

Visualizing the Quality of GNSS Multivariate Data

Bruno César Vani¹, Ivana Ivánová², João Francisco Galera Monico², Milton Hirokazu Shimabukuro³

¹Programa de Pós-Graduação em Ciências Cartográficas – Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista (FCT/UNESP) – Rua Roberto Simonsen, 305 – 19060-900 – P. Prudente – SP – Brazil

²Departamento de Cartografia – FCT/UNESP

³Departamento de Matemática e Computação – FCT/UNESP

brunovani22@gmail.com, {i.ivanova, galera, miltonhs}@fct.unesp.br

Abstract. *This paper presents aspects related to the definition, management, and communication of data quality with application to a multivariate data set from Global Navigation Satellite System (GNSS). We stated about the quality in a context of scientific research with GNSS data presenting quality elements that can be applied to scientific research. A case study focusing on time varying quality elements is presented, where interactive data quality visualization schemes are applied to visualize positional accuracy and attribute accuracy varying on time, allowing the users to perform queries. We highlight the main features of the implemented visualizations and conclude with suggestions for future applications.*

1. Introduction

Global Navigation Satellite System (GNSS) allows one to obtain positioning through reception of signals from the satellites of the corresponding system, like GPS, GLONASS and Galileo. The positioning can be achieved with just one system or in combination of more. Currently, GNSS applications extend far beyond positioning and navigation, since observations from GNSS may support research on Aeronomy, Meteorology, and other areas.

Even restricting for the main target of GNSS – the positioning – various methods to provide the position can be applied. Therefore, starting with data acquisition, continuing with data processing and obtaining the resulting position, distinct configurations can be established. These configurations can include, for example, the models to correct the effects that distort the GNSS signals and methods for adjusting the observations. Similarly, other research areas that apply GNSS data can have different configuration methods to reach the desired results. In this context, thinking about the quality of the obtained results may include several ramifications due to the various possibilities available to apply GNSS data.

Harding (2013) presents a central question in the context of data quality: how fit are the data for specific use and user? To answer this question, users need information about the datasets quality. Since it is provided, they can judge if the data is relevant for a specific use context.

In this paper we present about providing and communicating data quality information about GNSS multivariate data set in a context of scientific applications. The discussion covers data quality applied to researchers of GNSS scientific applications, in contrast of the general public who use GNSS positioning technology to navigate. The objective of this paper is to emphasize the importance of the data quality information with a practical case study. We selected a delimited context that covers a GNSS application of standard GPS positioning and also a related application concerning ionosphere monitoring, both being supported by a network of GNSS receivers. We also introduce aspects about the management of time varying data quality information, as well as data quality visualization and interactive manipulation schemes as efficient ways for communicating quality information to the users.

2. GNSS Applications

Nowadays, GNSS technology is ubiquitous. Receivers providing position in real-time and supporting navigation are embedded on cars, watches, tablets and other mobile devices. GNSS technology is also applied in several other professional activities. For example, one can mention the precision agriculture, where high accuracy GNSS positioning is applied allowing the farmers to reach the better application of resources in field plants, and consequently fetching financial advantages. Many other examples could be mentioned, such as Georeferencing, Meteorology, Reflectometry and others, having all of them benefits from the GNSS technology.

In addition of applications that use the GNSS positioning technology, GNSS is also important for scientists. Research related to GNSS technology may follow several ramifications. One example is the research related to improve the positioning accuracy provided by GNSS technology, as proposed for instance in Luo et al. (2008) and Aquino et al. (2011). Furthermore, as GNSS signals propagate by the Earth's atmosphere, the propagation properties of the signal allow investigations on other research areas, i.e., GNSS signals are used not only to obtain positioning. For example, GNSS measurements can support Meteorology on weather forecasting (Sapucci et al. 2009), Aeronomy applications, such as ionospheric studies (Muella et al. 2011), and Reflectometry for near surface applications (Nievinski & Larson, 2014). All of these research use GNSS data. Therefore, providing data quality information is an essential aspect since it may allow different researchers to choose data of desired quality then ensuring the reliability of the derived results according to the requirements of the different research areas.

2.1. Scientific GNSS Applications

From a technical point of view, GNSS are constituted of satellite constellations that allow one to get position based on range measurements between the antennas of receiver and satellites. Those range measurements are based on the propagation time of the electromagnetic signal from the orbital satellite to the ground receiver, and are well known on the literature as the GNSS observables (Seeber, 2003).

Apart from GNSS observables that provide positioning, some special receivers or add-ons can also provide different parameters like climatological indices and signal

processing metrics. Similarly, one can use the combination of these GNSS based parameters in conjunction with external information in order to produce other GNSS derived measurements. For this reason, in this paper, we denote GNSS multivariate data to refer to any kind of data that can be obtained, calculated or derived from GNSS measurements, then including the GNSS observables that provide positioning and all derived metrics or computed values.

General workflow presented in Figure 1 denotes data usage in scientific GNSS applications. The workflow states the data acquisition and the achievement of the results being linked by the data processing step. It is necessary to clarify that distinct workflows could be established covering different applications or points of view.

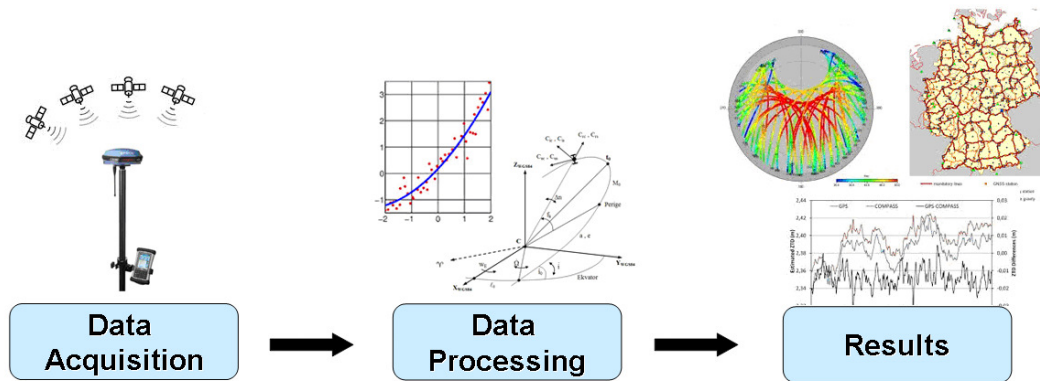


Figure 1. Scientific GNSS application workflow.

On the data acquisition, in spite of user's applications - where the own user's receiver collects the data - many research on GNSS are based on data collected by third part agencies/networks. One example is the network managed by the International GNSS Service (IGS), which is composed by receivers spread across the whole Earth surface. The receivers track continuously signals from the GNSS systems; these data are stored, constituting therefore a useful data repository that can be used on several research related to GNSS. In order to interchange data between different receivers, a common data format has been established since 1989: the Receiver Independent Exchange Format - RINEX (Gurtner et al. 2007). The data tracked by any station belonging to the IGS network can be downloaded in RINEX format by interested users. Other standard file formats are applied to GNSS product - like IONEX, SINEX and SP3 -, as well as standards for transmission of GNSS data in different transmission medium. These file formats specifications can be found at the IGS website (<http://igsceb.jpl.nasa.gov/igsceb/data/format/>).

Data processing is carried out after the data acquisition (see Figure 1). The data processing configuration may differ depending on the methods of GNSS positioning used. For example, one can mention the Standard Positioning Service (SPS) from GPS observables. SPS assumes the basic usage of GPS, which means the usage of a single receiver to get absolute position with accuracy from about 5 to 15m. Another example is given by the Precise Point Positioning (PPP), which allows obtaining centimeter accuracy by applying corrections for all systematic errors that affect the GNSS observables together with precise orbits and clocks. There are also other methods that

are based on differential corrections and relative positioning provided by base receivers in order to be as much accurate as possible (Seeber 2003). For the first case we have the differential positioning with metric accuracy while in the relative positioning using carrier phase, the accuracy is in the cm level.

Finally, the result of data processing is the position or the other desired output, like an ionospheric index (see Figure 1). These results are achieved from different mathematical models those can be applied during data processing, allowing one to obtain certain result.

3. Data Quality Review

Data quality is defined as “totality of characteristics of a data that bear on its ability to satisfy stated and implied needs” (ISO 19101:2002). The data quality information allows users to assess the fitness of the data for a specific purpose. In addition, it is possible to estimate the quality of derived products from a dataset (Goodchild & Clarke 2002). The aspect of allowing users to assess the fitness of the data for a specific purpose can be applicable in research using GNSS data by properly using data quality evaluation metrics.

GNSS data contains uncertainty, which arises from measurement errors, observation operator errors, misleading computation, data corruptions, etc. Consequently, processing uncertain data will propagate the uncertainty to subsequent levels (Williams et al. 2009). Therefore, the concept of data quality can also refer to the degree of uncertainty of the data (Zaixian et al. 2006).

We represent uncertainty in positions achieved with GNSS technology with positional accuracy, which is defined by ISO as “the degree of accuracy of positions within a spatial reference system” (ISO, 2013). If the ground truth of an expected position is available, the positional accuracy can be expressed with classical statistical measures, such as the root mean square error (RMSE). Another indicator is the absolute positional error for a given position, that can be estimated by comparing the surveyed coordinates (for instance X, Y and Z) against the known coordinates (X', Y' and Z'), where one can obtain the resulting tri-dimensional error as follows:

$$Error_{3D} = \sqrt{(X'-X)^2 + (Y'-Y)^2 + (Z'-Z)^2} \quad (1)$$

Considering the GNSS derived metrics - such as ionosphere monitoring indices based on GNSS observations - for the cases where the expected value for the attribute is available, the RMSE can also be applied, now referring to the quality element of attribute accuracy, i.e., how correct these attribute values are (Hacklay, 2010).

When the expected value for the attribute is not available, the classical statistical measurement of the Standard Deviation (SD) can be used to get an indicator of the precision of the variables. Considering a GNSS derived metric x been computed n times, with an average value of \bar{x} , the SD can be computed as follows:

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (2)$$

Other data quality element applicable to our context is the completeness. With completeness, we measure the absence of data, i.e., how many data is missing (ISO 19157:2013). Considering that GNSS data is tracked by networks over time, there is a challenge to manage data quality elements (positional accuracy, attribute accuracy and completeness), because they are also time varying.

Other elements of quality described and discussed in (ISO 19115:2003, ISO 19157:2013, Hacklay 2010, Harding 2013) are also applicable for multivariate GNSS. These elements may cover possible requirements of researchers of scientific GNSS applications using data provided by GNSS networks. We denote these elements as descriptive (containing plain text information) data quality elements :

- Provenance: describes how the GNSS data is created containing description of the GNSS receiver use for data acquisition and the computations performed in the case of GNSS derived parameters;
- Temporal validity: the time or the time range to which the measurement corresponds;
- Resolution: temporal and spatial resolution of the measurement;
- Availability and access: describe how one can find and access the data;
- Usage: describe the basic usage of the data.

3.1. Data Quality Visualization

Data visualization provides effective and attractive means to process information in massive datasets, as well as it provides communicate data quality to the users (Goodchild & Clarke 2002. Zaixian et al. (2006) highlights the importance of designing visualizations that cover not only the data, but also its quality. In that case, appropriate data quality visualization schemes can allow one to get more insight to the data and its quality in a straightforward way. Providing data quality information visualization schemes can be reached by managing data quality as a new attribute on the visualization process or trying to superimpose quality information on the current visualizations.

One example of managing quality elements as new attributes on visualizations is available at the SIRGAS (Sistema de Referencia Geocentrico para as Americas) website (available at <http://www.sirgas.org/index.php?id=206>). A map identifies the stations position on the territory, as illustrated on Figure 2 a). By selecting one station, the positional accuracy information over time is presented on a triad of scatter plots - Figure 2 b). Users can interactively select the stations to check its coordinates' behavior along the time.

4. Architecture and Technological Components of the System for GNSS Data Quality Visualization

The objective in our case study is to manage the quality of the results derived from GNSS data processing. In this section, we present the architecture of a system for visualizing the quality of GNSS data.

The main components of the system are presented on Figure 3. The GNSS constellation is a set of several GNSS systems - GPS, GLONASS and Galileo - with satellites broadcasting signals that propagate through the Earth atmosphere. These signals are tracked by GNSS receivers on the ground segment and are organized in

networks administered by several institutions. Receivers transform signals into data that can be stored in their own format; those receivers for ionosphere monitoring also calculate related parameters. Data are transmitted using FTP protocol to the central data repository where they are written in standard format files before being stored in a database. Using Data Base Management System (DBMS) to organize data makes easier the access by end user. This information is managed by the implemented system and displayed through a web interface which implements dynamic and interactive visualization schemes. We apply quality evaluation metrics presented in section 3 to provide the data quality information to the users. This allows users searching for data of desired quality and selecting it for further use in their own applications. More details are presented with the case study in Section 5.

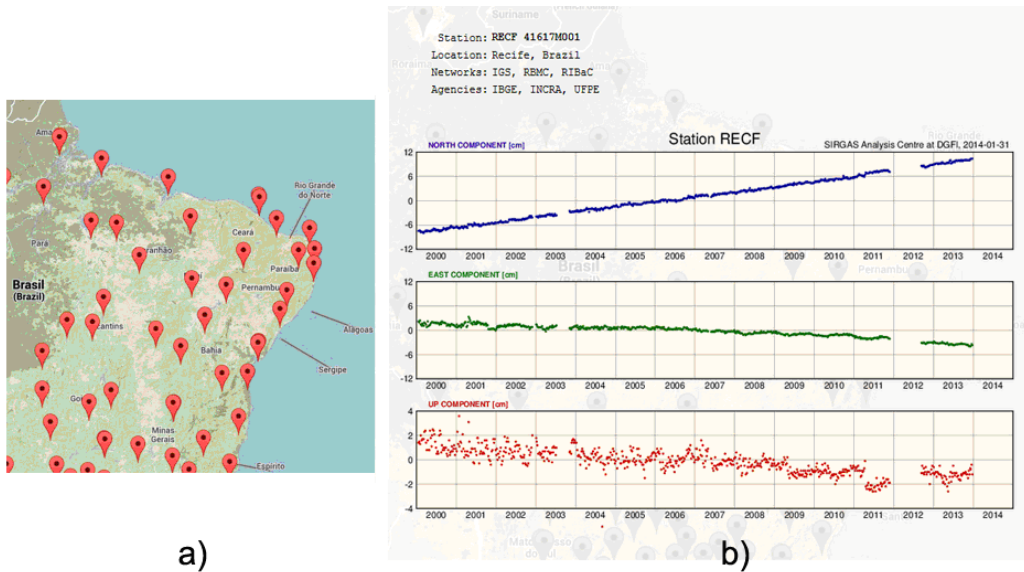


Figure 2. Example of displaying positional accuracy information over time given at the SIRGAS website, available at <http://www.sirgas.org/index.php?id=206>. The map identifies the stations position on the territory - a). By selecting one station, positional accuracy information over time is presented on a triad of scatter plots - b).

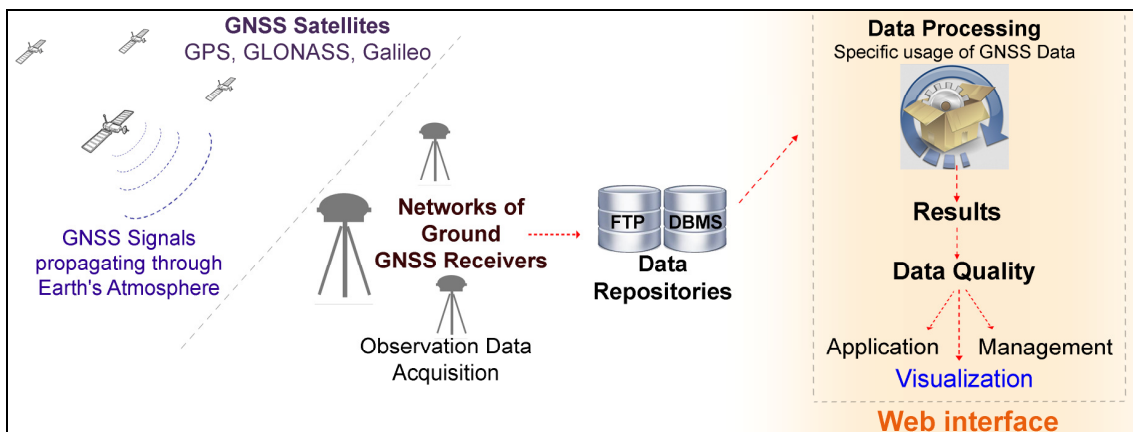


Figure 3. Organization of software and hardware components and data flowing through them.

5. Case Study

The objective of our case study is presenting the definition, management and communication of data quality information on scientific GNSS applications. We focus on the communication of positional accuracy and attribute accuracy in a time varying context. The results presented in this section were implemented within the software ISMR Query Tool – a web application for visualization and mining of GNSS derived data with emphasis on ionosphere monitoring data. The acronym ISMR denotes “Ionospheric Scintillation Monitor Receiver”, a special type of GNSS receiver that provides ionosphere monitoring parameters. The system is accessible at <http://is-cigala-calibra.fct.unesp.br/>.

5.1. Representing uncertainty in positions achieved from GNSS

The subject of our case study is the GNSS data available from the Brazilian Network of Continuous Monitoring of GNSS Systems (RBMC – acronym in Portuguese). This network is managed by the Brazilian national geographic and statistics institute (Instituto Brasileiro de Geografia e Estatística - IBGE). The stations’ configuration of RBMC – that provide RINEX data in real time or post-processed – is presented on Figure 4.



Figure 4. Stations configuration of RBMC. Blue points: real time stations; black points: post-processed stations. Image from IBGE website, available at <http://www.ibge.gov.br/home/geociencias/geodesia/rbmc/rbmc_est.php>.

In our case study, RINEX data for each RBMC station at a given day were processed through the PPS on-line software, which performs Standard Point Positioning with GPS data. The results achieved by the software are the position (X, Y and Z) and associated precision (σ_x , σ_y and σ_z), acquired in each minute of the day. To obtain a

manageable way to manipulate the results, the positioning solutions provided by the software were inserted in a database managed with PostgreSQL DBMS.

The positional error at a given epoch was estimated by comparing the processed output coordinates against the known coordinates (X', Y' and Z') of the stations of the network, obtaining the resulting 3-dimensional error as presented by Eq. 1. One can also obtain the 3-dimensional standard deviation – another quality indicator of the positioning achievement – as follows:

$$SD_{3D} = \sqrt{(\sigma_x)^2 + (\sigma_y)^2 + (\sigma_z)^2} \quad (3)$$

The interactive visualization scheme at a selected epoch is presented on a map as shown in the example of Figure 5. The user can choose between the positional accuracy indicators available ($Error_{3D}$ or SD_{3D}). Red circles are drawn over the station coordinates where the radius' size of each circle is mapped according to the value of the chosen indicator. In order to assure a user-controlled visualization scheme, the user can specify manually the scale range of the chosen parameter as well as the expected range of the representative circles. The circles are drawn in a linear scale according to the specified range values.

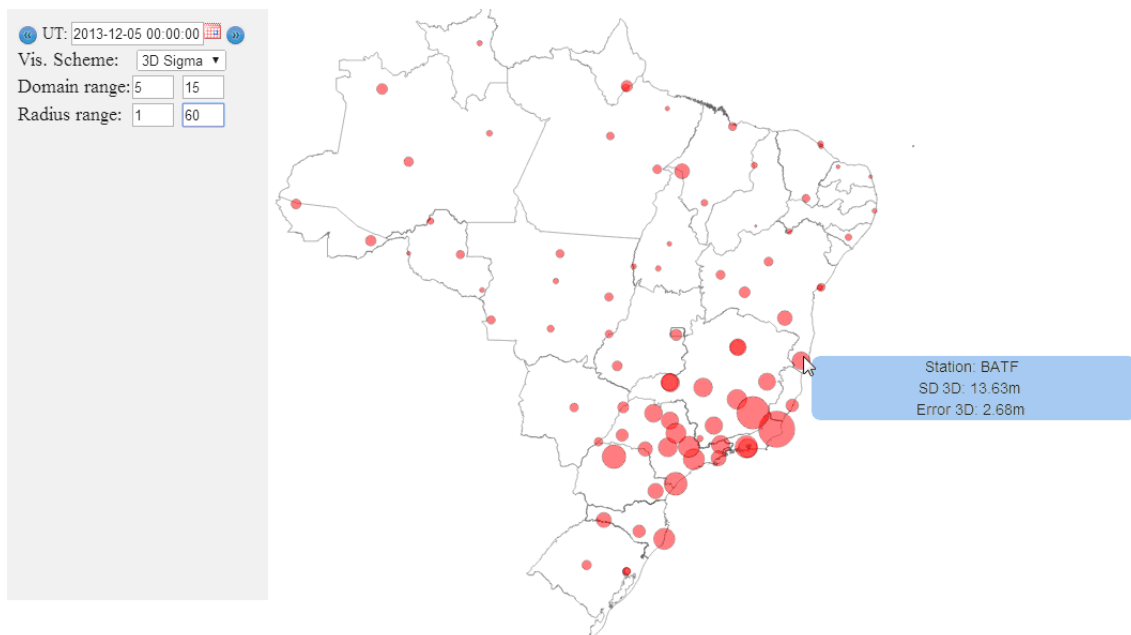


Figure 5. Interactive visualization scheme for Sigma 3D for the positioning solutions for 2013-13-05, 0h (UT).

The interactive visualization can be zoomed in our out, the scales can be changed, and interactions with the mouse give more details about a selected station.

Next to the representation of the results achieved at a single epoch, another possibility is to represent the aggregated positional accuracy information, such as average or standard deviation over a specific time interval. One example is shown in Figure 6, where the average of values of 3D error for the whole selected day is represented. The domain range of the values of 3D error was adjusted from 5 to 10 meters, i.e., the user is expecting that the 3D error has amplitude from 5 to 10 meters.

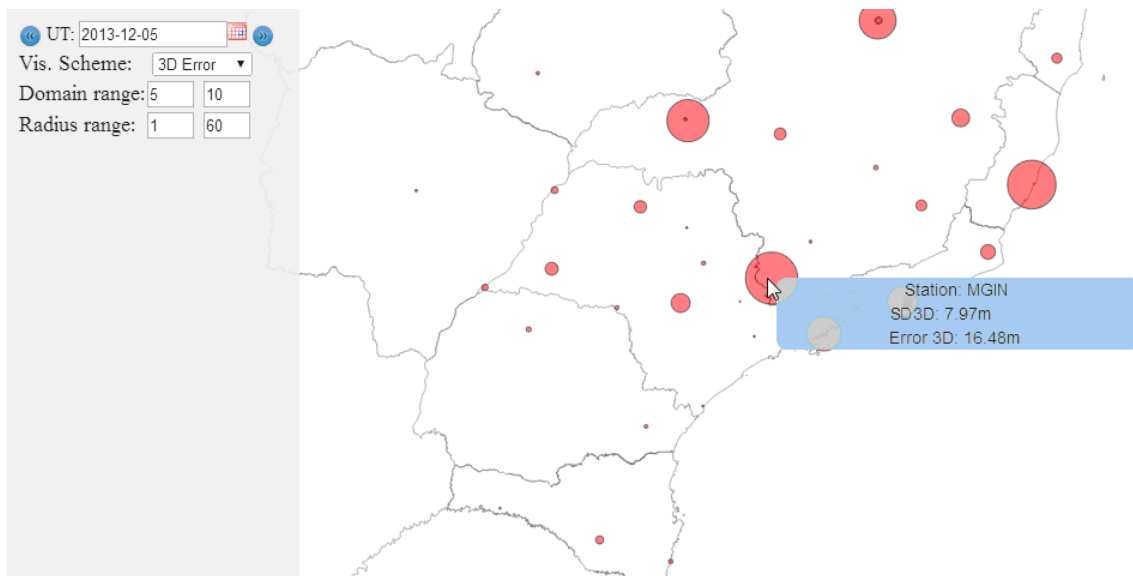


Figure 6. Interactive visualization scheme for averaged Error 3D for all the positioning solutions of 2013-13-05.

5.2. Representing uncertainty in attribute derived from GNSS data

We choose the CIGALA/CALIBRA network's dataset to provide an example for definition, management and display of uncertainty in attributes. This network – that is managed by FCT/UNESP in conjunction with many partners – is currently composed by eleven receivers spread across the Brazilian territory (see Figure 7).



Figure 7. Distribution of stations of CIGALA/CALIBRA Network.

The receivers in CIGALA/CALIBRA network provide derived metrics based on GNSS observations. An example of such observation is the ISMR data, which consists of a subset of several parameters that can be extracted every minute. Moreover, besides the GNSS observables available via the RINEX format, the CIGALA/CALIBRA

Network provide ISMR files, and both can be used by researchers of GNSS scientific applications.

ISMR file contains indices that describe ionospheric activity, as well as additional signal properties metrics those could characterize other effects, like noisy or reflections on the GNSS signals. The ISMR data provided by CIGALA/CALIBRA network support several researches in monitoring and modeling the South American ionosphere.

When dealing with such kind of ionospheric monitoring data, one can use the Ionospheric Pierce Point (IPP) representation in order to visualize the projection of the monitoring parameter over a map. The IPP is a virtual point where the signal passes through a layer of about 350 Km above the Earth surface. The IPP scheme is presented in Figure 8.

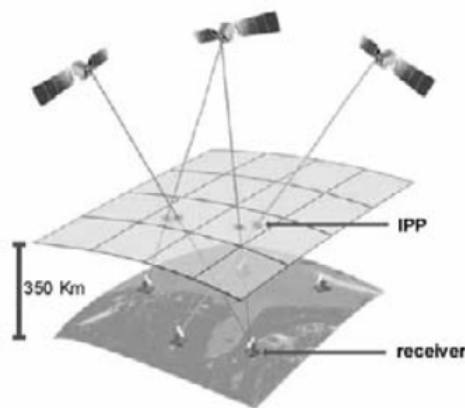


Figure 8. IPP representation. Image from Rezende et al. (2007).

Differently of the example presented on subsection 4.1, in this example we do not store the quality information in a database. Instead, the uncertainty is represented as an attribute (like an average or standard deviation) that can be computed dynamically and rendered with a map server engine. Therefore the result is shown over the Brazilian map with customizable grid resolution (with regular patterns in time and space) and output parameters.

One example is presented in Figure 9. The standard deviation of one selected attribute of ionosphere monitoring (the S4 scintillation index achieved from GNSS observations achieved by receivers of CIGALA/CALIBRA network) is represented for a period of one month (January 2014) with the IPP map, where users can identify an area of more frequently varying quality. On the example presented in Figure 9, values of higher standard deviation are mainly on the bottom of the station. Furthermore, one can also choose different time intervals to get the desired feature on the time varying domain.

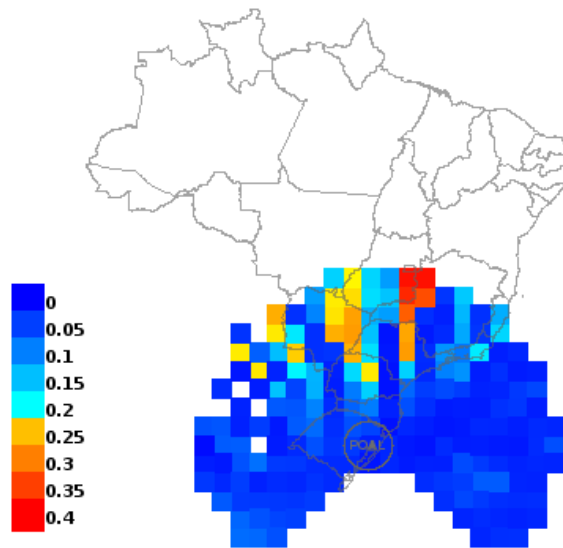


Figure 9. Example of attribute uncertainty visualization with IPP map: standard deviation of S4 ionosphere monitoring index on January, 2014 (data collected at the station POAL - Porto Alegre/RS, Brazil).

6. Conclusions and Future Work

In this paper, we discussed a system for visualizing quality information in the context of scientific applications with GNSS multivariate data. We assumed a general application workflow of data acquisition, data processing and results achievement, and defined quality elements applicable in this scenario.

We presented a system for management and visualization of positional accuracy and attribute accuracy of GNSS data for scientific applications. and demonstrated this system with a small case study. With our system researchers from different areas can visualize quality GNSS data with interactive data visualization schemes.

Our future work includes definitions and implementation of other relevant data quality parameters and improvements related to the usability of the quality visualization schemes. More possibilities to perform queries on data quality information will be explored, for instance by introducing uncertainty information into the dataset allows performing queries and find data of desired quality. In addition, the UncertML (Williams et al. 2009) concept could be applied in order to ensure a proper communication of uncertainty that could allow interoperability between systems.

References

- Aquino, M., Monico, J. F. G., Dodson, A. H., Marques, H., De Franceschi, G., Alfonsi, L., Romano, V. & Andreotti, M. (2009), 'Improving the GNSS positioning stochastic model in the presence of ionospheric scintillation'. *Journal of Geodesy*, 83(10), 953-966.
- Goodchild, M. F. & Clarke, K. C. (2002). 'Data quality in massive data sets'. In *Handbook of massive data sets* (pp. 643-659), Springer US.

- Gurtner, W. & Estery, L. (2007), 'RINEX-The Receiver Independent Exchange Format-Version 3.00', Astronomical Institute, University of Bern and UNAVCO, Boulder, Colorado.
- Harding, J. L. (2013), 'Data Quality in the Integration and Analysis of Data from Multiple Sources: Some Research Challenges', *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(1), 59-63.
- ISO, 2001, 'ISO19101:2001 Geographic Information – Reference model', International Standards Organisation.
- ISO, 2003, 'ISO19115:2003 Geographic information – Metadata', International Standards Organisation.
- ISO, 2013, 'ISO19157:2013 Geographic information – Data Quality', International Standards Organisation.
- Luo, X., Mayer, M., & Heck, B. (2008), 'Improving the stochastic model of GNSS observations by means of SNR-based weighting', *Observing our Changing Earth* (pp. 725-734), Springer Berlin Heidelberg.
- Muella, M. T. A. H., de Paula, E. R., Mitchell, C. N., Kintner, P. M., Paes, R. R. & Batista, I. S. (2011), 'Tomographic imaging of the equatorial and low-latitude ionosphere over central-eastern Brazil', *Earth Planets and Space* 63(2), 129.
- Nievinski, F. G., & Larson, K. M. (2014). Forward modeling of GPS multipath for near-surface reflectometry and positioning applications. 'GPS solutions', 18(2), 309-322.
- Rezende, L. F. C., Paula, E. R., Kantor, I. J. & Kintner, P. M. (2007), 'Mapping and survey of plasma bubbles over Brazilian territory', *Journal of Navigation* 60(01), 69-81.
- Sapucci, L. F., Machado, L. A. T. & Monico, J. F. G. (2009), 'Previsões do atraso zenital troposférico para a América do Sul: variabilidade sazonal e avaliação da qualidade', *Revista Brasileira de Cartografia*, 58(3).
- Williams, M., Cornford, D., Bastin, L. & Pebesma, E. (2009), 'OGC Discussion Paper 08-122r2: Uncertainty Markup Language (UnCertML)'.
(http://www.opengeospatial.org/standards/uncertml)
- Zaixian, X., Shiping, H., Ward, M. O. & Rundensteiner, E. A. (2006), 'Exploratory visualization of multivariate data with variable quality', in *Visual Analytics Science And Technology*, 2006 IEEE Symposium On (pp. 183-190).