indian Soc Remot. 35(1), 11-20. https://doi.org/10.1007/BF02991829

Liu, X. and Jiang, B. (2011). A novel approach to the identification of urban sprawl patches based on the scaling of geographic space. Int J Geomat Geosci. 2(2), 415-429. URL: http://www.diva-portal.org/smash/get/diva2:502047/FULLTEXT01.pdf

Musa, S.I. Hashim, M. Reba, M.N.M. (2017). A review of geospatial-based urban growth models and modelling initiatives. Geocarto Int. 32(8), 813-833. https://doi.org/10.1080/10106049.2016.1213891

Neis, P. and Zipf, A. (2012). Analysing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. ISPRS Int J Geo-Inf. 1(2), 146-165. https://doi.org/10.3390/ijgi1020146

Owen, K.K. and Wong, D.W. (2013). An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics. Appl Geogr. 38, 107-118. https://doi.org/10.1016/j.apgeog.2012.11.016

Polat, Z.A. Memduhoğlu, A. Hacar, M. and Duman, H. (2017). [in Turkish: Kentsel Büyüme İle Motorlu Araç Trafiği Yoğunluğu Arasindaki İlişkinin Belirlenmesi: İstanbul Örneği]. Omer Halisdemir University J Eng Sci, 6(2), 442-451. https://doi.org/10.28948/ngumuh.341275

Semboloni, F. (2000). The growth of an urban cluster into a dynamic self-modifying spatial pattern. Environ Plann B. 27(4), 549-564. https://doi.org/10.1068%2Fb2673

Triantakonstantis, D. and Stathakis, D. (2015). Examining urban sprawl in Europe using spatial metrics. Geocarto Int. 30(10), 1092-1112. http://dx.doi.org/10.1080/10106049.2015.1027289

Tsai, Y.H. (2005). Quantifying urban form: compactness versus 'sprawl'. Urban Stud. 42(1), 141-161. https://doi.org/10.1080%2F0042098042000309748

Turkish Statistical Institute (TUIK). (2019). Address based population registration system. [accessed on 20 September 2019]. URL: http://www.turkstat.gov.tr/Start.do

WCC. (2019). WCC District Plan Zones. Wellington City Council. [accessed on 20 September 2019]. URL: https://data-wcc.opendata.arcgis.com/datasets/6c3aaccfdbbf470491fb688595cf5b7e_0

Yang, B. Li, Q. and Shi, W. (2005). Constructing multi-resolution triangulated irregular network model for visualisation. Comput Geosci. 31(1), 77-86. https://doi.org/10.1016/j.cageo.2004.09.011

Zhao, P. Jia, T. Qin, K. Shan, J. and Jiao, C. (2015). Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. Physica A. 420, 59-72. https://doi.org/10.1016/j.physa.2014.10.076

International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 109-119, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.629381

ACCURACY AND SIMILARITY ASPECTS IN ONLINE GEOCODING SERVICES: A COMPARATIVE EVALUATION FOR GOOGLE AND BING MAPS

Batuhan Kilic[1], Fatih Gulgen[1]*

[1]Yildiz Technical University, Civil Engineering Faculty, Department of Geomatic Engineering, Istanbul, Turkey
(fgulgen/batuhank@yildiz.edu.tr); **ORCID 0000-0002-8754-9017, ORCID 0000-0002-0529-8569**

*Corresponding Author, Received: 04/10/2019, Accepted: 12/11/2019

ABSTRACT: Geocoding is a method used to convert address information into geographical coordinates. It plays a vital role in displaying the relationship between geographic features and semantic information expressed in texts. The objective of this study is to reveal the quality of online geocoding from postal addresses in Turkey provided by Google Maps and Bing Maps services. The quality of geocoding services in urban areas is evaluated using two particular metrics; positional accuracy and address similarity. Positional accuracy measures the distances between point features obtained through the online geocoding and reference data. Address similarity indicates the relationship between two postal addresses based on a similarity index known as the Levenshtein distance. The same performance assessment was also made with the United States' address data to make comparisons and discussions. The results show that services have different geocoding capabilities in both countries because of the differences in the addressing formats.

Keywords: *Geocoding, Google Maps, Bing Maps, Positional accuracy, Levenshtein distance*

1. INTRODUCTION

One of the essential processes used in the geographical analysis is geocoding, which assigns a coordinate pair to the description of a place by comparing the descriptive location-specific elements to those in reference data (Zandbergen, 2008). Researchers often use geocoding services to conduct the geographical analysis in various areas such as epidemiology (McElroy et al. 2003; Ward et al. 2005; Rushton et al. 2006), public safety (Ratcliffe 2004; Bichler and Balchak 2007; Hart and Zandbergen 2013), and traffic accident detection (Qin et al. 2013). There are two types of geocoding tools known as offline or online. Software packages in Geographic Information System (GIS) include offline geocoding tools. Online geocoding tools depend on map services. The popular services which provide online geocoding process use street or rooftop geocoding techniques. Street geocoding utilizes street network data, while the rooftop technique employs address (point) data. Considering that geocoding performance relies heavily on the quality of the reference database, the rooftop geocoding would produce higher accurate results than street geocoding (Roongpiboonsopit and Karimi 2010b).

Several cartographers and GIS specialists studying on geocoding were focused on the creation of model proposals and the assessment of positional quality between the different service results. The model proposal studies usually aimed to solve complex address structures. Li et al. (2010) proposed an address model using a method in China called association rule mining based on machine learning to eliminate the adverse effect of the irregular address structure. Tian et al. (2016) proposed a geocoding service based on an optimized address mapping method by designing a particular address model for China and revealing the differences between the address structures of China and Western countries. The geocoding quality assessment studies also aim to compare the results of different services in different regions. Cayo and Talbot (2003) performed geocoding with MapMarker software using residential addresses. They evaluated the positional accuracy of the obtained points by comparing them with orthophoto maps. Yang et al. (2004) identified the error sources that resulted in geocoding by revealing the pros and cons of commercial (i.e., ArcView, Automatch, and ZIP/GeoLytics) geocoding systems. Karimi et al. (2004) investigated the matching rate and positional accuracy uncertainties arising from geocoding techniques by comparing them with the data collected by the Global Navigation Satellite System (GNSS). Roongpiboonsopit and Karimi (2010a) intended to provide an overview of the appropriate services by examining the matching quality of the online geocoding services: Geocoder.us, Google Maps, MapPoint, MapQuest, and Yahoo. The quality of each service was assessed based on three different measures, defined as matching rate, positional accuracy, and similarity between the services as a result of matching. Zandbergen (2011) compared the geocoding results obtained from three different datasets to determine the influence of road data on geocoding studies. Goldberg et al. (2013) developed an evaluation framework for comparing geocoding systems without mentioning the name because of the non-disclosure agreements. Chow et al. (2016) compared the results of geocoding services (i.e. desktop geocoding software: ESRI and CoreLogic PxPoint and online geocoding software: Google Maps, Yahoo, Bing Maps, Geocoder.us, Texas A&M University Geocoder and OpenStreetMap) with GNSS metrics by gathering residential addresses in the state of Texas, United States of America (USA). Cetl et al. (2016) compared two online geocoding services (Google Maps and OSM Nominatim) results, taking into account population density in the city of Zagreb, Croatia.

Turkey is one of the countries which spread over a wide area and has a high population. The official institutions have tried to reorganize the unusual address expressions in a standard form for many years. However, abundant address components, improper and incorrect implementation of numerating studies, the failure of institutions to use the same address format even in their documents, and institutions' unawareness of new address components bring about confusion related to the addresses in Turkey (Yildirim et al., 2014; Kilic and Gulgen, 2019). Since the irregular expressions obscure understanding the components of an address specified by a user, the geocoding services have trouble matching the input address with the standard reference dataset. Kilic and Gulgen (2017) have compared the postal addresses used in Turkey and the USA within the scope of the geocoding process. They revealed that the geocoding accuracy in Turkey is less because of the following reasons: external door number, incomplete address, misspellings, typographical errors, and incorrect format.

A reference address is composed of a group of address components. The general components of a formatted address used in many countries are address number, street name, town, state/abbreviation, and postal code. Parsing them from an address description is a complicated process because users cannot usually define addresses in a standard form. The difficulty in converting the user's entries to the standard address form affects the quality of the geocoding process. The geocoding quality has been frequently evaluated in countries that have a high standard addressing system. However, the quality of geocoding issues in Turkey has not been adequately investigated yet. This study assesses the quality of the geocoding process used in Google Maps and Bing Maps services for Turkey and the USA. Both services, which are widely used throughout the world, are freely available to everyone. The purpose of this paper is to reveal the quality of geocoding by (1) measuring the positional accuracy between the retrieved and reference data and (2) comparing the text of retrieved with the reference using Levenshtein distance, which is a text similarity technique. While the first type of evaluation is not new, the later evaluation has never before been reported in the literature.

2. STUDY AREA AND REFERENCE DATA

Evaluating the performance of the geocoding process with dependable and accurate data sets for a large geographic area is quite tricky (Roongpiboonsopit and Karimi 2010b). Therefore, two distinct experimental tests in the urban centers of the USA and Turkey with a limited number of reference data that are reliable and

have similar characteristics were implemented.

A point of interest (POI) is a specific point location that someone may find useful or interesting. People use POIs for various purposes, such as navigating between different locations, determining the characteristics of a place, studying urban sociology, analyzing city dynamics, and geo-referencing of the texts (Rodrigues 2010). In this study, the accommodation facilities which supply communication between the dynamics of the cities (Bilgi et al. 2019) were used as the reference data. They are usually stored by point symbols in a POI layer of various online maps (Mulazimoglu and Basaraner, 2019). Accordingly, two test regions were selected: the district of Fatih in Turkey and the city of Miami Beach in the USA. Fatih is located in the historical peninsula of Istanbul, surrounded by the Byzantine city walls, the Golden Horn, and the Sea of Marmara (Figure 1). This region, which has harbored many civilizations throughout its thousands of year's history, is currently the most significant historical tourism center of Istanbul. The other test region, Miami Beach, is a famous coastal resort city in Florida (Figure 2). The city is a major international entertainment and cultural destination. It is a widely visited tourist area and offers various accommodation facilities.

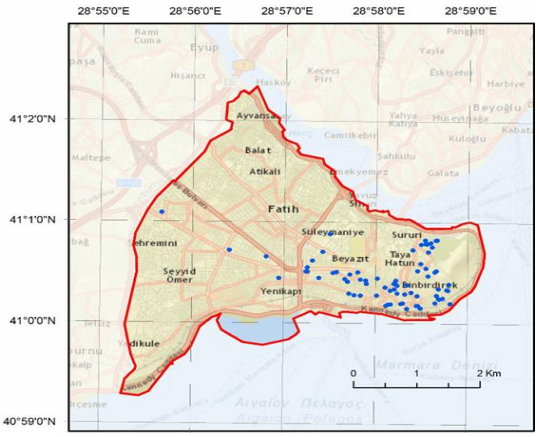


Figure 1. Approximate boundary of the study area within Fatih and accommodation facilities (blue points).

The postal addresses of the accommodation facilities forming the basis of this study were obtained from *www.booking.com* website automatically by using the web scraping method. Web scraping is a data mining process used to collect standard information stored in a specific location of a web page (Mitchell 2015). There are several commercial web scraping software such as Visual Web Ripper, Web Scraper Plus+, and many more (Rodrigues, 2010). In this study, *Web Content Extractor v.8.4* software developed by Newprosoft Company was used to collect the accommodation data. As a result of the scraping process, name and address information of 250 and 271 hotels with various types of building were obtained from Fatih and Miami Beach, respectively. Three scraped postal addresses are given in Table 1 to demonstrate the general address characteristics of the test regions. The first sample address is from the USA, while the other two are from Turkey.

The majority of scraped addresses contained repetitive and inconsistent information. Therefore, the problematic ones were omitted from the list after reviewing the entire data. The remaining addresses were considered identifiers of POIs, and their approximate locations were geocoded from the facility name by using Google Maps and Bing Maps services. Then, the advertising signs of the facilities over the street perspectives were checked to confirm the accuracy of the geocoding process based on the facility names (Figure 3a and Figure 4a). Thus, the building footprints of 74 hotels in Fatih and 82 hotels in Miami Beach were verified as reference data for test processes (Figure 3b and Figure 4b).

Figure 2. Approximate boundary of the study area within Miami Beach and accommodation facilities (blue points).

Table 1. Examples of postal addresses scraped from www.booking.com website automatically.

Name	Postal Address
Aqua Hotel&Suits	1530 Collins Avenue, Miami Beach, FL 33139, United States of America
Angel's Home Hotel	Sultan Ahmet Mh. Amiral Tafdil Sk. No:26, 34122 Fatih/Istanbul, Turkey
Hotel Romantic	Cankurtaran Mh. Amiral Tafdil Sk. No:17, 34122 Fatih/Istanbul, Turkey

Figure 3. a) Street view and b) Map view of a reference hotel in Fatih.

Figure 4. a) Street view and b) Map view of a reference hotel in Miami Beach.

The address point of a building can be extracted from a combination of various data sources including, but not limited to, parcel maps, building footprints, orthophotos, satellite imagery, and field survey (Roongpiboonsopit and Karimi 2010b). It represents the location of an addressable structure in a jurisdiction and is usually placed directly on top of the specific footprint or directly in front of it (Hart and Zandbergen 2013). These are valid and acceptable representations. In this study, the POIs were placed at the centroid of each digitized polygons. Then, their locations were stored in individual datasets for the test regions. (Figure 3b and Figure 4b). The centroid of a building polygon is the mean position of all vertices in two coordinate directions. If it falls outside the boundaries, the point can be displaced to what is considered a center of gravity within the boundaries.

3. ONLINE GEOCODING FROM POSTAL ADDRESS

Online geocoding is a network-accessible component which is available on the Internet with a Web Application Programming Interface (API) as a service. The geocoding API converts an entry into coordinates and then delivers the result, which includes the coordinates, the address used in geocoding, and the level of accuracy back to the user over the internet (Roongpiboonsopit and Karimi 2010a). During the online geocoding process, a service parses an address data into its natural elements and compares them to the potential candidates within the reference database of service. Then, it identifies and delivers the normalized address and geographical coordinates of the best-matched address as the output. The purpose of parsing is to segment an input address string into meaningful address elements with exact spatial semantics; in address normalization, any informal or abbreviated address elements are converted into a standard format, and address elements that are miswritten are re-recorded in the correct form (Tian et al. 2016).

In this study, the reference addresses were geocoded using a library for the Google Maps Geocoding API and the Bing Maps Locations API in the Python programming language. As with other online geocoding services, Google Maps and Bing Maps services have different limits and constraints concerning licensing of service usage (URL 1 and URL 2). Both services produce their results using a rooftop geocoding technique for all postal addresses in Miami Beach, while 51 of 74 postal addresses in Fatih are coming from the roof. Of the remaining 23 postal addresses, Bing Maps and Google Maps generated 14 and 8 using the rooftop technique, respectively. There was only one address that is geocoded by the street technique in both services.

The main address components of Google Maps and Bing Maps with the parsed elements belonging to three sample hotels presented in Table 1 were given in Table 2. These services had many parts such as house number, street name, postal code, neighborhood, county, city, state, and country, though some of them were defined in a combined form (e.g., Bing Maps converts the house number and street name into one component).

Table 3 contains the normalized address data and the related geographic coordinates in the standard form after the parsing process. Both services for the first input data returned pretty similar normalized address information, although Bing Maps could not give the parsed elements for neighborhood and county. For the second and third input data, neighborhood, county, and city features were not parsed correctly by the geocoding services because of the complexity or deficiency in sub-region information regarding the addressing system in Fatih. Google Maps converted the postal code element in the third parsed address because of the undefined neighborhood element. Google Maps' normalized format contained a neighborhood element that was incompatible with the postal code information. Bing Maps did not present any neighborhood information which was parsed correctly from input data (Table 3).

Table 2. Postal addresses parsed by services in a standard format.

Services	House Number	Street Name	Postal Code	Neighborhood	County	City	State	Country
Google Maps	1530	Collins Avenue	33139	South Beach	Miami-Dade County	Miami Beach	FL	US
	26	Amiral Tafdil Sk.	34122	-	Fatih	-	Istanbul	TR
	17	Amiral Tafdil Sk.	34122	-	Fatih	-	Istanbul	TR

Table 2. (Continued).

Bing Maps	1530 Collins Ave	33139	-	-	Miami Beach	FL	United States
	Amiral Taftil Sokak 26	34122	Sultan Ahmet	-	Istanbul	Istanbul	Turkey
	Amiral Taftil Sokak 17	34122	Cankurtaran	-	Istanbul	Istanbul	Turkey

Table 3. Postal addresses normalized by services in a standard format.

Services	Name	Address	Latitude and Longitude
Google Maps	Aqua Hotel&Suits	1530 Collins Ave, Miami Beach, FL 33139, USA	[25.78825, -80.13083]
	Angel's Home Hotel	Sultan Ahmet Mh., Amiral Tafdil Sk. No:26, 34122 Fatih/Istanbul, Turkey	[41.00415, 28.97945]
	Hotel Romantic	Cankurtaran Mh., Amiral Tafdil Sk. No:17, 34122 Fatih/Istanbul, Turkey	[41.00424, 28.97989]
Bing Maps	Aqua Hotel&Suits	1530 Collins Ave, Miami Beach, FL 33139	[25.78836, -80.13086]
	Angel's Home Hotel	Amiral Taftil Sokak 26, 34122 Istanbul, Turkey	[41.00421, 28.97962]
	Hotel Romantic	Amiral Taftil Sokak 17, 34122 Istanbul, Turkey	[41.00427, 28.97989]

In Turkish addressing format, the sub-region generally described by a neighborhood name is a crucial element. In some cases, buildings such as accommodation facilities handled in this study may be located within a different sub-region despite being on the same street. This paradox is the most fundamental issue that has not been solved by global geocoding services such as Google Maps and Bing Maps in Turkey. Such examples can be seen in the second and third addresses in Table 3 for both services in Fatih. On the other hand, both services use their particular address formats in Fatih. Bing Maps gives a house number to the structure after the street name and type, yet Google Maps always returns a normalized postal address together with an abbreviation "No:" for the same place used in Bing Maps.

4. QUALITY OF GEOCODING

All geocoding results consisting of normalized addresses and geographic coordinates in both test regions obtained by Google Maps and Bing Maps services are examined below in two sections: positional accuracy and similarity of address description.

4.1 Positional Accuracy

The positional accuracy of the online geocoding is evaluated compared to the reference data and dealt with in two stages:

- Analysis of the distance and the coordinate deviation between the geocoded point of each address and its actual location,
- Analysis of the topological relation between the geocoded point of each address and its building footprint.

In the first stage, the positional accuracy between the geocoded point and the actual location by measuring the straight-line distance in the Euclidean space was calculated. Several authors assumed that the precise location of an address is the centroid of the building footprint on a parcel (Zandbergen and Green 2007; Armstrong and Tiwari 2008; Zandbergen 2008; Roongpiboonsopit and Karimi 2010b; Chow et al. 2016). In this study, latitudes and longitudes were transformed into projected coordinates to calculate the positional deviation between the geocoded and centroid points. For Fatih and Miami Beach, the Gauss Kruger in ITRF96 datum and the Florida State Plane in NAD83 datum were the coordinate systems, respectively. The deviation for easting and northing coordinates was calculated by using Eq. (1) and Eq. (2), respectively.

$$\Delta E_i = E_{ref} - E_i \quad (1)$$

$$\Delta N_i = N_{ref} - N_i \quad (2)$$

where ΔE_i and ΔN_i refer to deviations in the easting and northing coordinates, E_{ref} and N_{ref} are the easting and northing coordinates of the centroid, E_i and N_i are the easting and northing coordinates of the geocoded point, respectively. The positional deviation ΔS_i between the two points is calculated from the values ΔE_i and ΔN_i by using Eq. (3).

$$\Delta S_i = \sqrt{\Delta E_i^{\,2} + \Delta N_i^{\,2}} \quad (3)$$

The calculated deviations in Fatih and Miami Beach are shown on the graphs in Figure 5 and Figure 6, respectively. The horizontal coordinate in all charts represents the rank, which increases depending on the positional deviation value calculated for each accommodation facility. The vertical coordinates of ΔE and ΔN graphs show positive or negative deviations between the geocoded point and the centroid in the easting and northing directions, respectively. The vertical coordinate of ΔS refers to the values of positional deviation. According to the deviation graphs, Bing Maps' geocoding results have a more significant difference than Google Maps' results in Fatih. The results of both services in Miami Beach are reasonably close to each other (Figure 5 and Figure 6).

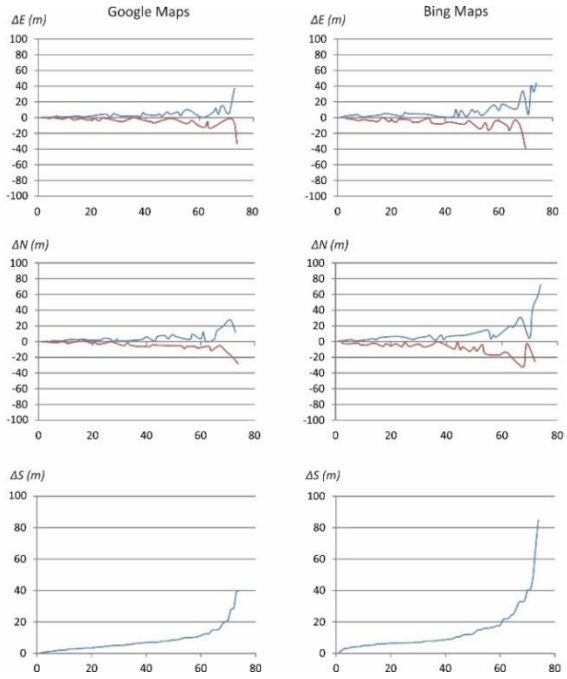


Figure 5. Easting, northing, and positional deviations for Fatih: blue and red lines show the positive and negative deviations, respectively.

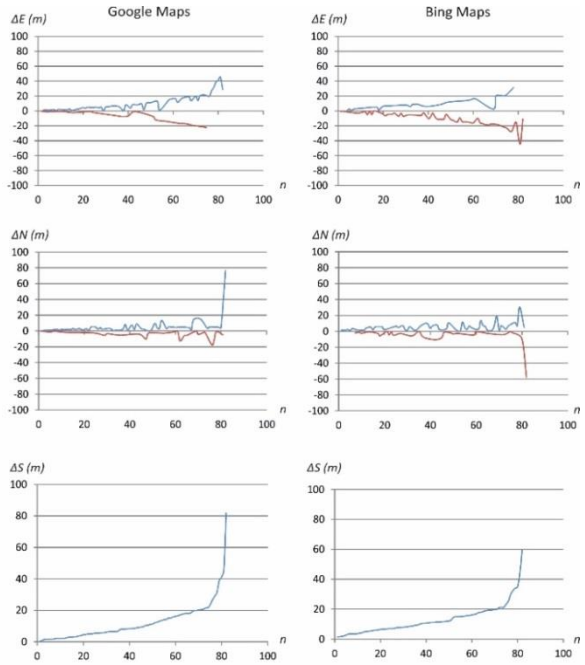


Figure 6. Easting, northing, and positional deviations for Miami Beach: blue and red lines show the positive and negative deviations, respectively.

Statistically, several measures known as minimum, maximum, median, mean, and norm are used to reveal the quantity of error between the geocoded point and the exact location. The minimum and maximum are the least and the highest samples of a set of observations. The median and mean are the conventional measures of central tendency in summary statistics. Equation (4) shows the norm value, which is the total magnitude of observation values. It increases depending on the number of observations in a set.

$$Norm = \sqrt{\sum_i^n {x_i}^2} \quad (4)$$

where x refers to each observation value, and n represents the number of observations.

The minimum, maximum, median, mean, and norm values of the deviations are given in Table 4. The absolute values are used when calculating the minimum, median, and mean. According to Table 4, all Google Maps' statistical values in Fatih are lower than Bing Maps' values. These values prove that Google Maps' geocoding results have higher accuracy than Bing Maps' results in Fatih. The statistical significances of the deviations in Miami Beach are quite varied, and the benefits of neither Google Maps nor Bing Maps are superior over the other. They represent that the positional accuracies of both services are approximately similar to each other in Miami Beach.

Table 4. Statistical measures calculated from deviation values.

Region	Measure	**ΔS** (m)		**ΔE** (m)		**ΔN** (m)	
		Google Maps	Bing Maps	Google Maps	Bing Maps	Google Maps	Bing Maps
Fatih	Min	0.1	0.4	0.1	0.3	0.1	0.3
	Max	39.7	85.1	37.3	44.3	28.6	72.6
	Median	6.4	8.2	3.1	5.2	4.4	5.8
	Mean	8.3	13.9	4.9	8.1	5.7	9.8
	Norm	97.5	171.2	67.8	105.4	70.1	134.9
Miami Beach	Min	0.3	1.4	0.1	0.1	0.2	0.1
	Max	82.0	59.7	45.3	43.7	76.9	58.7
	Median	8.4	10.9	6.8	8.2	2.8	3.5
	Mean	12.2	12.9	10.0	10.4	4.8	5.4
	Norm	155.3	146.5	126.2	120.6	90.5	83.3

On the other hand, when Google Maps' statistical values were examined, it is found that the amount of deviations within Fatih is lower than the differences within Miami Beach. The reason for better results of geocoding in Fatih is the fact that most of the building geometry in Miami Beach consists of non-rectangular and complex shapes, and also an enormous amount of buildings in Fatih has simple geometry. Besides, the statistical values are insufficient to make decisions in the test region, where Bing Maps results better than Google Maps.

In the second stage, the positional accuracy between the geocoded point and building footprint was measured by considering the point&polygon topological relationships. The point-in-polygon algorithm, which is an essential operation to determine whether a point is inside a complex polygon (Haines, 1984), was employed. Likewise, Roongpiboonsopit and Karimi (2010b) used this algorithm to verify the correctness of geocoded locations at the building level.

When geocoding an address using the rooftop technique, one can expect that the point will place within the building footprint. The point shown in blue in Figure 7 is adequately geocoded. However, this expectation is not always possible. The red spot is improper because the geocoding service locates it outside the building footprint. The numbers and percentages of point-in-polygon analysis in each test region used in this study are presented in Table 5.

Figure 7. Geocoding points located inside and outside the digitized buildings.

The results of the point-in-polygon analysis show

that the number of geocoded points obtained adequately from the Google Maps service is more than that obtained from the Bing Maps service in both regions. Similarly, the results of Miami Beach are better than those of Fatih. The differences in percentages between Miami Beach and Fatih are approximately 5% on Google Maps and 15% on Bing Maps.

Table 5. Point-in-polygon analysis for geocoding services.

Region		Google Maps		Bing Maps	
		Inside	Outside	Inside	Outside
Fatih	Number	61	13	44	30
	%	82.4	17.6	59.5	40.5
Miami Beach	Number	72	10	61	21
	%	87.8	12.2	74.4	25.6

On the other hand, the best geocoding location for an address is a place close to the front edge of the building footprint. The front side is determined by considering the actual status of the building's main entrance door. Considering that the proper geocoding point is on a building's footprint boundary, if a geocoded point is within the footprint, the address can be assumed to be geocoded correctly. It means that the location of the geocoded point can be varied from the centroid to the footprint of a building. The quantity of error for a geocoding point, which is outside the footprint, is calculated from the closest distance between the location and the related footprint edges (Figure 7). Table 6 provides some statistical values for the improper geocoding points in each test region. These show that geocoding results obtained from Google Maps are of higher quality than Bing Maps, and the results in Miami Beach are more reliable than the results in Fatih.

Table 6. Statistical measures calculated for improper (outside) geocoded points.

	Distances (m)			
	Fatih		Miami Beach	
	Google Maps	Bing Maps	Google Maps	Bing Maps
Minimum	0.6	0.1	0.3	0.7
Maximum	28.2	60.5	6.3	8.5
Median	5.0	6.6	1.8	2.5
Mean	9.3	11.1	2.1	3.0
Norm	48.1	92.8	8.6	17.3

4.2 Address Similarity

The similarity rate between two strings can be calculated by using a similarity measurement method known as string metric in information theory, mathematics, and computer science. The most widely known string metric is the Levenshtein distance. The minimum number of operations gives the Levenshtein distance between two string vectors, and that needed to transform one text into the other, where an action is an insertion, deletion, or substitution of a single character (Levenshtein 1966). In this study, The Levenshtein distance as an indicator of similarity between two postal addresses was used. The algorithm detects the address similarity in two stages. In the first, the similarity percentage between the postal address of an accommodation facility obtained from the scraped website and the best-matched address in each service database is calculated. Then, a calculation of the similarity percentage is done between the results of both services. Figure 8 shows the similarity graphs of computed values for both test regions. According to the charts containing scraped data, the similarity in Fatih for Google Maps changes approximately between 50-85% and between 35-82% for Bing Maps. The similarity in Miami Beach for Google Maps and Bing Maps is roughly between 67-80% and 60-75%, respectively. The increased difference between the highest and lowest percentage values (i.e., scale expanding on the vertical axis of the graphs) points out the issues that are growing related to the use of standard address format. Differences between the maximum and minimum values in the charts reveal that the address standard in Miami Beach is higher than that in Fatih. The similarity between the addresses obtained from both services is approximately 37-73% in Fatih and 66-95% in Miami Beach. Also, the average value calculated for both services in Miami Beach is hugely higher than the others (Table 7). These percentages indicate that Google Maps and Bing Maps return the addresses in a similar standard format for Miami Beach. However, the address format used by each service has different standards in Fatih due to the complications in the identification of the address components in Turkey.

Table 7. Similarity percentages for Google Maps, Bing Maps geocoded addresses, and reference addresses.

	Reference addresses		Google Maps
	Google Maps	Bing Maps	Bing Maps
Fatih	65.3	64.6	61.2
Miami Beach	72.3	66.9	93.2

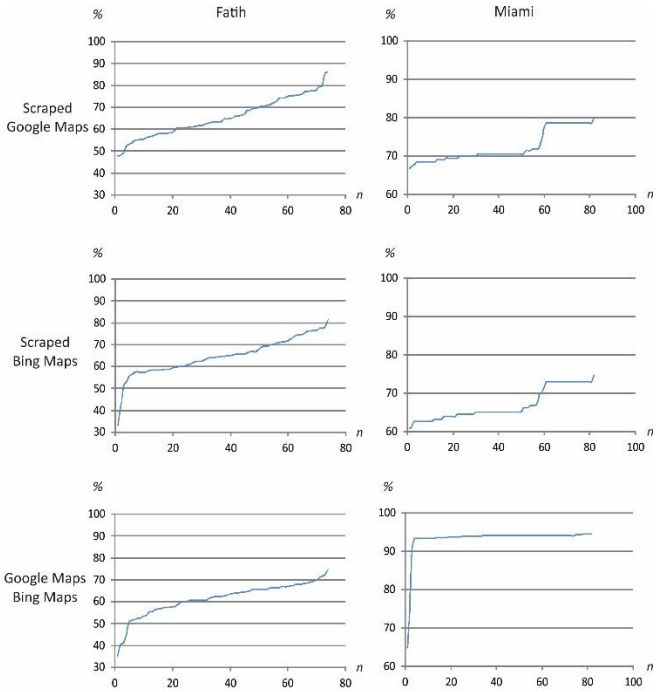


Figure 8. Address similarities for geocoding between a) Scraped and Google Maps data, b) Scraped and Bing Maps data c) Google Maps and Bing Maps data.

The similarity values between the postal addresses obtained with the geocoding of both services in Table 7 have significant differences of approximately 30% for Fatih and Miami Beach, individually. The main reason for the deficiency of geocoding in Fatih is that normalized addresses of both services cannot provide proper and sufficient information for sub-regions such as neighborhood, county, and city. Besides, Google Maps can return a neighborhood name that has not been parsed from a postal address, and uses various abbreviations for road types, while Bing Maps does not present any neighborhood information and abbreviations. As both services use similar abbreviations in Florida, such as Avenue: Ave, Street: St, Florida: FL, and present the addresses following the USA standard format, similarity values of their geocoding results are significantly high. The main factor that prevents reaching a similarity rate of 100% between the results in Google Maps and Bing Maps is the fact that Bing Maps service does not present country information in its normalized addresses.

5. CONCLUSION

The postal address is one of the most crucial input data used in the geocoding process. In this study, geocoding from postal addresses are comparatively analyzed through Google Maps and Bing Maps services. Test conducted in two individual study regions suggests that those services have different geocoding capabilities in Turkey and the USA. The positional accuracy of geocoding in the USA is almost similar for Google Maps and Bing Maps online geocoding services, whereas, in Turkey, Google Maps provides a few better-geocoding results than Bing Maps does.

One of the most critical factors that increase the quality of geocoding is the actuality of standard addresses stored in the database of services. The addresses in databases of each service are defined according to their standards. The address formats of both services in the USA have almost similar standards, yet they may be quite different in Turkey (e.g., the regular use of an abbreviation or the order of a parsed element in a normalized address). The discrepancies among geocoding services in Turkey concerning address standardization still exist among the public institutions. Many of them use their address format and share information with other stakeholders. It is a complex problem that needs to be solved in collaboration with the official institutions and local governments in Turkey as soon as possible.

One of the main problems for the online geocoding process is to determine the actual location of the point belonging to an address on a geospatial database. The correct location is the front edge of a building, depending on the actual status of the main door used to enter the building. However, the services mark the address points of the buildings generally, regardless of the main entrances. In the future, the authors aim to carry out a study on deriving the entrances of buildings automatically to determine the actual location of address points provided by the services more appropriately. Thus, the quality of the geocoding will be improved.

REFERENCES

Armstrong, M.P. and Tiwari, C. (2008). Geocoding methods, materials, and first steps toward a geocoding error budget. In Rushton G, Armstrong M P, Gittler J, Greene B R, Pavlik C E, West M M, and Zimmerman D

L (eds) Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice. Boca Raton, FL, CRC Press: 11–35.

Bichler, G. and Balchak, S. (2007). Address matching bias: Ignorance is not bliss. Policing: An International Journal of Police Strategies & Management, 30 (1), 32–60.

Bilgi, S., Gulnerman, A., Arslanoğlu, B., Karaman, H. and Ozturk, O. (2019). Complexity measures of sport amenities allocation in urban area by metric entropy and public demand compatibility. International Journal of Engineering and Geosciences, 4 (3), 141–148.

Cayo, M.R. and Talbot, T.O. (2003). Positional error in automated geocoding of residential addresses. International Journal of Health Geographics, 2 (10).

Cetl, V., Kliment, T. and Jogun, T. (2016). A comparison of address geocoding techniques–a case study of the city of Zagreb, Croatia. Survey Review, 50 (359), 97–106.

Chow, T.E., Dede-Bamfo, N. and Dahal, K.R. (2016). Geographic disparity of positional errors and matching rate of residential addresses among geocoding solutions. Annals of GIS, 22 (1), 29–42.

Goldberg, D.W., Ballard, M., Boyd, J.H., Mullan, N. Garfield, C., Rosman, D., Ferrante, A.M. and Semmens, J.B. (2013). An evaluation framework for comparing geocoding systems. International Journal of Health Geographics, 12 (1), 50.

Haines, E. (1994). Point in polygon strategies. Graphics gems IV, 994, 24–26.

Hart, T.C. and Zandbergen, P.A. (2013). Reference data and geocoding quality: Examining completeness and positional accuracy of street geocoded crime incidents. Policing: An International Journal of Police Strategies & Management, 36 (2), 263–294.

Karimi, H.A., Durcik, M. and Rasdorf, W. (2004). Evaluation of uncertainties associated with geocoding techniques. Computer-Aided Civil and Infrastructure Engineering, 19 (3), 170–185.

Kilic, B. and Gulgen F. (2017). Coğrafi Kodlama için Adres Standardizasyonu. 16. Türkiye Harita Bilimsel ve Teknik Kurultayı. 3-6 May, Ankara, Turkey.

Kilic, B. and Gulgen F. (2019). Türkiye'de Kullanılan Posta Adres Bilgilerinde Uyum Düzeylerinin Belirlenmesi. Harita Dergisi, 161, 26–34.

Levenshtein, V.I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics Doklady, 10 (8), 707–710.

Li, B., Zhang, X. and Chen, Y. (2010). Automatic construction and visualization of address models. In 2010 Sixth International Conference on Natural Computation (Vol. 6, pp. 2894-2897), IEEE.

McElroy, J.A., Remington, P.L., Trentham-Dietz, A., Robert, S.A. and Newcomb, P.A. (2003). Geocoding addresses from a large population-based study: lessons learned. Epidemiology, 14 (4), 399–407.

Mitchell, R. (2015). Web scraping with Python: collecting data from the modern web. O'Reilly Media, Inc.

Mulazimoglu, E. and Basaraner, M. (2019). User-centred design and evaluation of multimodal tourist maps. International Journal of Engineering and Geosciences, 4 (3), 115–128.

Qin, X., Parker, S., Liu, Y., Graettinger, A.J. and Forde, S. (2013). Intelligent geocoding system to locate traffic crashes. Accident Analysis & Prevention, 50, 1034–1041.

Ratcliffe, J.H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. International Journal of Geographical Information Science, 18 (1), 61–72.

Rodrigues, F. (2010). POI Mining and Generation. PhD Thesis. Faculty of Sciences and Technology (FCTUC) Department of Informatics Engineering - University of Coimbra. Portugal.

Roongpiboonsopit, D. and Karimi, H.A. (2010a). Comparative evaluation and analysis of online geocoding services. International Journal of Geographical Information Science, 24 (7), 1081–1100.

Roongpiboonsopit, D. and Karimi, H.A. (2010b). Quality assessment of online street and rooftop geocoding services. Cartography and Geographic Information Science, 37 (4), 301–318.

Rushton, G., Armstrong, M.P., Gittler, J., Greene, B.R., Pavlik, C.E., West, M.M. and Zimmerman, D.L. (2006). Geocoding in cancer research: a review. American Journal of Preventive Medicine, 30 (2), S16–S24.

Tian, Q., Ren, F., Hu, T., Liu, J., Li, R. and Du, Q. (2016). Using an Optimized Chinese Address Matching Method to Develop a Geocoding Service: A Case Study of Shenzhen, China. ISPRS International Journal of Geo-Information, 5 (5), 65.

Ward, M.H., Nuckols, J.R., Giglierano, J., Bonner, M.R., Wolter, C., Airola, M., Mix, W., Colt, J.S. and Hartge, P. (2005). Positional accuracy of two methods of geocoding. Epidemiology, 16 (4), 542–547.

Yang, D.H., Bilaver, L.M., Hayes, O. and Goerge, R. (2004). Improving geocoding practices: evaluation of geocoding tools. Journal of Medical Systems, 28 (4), 361–370.

Yildirim, V., Yomralioglu, T., Nisanci, R. and Inan, H. (2014). Turkish street addressing system and geocoding challenges. Proceedings of the Institution of Civil Engineers, 167 (2), 99–107.

Zandbergen, P.A. and Green, J.W. (2007). Error and bias in determining exposure potential of children at school locations using proximity-based GIS techniques. Environmental Health Perspectives, 115 (9), 1363–1370.

Zandbergen, P.A. (2008). A comparison of address point, parcel and street geocoding techniques. Computers, Environment and Urban Systems, 32 (3), 214–232.

Zandbergen, P.A. (2011). Influence of street reference data on geocoding quality. Geocarto International, 26 (1), 35–47.

URL 1. Google Maps Platform – Geocoding API documentation, https://developers.google.com/maps/documentation/geocoding/start?hl+=+de, accessed on 3 October 2019.

URL 2. Bing Maps Dev Center – Find location by address (geocode), https://www.bing.com/api/maps/sdk/mapcontrol/isdk/searchbyaddress, accessed on 3 October 2019.

IJEG
International Journal of Engineering and Geosciences

IJEG
Volume 5 - Issue 2

ARTICLES