

A method for Assemblage of an Open Access Data Set for Research in Geomagnetic Effects on GPS/GNSS Ionospheric Delay in Sub-equatorial Regions

Ivan Hedi¹, Enes Ciriković¹, Željko Borković² and Renato Filjar^{1, 2, 3*}

¹Virovitica University of Applied Sciences, Virovitica, Croatia

²Laboratory for Spatial Intelligence, University of Applied Sciences Hrvatsko Zagorje Krapina, Krapina, Croatia

³Faculty of Engineering, University of Rijeka, Rijeka, Croatia

*E-mail of corresponding author: renato.filjar@gmail.com

Abstract: Space weather, geomagnetic, and ionospheric effects are major sources of Global Navigation Satellite System's (GNSS) Positioning, Navigation, and Timing (PNT) service degradation. Researchers studying the problem often face an obstacle in missing large data sets comprising ionospheric effects variables and space weather ones for their studies. Here a formal assemblage method is proposed for data set assemblage for GNSS ionospheric effects assessment that utilises experimental sets of space weather/geomagnetic observations, and raw GNSS pseudorange observations polluted with uncorrected ionospheric delay effects to estimate Total Electron Content (TEC) as a target variable outlining the ionospheric effects on GNSS. The five Observations Quality Principles (OQP) are proposed for the observations to be used in statistical learning model development to comply to. The proposed assemblage method is demonstrated in a particular case study of an assembled data set for Darwin, NT, Australia, in sub-equatorial zone, using a tailored software developed in the R environment for scientific computing, and a freely available TEC estimation software. The assembled massive data obtained in the demonstration of the proposed method is provided in the open-access manner in support of the international scientific community.

Received 10 April 2023

Accepted 20 May 2023

Published 30 May 2023

Keywords: GPS/GNSS, ionospheric delay, geomagnetic conditions, sub-equatorial region, data.

1. Introduction

Satellite navigation has become one of the few cornerstones of the modern civilisation, as it empowers, as an embedded and enabling technology, a vast range of technological, and socio-economic applications, systems and services [1]. Its provision of Positioning, Navigation, and Timing (PNT) services through Global Navigation Satellite Systems (GNSS), such as the US GPS, Russian Federation's Glonass, China's Beidou, and the EU's Galileo, has matured to the levels which require stable, sustainable, and standardised provision of PNT Quality of Service (QoS) [2]. Demand and requests for PNT mean the effort should be made in confinement and mitigation of inherited and emerging shortcomings and vulnerabilities of the system, spectrum, signals, and information being deployed in the provision, and used in exploitation of the GNSS PNT services [3].

Ionospheric effects, the ionospheric delay in particular, emerge as the primary source of the GNSS PNT QoS degradation, particularly of GNSS-based

positioning accuracy [4]. Approaches utilised for mitigation of ionospheric delay effects on the GNSS positioning accuracy may involve: (i) provision of standard global ionospheric correction models, such as Klobuchar model [4, 5], (ii) development and provision of local or regional ionospheric correction models [4, 6, 7], (iii) deployment of advanced satellite-based positioning methods, such as: Real Time Kinematics, RTK, or Precise Point Positioning, PPP, [4], and (iv) utilisation of self-adaptive positioning environment-aware GPS/GNSS position estimation [8]. Additionally, the awareness of ionospheric conditions may be utilised for the assessment of risk for GNSS utilisation in position-aware applications, such as electronic navigation, robotics, location-based services, autonomous vehicles and aircrafts, and others [9].

Current state-of-the-art in related disciplines, including statistics and computer science, offers prospect in the predictive ionospheric model development using methods of statistical/machine learning [6, 10], based on experimentally collected observations resulting from the ionospheric, geomagnetic, and space weather processes and conditions [8].

A growing number openly available internet-based databases allows researchers for embracing the trends in



development of new, perspective, and productive models to improve the GNSS PNT positioning performance through more accurate prediction of the GNSS ionospheric delay. However, a researcher may encounter a number of obstacles and shortcomings, including, but not limited to: partially restricted access to the original data, unknown or complicated formats of raw observations, different sampling methods/times used for different variables, and different formats used by different sources of observations. Finally, researchers who undergo the whole process of observations collection, aggregation, collation, data cleaning, parsing, and formatting rarely publish datasets in an open-access manner, leaving the other potentially interested research parties with the same challenging and daunting task for their research. For the reasons above, studies of ionospheric effects on GNSS positioning performance are usually constrained both geographically, to the place of data collection, and in time, spreading to a couple of days when an ionospheric anomaly lasts. Such approach downgrades prospects for learning from data and development of understanding of the long-term ionospheric effects on GNSS positioning performance.

Here a formal methodology for GNSS – space weather data sets assembly is proposed, and demonstrated in the case of a data set creation for the purpose of an ionospheric correction model development and GNSS positioning performance ionospheric effects assessment in a geographically constrained environment (Darwin, NT, Australia), and with the reasonably wide set of potential deterioration scenarios considered by extension of the time span to the whole calendar year of 2014.

The manuscript reads as follows. This Session sets the stage by stating the problem and motivation for research. Section 2 outlines the proposed methodology for data set assemblage, and the openly accessible internet sources that may be used for raw observations collection. Section 3 presents the results of the proposed method's demonstration in the particular case, as well as a massive data set assembled and provided to interested research parties in an open manner. Section 4 discusses the proposed methodology, identifies its shortcomings, and summarises the contribution of the presented research.

2. Method and material

Assembling a relevant and related data set for the GNSS ionospheric effects mitigation purposes emerges as a research problem. Here a method for such an assemblage is proposed, together with a set of openly accessible data sources, which will yield a data set suitable for the GNSS ionospheric correction modelling, ionospheric effects on GNSS positioning accuracy assessment, and risk assessment of GNSS ionospheric effects on the QoS of GNSS-based applications.

The GNSS ionospheric delay results from the passage of a satellite radio wave signals through non-homogeneous ionised layers of the upper Earth's atmosphere, known collectively as the ionosphere [11]. The amount of delay t_{iono} is directly proportional to a single parameter resulting from the ionospheric conditions, called Total Electron Content (TEC) given in (1), with $N(h)$ denoting the vertical ionospheric profile, a

surface density of electrons on height h [km] above the sea level, in [electrons/m²], and h_{min} and h_{max} denoting the height of the lower bound, and upper bound of the ionosphere, respectively.

$$TEC = \int_{h=h_{min}}^{h=h_{max}} N(h)dh \quad (1)$$

The GNSS ionospheric delay affects GNSS pseudorange measurements by creating an error which further propagates into GNSS position estimation error [4, 8]. GNSS ionospheric delay may be determined using the expression (2), with f denoting the frequency of the satellite radio wave carrier, as derived from the Appleton-Hartree formula [11, 12].

$$t_{iono} = \frac{40.3}{f^2} \cdot TEC \quad (2)$$

2.1. Method

Here a formal method for data set assemblage for GNSS ionospheric effects assessment is proposed that utilises experimental sets of space weather/geomagnetic observations and raw GNSS pseudorange observations polluted with uncorrected ionospheric delay effects.

Accomplishments of a successful data set assemblage calls for requirements on the raw observations, frequently obtained from various sources. It is proposed here that observations used in statistical learning model development should comply to the five Observations Quality Principles (OQP), as follows. (i) Observations should come from the trusted sources, who guarantee the objective and unbiased process of data collection. (ii) Observations should focus on phenomena of research interest, preserving the effects of the studied phenomena in data, while suppressing and eliminating the effects of the others, (iii) Observations should be of variables expected to serve as both predictors (input variables) and target (the output variable). (iv) Observations should be collected in all related, or anticipated scenarios, both in time and space. (v) Data should be structured in a known format to allow for machine reading.

Assuming compliance to aforementioned OQP principles, the assemblage method is proposed, as depicted in Figure 1.

Raw GPS pseudoranges collection comprises tasks of identification and collection of data, complying to the before mentioned five OQP. Data collected may either enter the assemblage procedure directly, or be stored locally on the computing system beforehand.

TEC estimation from GNSS pseudoranges may be performed using the common method utilising dual frequency GPS/GNSS observations. Let us assume a reference station performs simultaneous GPS pseudorange measurements on two carrier frequencies, L1 and L2. The process yields two pseudorange observations ρ_1 and ρ_2 , on frequencies L1 and L2, respectively, for every time instant.

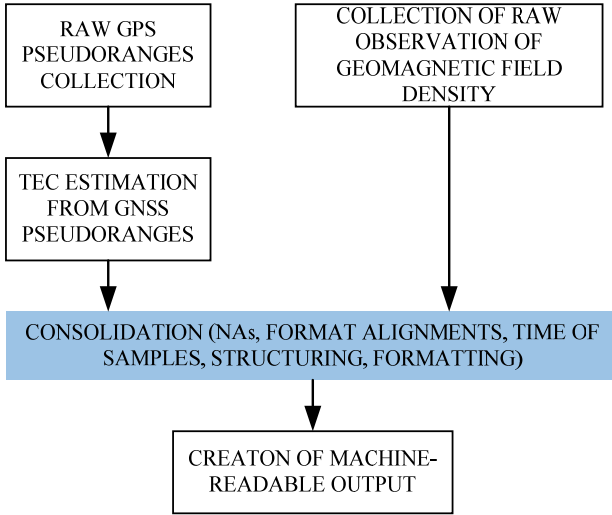


Fig. 1. Relationship between step size and path length.

With two time series of observations, TEC may be estimated using the method depicted with (3) ... (5), where STEC denotes slant (actually observed) TEC, at elevation angle E in [rad], b_s denotes satellite bias in [m], and b_r denotes receiver bias in [m].

$$STEC = \frac{\rho_2 - \rho_1 - b_s - b_r}{40.31 \cdot \left[\frac{1}{f_2^2} - \frac{1}{f_1^2} \right]} \quad (3)$$

Satellite bias may be determined using trusted services, such as the one provided to the international research community by University of Bern in Switzerland [4]. Receiver bias may be estimated using various methods involving post-analysis of time series of TEC estimates [4].

Furthermore, TEC is determined in a standardised manner for the comparison purpose, as a so-called Vertical TEC (VTEC) given as a projection of STEC on a vertical at the point of observer (GNSS receiver observing pseudoranges), using the mapping function $m(E)$, as given in (4) and (5) respectively.

$$m(E) = \frac{1}{\sqrt{1 - \left(\frac{R_{Earth}}{R_{Earth} + h} \cdot \cos(E) \right)^2}} \quad (4)$$

$$TEC = \frac{STEC}{m(E)} \quad (5)$$

Referring to (4) and (5), R_{Earth} denotes radius of Earth's sphere, in [km], and h denotes height above the average sea level, in [km]

Collection of raw observations of geomagnetic field density components is requested as the result of hypothesis of TEC relation to geomagnetic conditions, which also determines ionospheric conditions [4, 7, 13]. Collected data should comply to the five OQC principles, including the condition of observations collection in the close geographical vicinity of the GNSS pseudorange collection site.

Consolidation tasks of the proposed data assemblage process comprise dealing with the NA (Not Available) observations, formatting alignments for data provided by different sources with differing formatting standards,

resolving problems of data observed at different time instants or with the different sampling frequency, and definition of the conversion into output structure and data format.

As the result of the assemblage process, the output data set is created, with the requested structure and machine-readable data format.

2.2. Material

Observations should comply to the five OQP. Here the following sources are used for experimental observations access: (i) International GNSS Service (IGS) raw GNSS pseudorange repository [14], with the observations taken by world-wide network of stationary GNSS reference stations, and (ii) INTERMAGNET repository [15], with the observations taken by the world-wide network of stationary geomagnetic reference stations.

The IGS reference stations network has been established to allow for continuous raw GNSS observations collection, with suppression of all sources of pseudorange errors, apart from the ionospheric ones. Strict requirements are imposed to operators of IGS reference station to ensure the quality of observations. Continuous 24-hours-a-day observations ensure collection of data in all real scenarios of GNSS positioning degradation scenarios due to ionospheric, geomagnetic, and space weather conditions. Methodology of observation collection, including description of utilised devices, is transparently presented, and the quality of data monitoring process consistently and transparently operated. In that sense, the IGS is considered a trusted source of raw GNSS pseudorange observations from which TEC may be estimated as a target variable, in compliance to the five OQP.

The assessment returns the same results for the INTERMAGNET network of geomagnetic observations, rendering it a trusted sources of geomagnetic field density components observations, as presumed predictor variables of TEC.

3. Research results

The GNSS data assemblage method proposed in Section 2 is implemented in a set of software packages, and demonstrated in the case of GNSS and geomagnetic observations assemblage for the case of sub-equatorial region GPS ionospheric delay assessment in relation to components of geomagnetic field density. Sources of data involves the IGS reference station at Darwin, NT, Australia, as the source of daily raw dual-frequency GPS observations at 30 s sampling time, and the INTERMAGNET reference station at Kakadu, NT, Australia, as the source of daily geomagnetic field density components observations taken at 1 min sampling period using precise magnetometers. The two reference stations are set at distance of approx. 195 km, which can be seen as a very close distance, considering the nature of the observed phenomena.

The data assemblage procedure is implemented, as follows.

The raw GPS pseudorange observation collection is performed for the whole year 2014 and for the IGS

Darwin data using a bespoke software developed by authors in the open source R environment for statistical computing in version 4.2.2 [15]. Raw GPS observations were taken every 30 s, 24 hours a day. Collected daily data sets are stored on a local disc of a computer, where preparatory analysis and consolidation take place.

The raw geomagnetic field density components observation collection is performed for the whole year 2014 and for the INTERMAGNET Kakadu data using another bespoke software developed by authors in the open source R environment for statistical computing in version 4.2.2 [16]. Geomagnetic field density components observation s were taken every minute, 24 hours a day. Collected daily data sets are again stored on a local disc of a computer, where preparatory analysis and consolidation take place.

The TEC estimation from dual frequency GPS pseudorange observations is performed using the GPS-TEC [17, 18], a third party software developed by Dr. Gopi Seemala, and provided freely to the international research community. Dr. Seemala's software performs the TEC estimation, as presented in Section 2, using satellite bias estimates from University of Bern, and determining the receiver bias by setting the lowest estimated nightly TEC of a day under consideration to zero.

Consolidation tasks of the data set assemblage is again conducted by a tailored R-based software developed by authors. The tasks involve the removal of cases with NA, for whatever reason they appear, time formats alignment, harmonisation of data by time stamps (observations without counterparts taken at the same time are removed from the set), as well as structuring data by variables a storing them in a table format. Finally, the obtained table is stored to a Comma Separated Value (CSV) textual file, which is proven to be machine-readable. As the result of the proposed data assemblage procedure, a massive data set of 522 298 individual instances is created, providing TEC and geomagnetic field density data in Darwin, NT, Australia throughout 2014 for the purpose of the research.

The file comprising the assembled data set carries approx 28.5 MB of data, and is provided as a Supplementary Material to this manuscript in the open access manner to the international research community at the internationally recognised repository *figshare* [19].

4. Discussion and conclusion

Presented research addresses the problem GNSS PNT performance researchers face in modern times, where diversity of data resources and formats prevent researchers from having a structured and consistent data sets to utilise for accomplishment of improved understanding of geomagnetic and ionospheric effects on GNSS positioning accuracy. Here a formal method is proposed that outlines tasks for multi-source GNSS and geomagnetic data assemblage into a single data set, which can be used for the GNSS ionospheric correction model development using statistical/machine learning methods, as well as for risk assessment of the ionospheric/geomagnetic impact on Quality of Service of

GNSS-based applications, systems and services. Further to this, the set-of-five Observation Quality Principles are proposed as the set of standard requirements for data sets to be assembled.

The proposed method is implemented in a combination of tailored software developed in the R environment for statistical computing, and a freely available GPS-TEC estimation software. As a result, a massive data set is assembled for the case of sub-equatorial region, comprising Total Electron Content (TEC) values derived from experimentally collected raw (un-corrected) dual frequency GPS pseudorange observations of a trusted third-party, as well as observations of geomagnetic field density components for the same area, produced by another trusted third-party. The assembled data set for Darwin, NT, Australia in 2014, obtained as a result of demonstration of the principle, is rendered available in the open-access manner, and presented as a Supplementary Material to this manuscript.

Authors continue work on the software consolidation to publish the R-based implementation of the proposed data assemblage method as an open-source and open-access software, available to the international research community.

References

- [1] Filjar, R, Damas, M C, Iliev, T B. (2020). Resilient Satellite Navigation Empowers Modern Science, Economy, and Society. *IOP Conf. Ser: Mater Sci Eng* vol. 1032, 012001, doi:10.1088/1757-899X/1032/1/012001
- [2] EUSPA. (2021). EUSPA User Consultation Platform (UCP) GNSS User Needs and Requirements Library of books for: Aviation, Agriculture, Location-Based Services, Maritime and Inland Waterways, Rail, Public Transport, Road, Surveying, Time & Synchronisation, and Space. European Space Programme Agency (EUSPA). Prague, Czechia. Available at: <https://www.euspa.europa.eu/euspace-applications/euspace-users/user-needs-and-requirements>
- [3] HM Government Office for Science. (2018). Satellite-Derived Time and Position: A Study of Critical Dependencies. HM Government of the UK and NI. Dostupno na: <https://bit.ly/2E2STn>
- [4] Sanz Subirana, J., Juan Zornoza, J. M., and Hernandez-Pajares, M. (2013). *GNSS Data Processing – Volume I: Fundamentals and Algorithms*. ESA Communications. Available at: <http://bit.ly/1tDzJIQ>
- [5] Strang, G, Borre, K. (1997). *Linear algebra, geodesy, and GPS*. Wellesley-Cambridge Press. Wellesley, MA
- [6] Filić, M, Filjar, R. (2018). Modelling the Relation between GNSS Positioning Performance Degradation, and Space Weather and Ionospheric Conditions using RReliefF Features Selection. *Proc of 31st International Technical Meeting ION GNSS+ 2018*, pp. 1999-2006. Miami, FL. doi: <https://doi.org/10.33012/2018.16016>
- [7] Filjar, R, Weintrit, A, Iliev, T, Malčić, G, Jukić, O, Sikirica, N. (2020). Predictive Model of Total Electron Content during Moderately Disturbed Geomagnetic Conditions for GNSS Positioning Performance Improvement. *Proc FUSION2020*, pp. 256-262. Pretoria, South Africa. doi: <https://doi.org/10.23919/FUSION45008.2020.9190264>
- [8] Filjar, R. (2022). An application-centred resilient GNSS position estimation algorithm based on positioning environment conditions awareness. *Proc ION International Technical Meeting (ITM) 2022*. Long Beach, CA.

- [9] Malić, E, Sikirica, N, Špoljar, D, Filjar, R. (2023). A method and a model for risk assessment of GNSS utilisation with a proof-of-principle demonstration for polar GNSS maritime applications. *Conference Proceedings of TransNav 2023 Conference*. Gdynia, Poland.
- [10] James, G, Witten, D, Hastie, T, and Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R*, 2nd ed. Springer New York. New York, NY. Available at: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf
- [11] Davies, K. (1990). *Ionospheric Radio*. Peter Peregrinus Ltd. London, UK.
- [12] Appleton, E. V. (1947). *The ionosphere*. Nobel Lecture, Available at: <https://www.nobelprize.org/uploads/2017/01/appleton-lecture-new.pdf>
- [13] Filjar, R, Kos, S, Krajnovic, S. (2013). Dst index as a potential indicator of approaching GNSS performance deterioration. *Journal of Navigation*, vol. 66, issue 1, pp. 149-160. doi:10.1017/S037346331200029X
- [14] IGS. (2023). International GNSS Service GNSS RINEX data archive (repository hosted by NASA). Available at: <ftp://cddis.nasa.gov/gnss/data/daily/>
- [15] INTERMAGNET. (2021). The International Real-time Magnetic Observatory Network data archive. Available at: <https://intermagnet.github.io>
- [16] R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>
- [17] Seemala, G. (2023). *GPS-TEC Programme rev 3.00*. Available at: <https://seemala.blogspot.com/>
- [18] Sikirica, N, Dimc, F, Jukić, O, Iliev, T B, Špoljar, D, Filjar, R. (2021). A Risk Assessment of Geomagnetic Conditions Impact on GPS Positioning Accuracy Degradation in Tropical Regions Using Dst Index. *Proc ION ITM 2021*, pp. 606-615. San Diego, CA. doi: 10.33012/2021.17852
- [19] Hedi, I, Ciriković, E, Borković, Ž, Filjar, R. (2023). Hedji, Cirikovic, Borkovic, Filjar, JCIEES 2023 manuscript, Supplementary material. doi:10.6084/m9.figshare.22579786. Available at: https://figshare.com/articles/dataset/Hedji_Cirikovic_Borkovic_Filjar_JCIEES_2023_manuscript_Supplementary_material/22579786