

DINData: A Windows software to reformat and clean digisonde numeric dataset

Bello Saeed Abioye¹, Yusuf K A¹, Ige S O¹, Joseph O A¹, Oluwafemi L D¹, Abdullah Mardina², and Abdul Hamid Nurul Shazana³

¹University of Ilorin

²University Kebangsaan Malaysia

³Universiti Kebangsaan Malaysia

November 16, 2022

Abstract

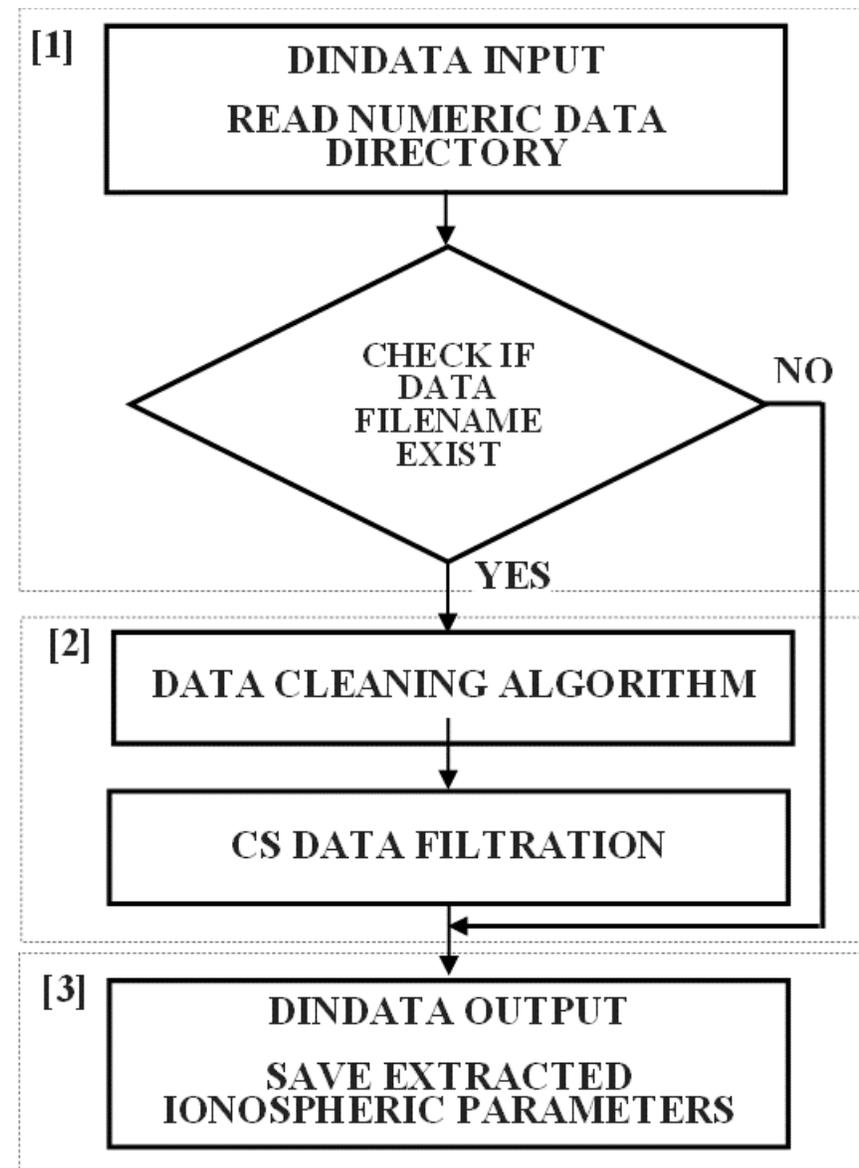
The ionospheric numerical ionogram dataset on the Digisonde Ionogram Database (DIDBase) hosted by the Global Ionospheric Radio Observatory (GIRO) provides an important opportunity to study the bottomside ionospheric electron density profile. The study presents the description of software developed for digisonde numerical dataset that can be obtained from DIDBase. The developed DIDBase numerical data reader and cleaner named Digisonde Numeric Data (DINData) software can find missing data periods in the dataset and resequencing the date and time stamp on the dataset. The DINData software also allows users to apply settings to check the integrity of the scaled ionogram using the ionogram confidence score (CS) value. The DINData application is written in the Python programming language. The graphical user interface (GUI) of the software was developed using Kivy, a cross-platform Python GUI framework that is built on OpenGL. The Kivy framework allows the source code of the application to be packaged into various operating systems. The software currently runs on the Windows operating system and has a user-friendly graphical user interface to accept alphanumeric input entries.

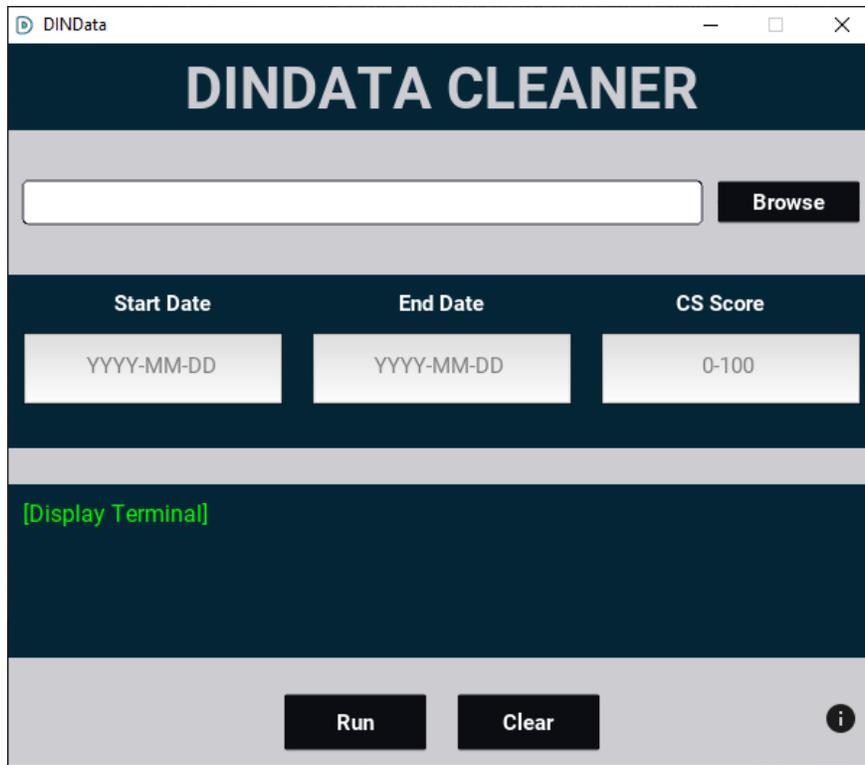
Hosted file

essoar.10512053.1.docx available at <https://authorea.com/users/524338/articles/595446-dindata-a-windows-software-to-reformat-and-clean-digisonde-numeric-dataset>

Hosted file

table.docx available at <https://authorea.com/users/524338/articles/595446-dindata-a-windows-software-to-reformat-and-clean-digisonde-numeric-dataset>





```

DIDBGetValues - Notepad
File Edit Format View Help
# Global Ionospheric Radio Observatory
# GIRO Tabulated Ionospheric Characteristics, Version 1.0 Revision B
# Generated by DIDBGetValues on 2021-12-22T16:59:36.269Z
#
# Location: GEO 8.5N 4.5E, URSI-Code IL008 ILORIN
# Instrument: Ionosonde, Model: DPS-4
#
# Query for measurement intervals of time:
# 2010-03-02T21:00:00.000Z - 2010-12-03T03:00:00.000Z
#
# Data Selection:
# CS is Autoscaling Confidence Score (from 0 to 100, 999 if manual scaling, -1 if unknown)
# foF2 [MHz] - F2 layer critical frequency
# hmF2 [km] - Peak height F2-layer
# TEC [10^16 m^-2] - Total electron content
# B0 [km] - IRI thickness parameter
# B1 [] - IRI profile shape parameter
#
# All GIRO measurements are released under CC-BY-NC-SA 4.0 license
# Please follow the Lowell GIRO Data Center RULES OF THE ROAD
# https://ulcar.uml.edu/DIDBase/RulesOfTheRoadForDIDBase.htm
# Requires acknowledgement of IL008 data provider
#
#Time CS foF2 QD hmF2 QD TEC QD B0 QD B1 QD
2010-03-26T16:45:00.000Z 100 8.775 // 376.2 // 38.4 // 168.3 // 2.53 //
2010-03-26T17:00:00.000Z 100 8.806 // 372.6 // 35.1 // 161.6 // 2.11 //
2010-03-26T17:15:00.000Z 50 6.863 // 288.5 // 10.9 // 80.3 // 1.74 //
2010-03-26T17:30:00.000Z 75 8.950 // 359.1 // 28.3 // 121.5 // 2.30 //
2010-03-26T17:45:00.000Z 100 9.050 // 357.9 // 27.1 // 110.8 // 2.46 //
2010-03-26T18:00:00.000Z 100 9.050 // 362.8 // 26.5 // 103.4 // 2.89 //
2010-03-26T18:45:00.000Z 100 8.600 // 390.3 // 23.2 // 103.4 // 3.08 //
2010-03-26T19:00:00.000Z 100 8.400 // 414.0 // 30.2 // 123.6 // 4.82 //
2010-03-26T19:15:00.000Z 75 8.100 /_ 412.0 /_ 21.1 /_ 107.6 /_ 2.85 /_
2010-03-26T19:30:00.000Z 999 8.050 /_ 426.4 /_ 20.0 /_ 108.7 /_ 2.36 /_
2010-03-26T19:45:00.000Z 75 3.500 // 338.9 // 2.0 // 64.9 // 1.39 //
2010-03-26T20:15:00.000Z 0 6.700 // --- --- --- --- ---
2010-03-26T20:30:00.000Z 0 6.750 // --- --- --- --- ---
2010-03-26T20:45:00.000Z 0 6.300 // --- --- --- --- ---
2010-03-26T21:00:00.000Z 75 8.000 // 290.1 // 5.6 // 16.8 // 5.61 //
2010-03-26T21:15:00.000Z 50 7.850 // 307.7 // 7.8 // 42.1 // 2.97 //
2010-03-26T21:30:00.000Z 0 8.700 // --- --- --- --- ---
2010-03-26T21:45:00.000Z 50 10.700 // 298.3 // 13.5 // 48.8 // 1.73 //
2010-03-26T22:00:00.000Z 50 6.500 // 258.5 // 5.5 // 36.8 // 4.65 //
2010-03-26T22:15:00.000Z 33 6.650 // 255.4 // 4.2 // 35.0 // 2.06 //
2010-03-26T22:30:00.000Z 33 7.350 // 280.2 // 3.0 // 21.9 // 1.07 //
2010-03-26T23:00:00.000Z 50 2.850 // 244.1 // 0.4 // 17.6 // 1.00 //
2010-03-26T23:15:00.000Z 50 8.300 // 292.2 // 7.0 // 35.0 // 2.83 //
2010-03-26T23:30:00.000Z 50 6.950 // 271.2 // 4.8 // 36.7 // 1.98 //
2010-03-26T23:45:00.000Z 0 7.950 // --- --- --- --- ---
2010-03-27T00:00:00.000Z 50 8.400 // 256.6 // 5.2 // 29.9 // 1.55 //
2010-03-27T00:15:00.000Z 33 7.100 // 252.7 // 5.4 // 37.6 // 2.28 //
2010-03-27T00:30:00.000Z 33 5.800 // 255.7 // 5.4 // 48.0 // 4.52 //
2010-03-27T00:45:00.000Z 50 5.700 // 245.2 // 4.0 // 43.1 // 2.50 //
2010-03-27T01:00:00.000Z 50 6.200 // 262.3 // 4.4 // 45.2 // 1.73 //
2010-03-27T01:15:00.000Z 50 5.300 // 275.7 // 4.9 // 56.0 // 3.51 //
2010-03-27T01:30:00.000Z 33 4.750 // 257.9 // 2.7 // 42.4 // 2.33 //

```

DINData: A Windows software to reformat and clean digisonde numeric dataset

S.A. Bello¹, K.A. Yusuf¹, S.O. Ige¹, O.A. Joseph², L.D.Oluwafemi¹, M. Abdullah^{3,4}, N.S.A. Hamid^{3,5}

¹Department of Physics, Faculty of Physical Sciences, University of Ilorin, Nigeria

²Department of Mathematics, Faculty of Physical Sciences, University of Ilorin, Nigeria

³Space Science Centre (ANGKASA), Institute of Climate Changes, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor, Malaysia

⁴Department of Electrical, Electronics & Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia.

⁵Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

Abstract

The ionospheric numerical ionogram dataset on the Digisonde Ionogram Database (DIDBase) hosted by the Global Ionospheric Radio Observatory (GIRO) provides an important opportunity to study the bottomside ionospheric electron density profile. The study presents the description of software developed for digisonde numerical dataset that can be obtained from DIDBase. The developed DIDBase numerical data reader and cleaner named Digisonde Numeric Data (DINData) software can find missing data periods in the dataset and resequencing the date and time stamp on the dataset. The DINData software also allows users to apply settings to check the integrity of the scaled ionogram using the ionogram confidence score (CS) value. The DINData application is written in the Python programming language. The graphical user interface (GUI) of the software was developed using Kivy, a cross-platform Python GUI framework that is built on OpenGL. The Kivy framework allows the source code of the application to be packaged into various operating systems. The software currently runs on the Windows operating system and has a user-friendly graphical user interface to accept alphanumeric input entries.

Keyword: Ionosphere, Ionosonde, Digisonde, GIRO, DIDBASE, GUI

1. Introduction

Close to a century now, the radio sounding observational procedure of Breit and Tuve (1925) is still a relevant and important observational technique in today's

ionospheric physics study. An ionospheric sounder or ionosonde as typically known is one of the oldest and widely used ionospheric remote sensing devices which works by transmitting radio pulses vertically and measuring the time t taken for the radio pulse to travel up to the ionosphere and return from it. (Schunk and Nagy, 2009). The height of reflection of the radio pulses can be calculated using the time delay, the electron density profile can be obtained from the corresponding frequency. The ionospheric round-trip delay time, t in the absence of magnetic field and collisions as a function of frequency, f can be expressed as

$$t(f) = \frac{2}{c} \int_0^{z(f_p)} \frac{dz}{\sqrt{1 - \frac{\omega_p^2(z)}{\omega^2}}} \quad (1)$$

The integration Eq. (1) is done from the location of the ionosonde transmitter to the height of the reflection point, $z(f_p)$. The symbols c , ω_p and ω indicates the velocity of light, plasma frequency and the frequency of the propagating radio signal. The altitude of reflection points and the corresponding frequencies form what is known as the ionogram. This ionogram is the principle of a complicated ionospheric record to interpret. One reason for this is due to the propagating radio signal emanating from the ionosonde do not travel with the speed of light rather with the group velocity. Also, the presence of a magnetic field creates magnetic birefracton and velocities due to multiple values of the refractive index. These, therefore, make the data extraction and interpretation of ionogram traces a complicated routine. Over the years, the digisonde research scientists and engineers at the University of Massachusetts Lowell, Center for Atmospheric Research (UMLCAR) have contributed enormously in the discovery and inventions of a series of technologies and mathematical equations that can enhance scaling and inverting the ionogram. For instance, the ‘‘Automatic Real-Time Ionogram Scaler with True Height’’ (ARTIST) software (Reinisch et al., 2005, Reinisch and Huang, 1983) and the true height inversion program NHPC (Huang and Reinisch, 1996, Huang and Reinisch, 2001) can invert virtual height, $h'(f)$ digitized traces from both digital or analogue ionogram. Each ionospheric layer has an electron density, $N(h)$ profile that is constructed in terms of the Chebyshev polynomials fitting algorithm. Both the ARTIST and NHPC have now been embedded in the SAO Explorer software (Khmyrov et al., 2008). The SAO Explorer software is an interactive ionogram processing tool for data visualization, interactive scaling, profile inversion and time-series data formatting.

Using a central data repository, Galkin et al. (2006) reported the possibility to manage both the Digisonde Ionogram Database (DIDBase) and the Drift Database (DriftBase) databases and make them accessible over the internet through SAO Explorer software. In recent times, the Global Ionospheric Radio Observatory (GIRO, Reinisch and Galkin, 2011) has become an important and essential virtual observatory for ionospheric studies. The GIRO portal comprises of several components namely; the digisonde network of stations, the DIDBase, DriftBase for plasma drift observational studies, travelling ionospheric distur-

bance (TID) (Reinisch et al., 2018, Paznukhov et al., 2012), Global Assimilative Modelling of Bottomside Ionospheric Timelines, GAMBIT (Galkin et al., 2015), Ray-Tracing through Realistic Ionosphere Explorer, RayTRIX, (Galkin et al., 2018) and a good numbers of UMLCAR developed software’s for ionospheric studies (e.g., Galkin et al., 2008, Khmyrov et al., 2008, Kozlov and Paznukhov, 2008). According to Galkin et al. (2018), GIRO archives measurement and ionospheric derived data in near real-time from more than sixty (60) ionosonde across twenty-six (26) countries and are available online at <http://giro.uml.edu>. The scaled numerical ionospheric characteristics data, which are the ionogram-derived data are accessible on the FastChar section on GIRO on the web link at <https://giro.uml.edu/didbase/scaled.php>. The numeric data provide easy access to all the ionosondes data parameters on the DIDBase which are typically downloaded as an ASCII file or CSV file. The numerical data also have a confidence score value (CS, Galkin et al., 2013) for each data entry. The CS grades the integrity of the data available for each period of observation. The timeseries data representation of these numerical dataset follows the ISO-8601 date format (Wolf and Wicksteed, 1998). This makes it easy to check for missing data periods and data ‘interval aberration’. Since, ionosondes typically sounds between 5 to 15 minutes, each dataset is expected within this interval timestamp, anything short of this can be referred to as data timestamp-interval aberration. However, there are few situations where the timestamp on the downloaded numerical ionospheric dataset span beyond this interval. This perhaps might be due to operational challenges caused by the equipment error, glitches or equipment reset/restart sequences. For the end-user of the dataset, the need for consistency in the data interval is imperative for ionospheric research work. For this reason, a Digisonde Numeric Data (DINData) software is developed to make necessary data manipulations that involves, sequencing of the data date-time range interval, filling up data gap and data CS selection. The selection of CS above 70% have been reported in a few our previous studies on the variations in the ionospheric bottomside profile parameters (Bello et al., 2019b, Bello et al., 2017a, Bello et al., 2017b, Bello et al., 2017c).

2. DINData: Description of the Software program

The quarterly hourly ionospheric numerical characteristics data published on GIRO can be downloaded and fed into the DINData software. The downloaded numeric data stored as a text file contains the standard ionospheric profile parameters below the peak of the ionospheric F2 layer. These parameters are important for the description of the bottomside electron density of the ionosphere. Table 1 gives the complete description of these datasets and their usual symbol or abbreviations. The software is written in Python programming language with Kivy framework (Phillips, 2014, Ulloa, 2013) used to develop the graphical user interface (GUI). Kivy is a Python GUI framework and is an open-source

library for developing cross-platform software (Phillips, 2014, Ulloa, 2013).

The DINData program is structured into three sections, namely the user input section, data cleaning algorithm and data output. The structure of the program is an adaptation of SEAIONO (Bello et al., 2019a) software. The SEAIONO is a software application developed at the Space Science Center (ANGKASA), Universiti Kebangsaan Malaysia for extracting numerical ionogram datasets from the Frequency Modulated Continuous Wave (FMCW) ionosonde. The FMCW ionosonde is a portable analogue type ionosonde in the South East Asia Low-Latitude Ionospheric Network, SEALION (Nozaki, 2009). Although, the data structure of the DIDBase and FMCW ionosondes are different, however, some important logic in SEAIONO software are borrowed for this currently developed software. The DINData program process flow is illustrated in Figure 1. The first section handles user-inputted data. It comprises the text input fields for accepting the filename (absolute full path) and date range (start and end date) and CS values. The screenshot of the DINData GUI is shown in Figure 2. The dateValidator and CSValidator methods are the first executed functions in the software. The dateValidator method ensures that the values of the text fields conform with the expected format. The method also initiates logic to check the dates range differences, by ensuring the start dates precede the end date.

The CSValidator method checks users inputted CS score is within 0 to 100 as expected. CS score closer to 100 indicates reliable scaled ionogram data. The correct entered date values are then passed to the dateGenerator method. The dateGenerator creates a sequence of date-time stamps using the user's inputted start and end date in the time interval of 15 minutes. The generated date periods follow the ISO-8601 date format (Wolf and Wicksteed, 1998). These dates are important for finding missing periods in the dataset. Before section 2 is activated, the filename supplied is verified. If the file exists, the ionogram-NumReader method is invoked to read the numerical ionogram into a Pandas DataFrame. The Pandas is a foundational and important Python library for data science (McKinney, 2011, McKinney and Team, 2015). The Pandas DataFrame tabulates the data for easy data manipulations. Figure 3 shows an excerpt of the numerical ionogram data for the Ilorin station in Kwara state, Nigeria. Using the generated dates, another DataFrame is created. This newly created DataFrame will be primed to have the completed and cleaned datasets. The procedure to achieve this is by making use of the Pandas' in-built merge method for tables comparisons. Those periods of observations that are found on the extracted DataFrame are included in the date generated Dataframe and a NaN value returned for the rows in which there were no records. As can be seen on the data sample shown in Figure 3, non-numeric columns (e.g., QD) are drops and non-numeric values are replaced with NaN. This way, the whole dataset gets cleaned and the missing days is well represented. The last stage of section 2 is the implementation of the CS data filtration. Based on the user's specified CS value, a mask is created and used to select the data needed and the portion not needed for the analysis is turned to a NaN value. For example, if a user inputted a CS value of 70, this indicates data less than 70 will be returned

as a NaN value on the table.

Section 3 handles the data saving routine of the software. The outputted file is formatted into a comma-separated value (CSV) file and stored with a filename as DINData with a timestamp concatenated to it.

3. Software Packaging and Installation

The various methods and functions mentioned in the previous sections are executed using a graphical user interface (GUI) that allows for a "point-and-click" control of the built software program. The finished codebase is packaged as a single installer file using PyInstaller. The PyInstaller is a Python library that bundles a Python application and all its dependencies into a single package (Cortesi, 2021). Using a single source code, the application can be packaged into different operating systems. Owing to limited resources, the DINData has been distributed in the form of a Windows operating system. The source code will be available for packaging into other different operating systems.

4. Conclusion

In this study, the DINData application software task to read, clean and format DIDBase ionospheric numerical ionogram dataset is presented and has been packaged to run under the Windows operating system. The software was developed using different Python libraries. The DINData source code and the compiled exe file of the application is available for download on <https://github.com/bioyesaeed/dindata.git>. The DINData software, on the other hand, is not dependent on Python for its execution. The software's graphical user interface allows users to enter alphanumerical data, apply CS data filtration and store the outputted data into a comma-separated value (CSV) file.

Acknowledgement

The authors are grateful to the Global Ionospheric Radio Observatory for providing the ionogram dataset. The concept of filtering ionogram datasets using CS values was taught by Galkin I. during one of his presentations at the International Reference Ionosphere 2015 (IRI 2015) Workshop, King Mongkut's Institute of Technology Ladkrabang, Thailand, between 2nd -13th November 2015.

References

<http://www.pyinstaller.org/>

Bello, S., Abdullah, M., Hamid, N. & Yokoyama, T. SEAIONO: A Windows software tool to extract ionogram dataset from South East Asia Low-Latitude Ionospheric Observation Network. *Journal of Physics: Conference Series*, 2019a. IOP Publishing, 012015.

Bello, S., Abdullah, M., Hamid, N. A., Reinisch, B., Yoshikawa, A. & Fujimoto, A. 2019b. Response of ionospheric profile parameters to equatorial electrojet over Peruvian station. *Earth and Space Science*, 6, 617-628.

Bello, S. A., Abdullah, M. & Hamid, N. S. A. 2017a. Investigation of ionospheric minimum frequency near dip equator. *Advanced Science Letters*, 23, 1329-1332.

Bello, S. A., Abdullah, M., Hamid, N. S. A. & Reinisch, B. W. 2017b. Comparison of ionospheric profile parameters with IRI-2012 model over Jicamarca. *Journal of Physics: Conf. Series*, 852.

Bello, S. A., Abdullah, M., Hamid, N. S. A., Yoshikawa, A. & Olawepo, A. 2017c. Variations of B0 and B1 with the solar quiet Sq-current system and comparison with IRI-2012 model at Ilorin. *Advances in Space Research*, 60, 307–316.

Breit, G. & Tuve, M. A. 1925. A radio method of estimating the height of the conducting layer. *Nature*, 116, 357-357.

Cortesi, D. 2021. *PyInstaller Manual* [Online]. Available: [Accessed 21 December 2021 2021].

Galkin, I., Khmyrov, G., Kozlov, A., Reinisch, B., Huang, X. & Kitrosser, D. 2006. Ionosonde networking, databasing, and Web serving. *Radio science*, 41.

Galkin, I., Reinisch, B. W., Huang, X. & Khmyrov, G. M. 2013. Confidence score of ARTIST-5 ionogram autoscaling. *Ionosonde Network Advisory Group (INAG) Bulletin No*, 73, 1-7.

Galkin, I., Vesnin, A., Huang, X., Kozlov, A., Song, P. & Reinisch, B. 2015. GAMBIT Database and Explorer for Real-Time IRI Maps of F2 Layer Peak Height and Density.

Galkin, I. A., Khmyrov, G. M., Reinisch, B. W. & McElroy, J. The SAOXML 5: New format for ionogram-derived data. AIP Conference Proceedings, 2008. American Institute of Physics, 160-166.

Galkin, I. A., Reinisch, B. W. & Bilitza, D. 2018. Realistic Ionosphere: real-time ionosonde service for ISWI. *SunGe*, 13, 173-178.

Huang, X. & Reinisch, B. W. 1996. Vertical electron density profiles from the Digisonde network. *Advances in Space Research*, 18, 121-129.

Huang, X. & Reinisch, B. W. 2001. Vertical electron content from ionograms in real time. *Radio Science*, 36, 335-342.

Khmyrov, G. M., Galkin, I. A., Kozlov, A. V., Reinisch, B. W., McElroy, J. & Dozois, C. Exploring digisonde ionogram data with SAO-X and DIDBase. AIP Conference Proceedings, 2008. American Institute of Physics, 175-185.

Kozlov, A. V. & Paznukhov, V. V. Digisonde drift analysis software. AIP Conference Proceedings, 2008. American Institute of Physics, 167-174.

McKinney, W. 2011. pandas: a foundational Python library for data analysis and statistics. *Python for high performance and scientific computing*, 14, 1-9.

McKinney, W. & Team, P. 2015. Pandas-Powerful python data analysis toolkit. *Pandas—Powerful Python Data Anal Toolkit*, 1625.

Nozaki, K. 2009. FMCW Ionosonde for the SEALION Project. *Journal of the National Institute of Information and Communications Technology*, 56, 287-298.

Paznukhov, V., Galushko, V. & Reinisch, B. 2012. Digisonde observations of TIDs with frequency and angular sounding technique. *Advances in Space Research*, 49, 700-710.

Phillips, D. 2014. *Creating Apps in Kivy: Mobile with Python*, " O'Reilly Media, Inc."

Reinisch, B., Galkin, I., Belehaki, A., Paznukhov, V., Huang, X., Altadill, D., Buresova, D., Mielich, J., Verhulst, T. & Stankov, S. 2018. Pi-

lot ionosonde network for identification of traveling ionospheric disturbances. *Radio Science*, 53, 365-378.Reinisch, B., Huang, X., Galkin, I., Paznukhov, V. & Kozlov, A. 2005. Recent advances in real-time analysis of ionograms and ionospheric drift measurements with digisondes. *Journal of Atmospheric and Solar-Terrestrial Physics*, 67, 1054-1062.Reinisch, B. W. & Galkin, I. A. 2011. Global ionospheric radio observatory (GIRO). *Earth, planets and space*, 63, 377-381.Reinisch, B. W. & Huang, X. 1983. Automatic calculation of electron density profiles from digital ionograms: 3. Processing of bottomside ionograms. *Radio Science*, 18, 477-492.Schunk, R. & Nagy, A. 2009. *Ionospheres: physics, plasma physics, and chemistry*, Cambridge university press.Ulloa, R. 2013. *Kivy: Interactive Applications in Python*, Packt Publishing Ltd.Wolf, M. & Wicksteed, C. 1998. Date and time formats. *W3C NOTE-datetime-19980827*, August.

Figure Caption

Figure 1. DINData software program flow process.

Figure 2. Screenshot of the DINDATA software

Figure 3. Excerpt of the ionogram numerical characteristics data from DID-Base