Lima Segantine, Paulo Cesar; da Silva, Irineu; Sarubo Dorth Domingues, Mélodie Kern

PROPOSAL OF A TABLE TO CLASSIFY THE RELIABILITY OF BASELINES
OBTAINED BY GNSS TECHNOLOGY

PROPOSAL OF A TABLE TO CLASSIFY THE RELIABILITY OF BASELINES OBTAINED BY GNSS TECHNOLOGY

Proposta de uma tabela para a classificação da confiabilidade de linhas de base obtidas com a tecnologia GNSS

Paulo Cesar Lima Segantine[1]
Irineu da Silva[1]
Mélodie Kern Sarubo Dorth Domingues[1]

[1] Escola de Engenharia de São Carlos – USP –São Carlos- SP – Brasil. Email: seganta@sc.usp.br; irineu@sc.usp.br

Abstract:

The correct data processing of GNSS measurements, as well as a correct interpretation of the results are fundamental factors for analysis of quality of land surveying works. In that sense, it is important to keep in mind that, although, the statistical data provided by the majority of commercials software used for GNSS data processing, describes the credibility of the work, they do not have consistent information about the reliability of the processed coordinates. Based on that assumption, this paper proposes a classification table to classify the reliability of baselines obtained through GNSS data processing. As data input, the GNSS measurements were performed during the years 2006 and 2008, considering different seasons of the year, geometric configurations of RBMC stations and baseline lengths. As demonstrated in this paper, parameters as baseline length, ambiguity solution, PDOP value and the precision of horizontal and vertical values of coordinates can be used as reliability parameters. The proposed classification guarantees the requirements of the Brazilian Law N° 10.267/2001 of the National Institute of Colonization and Agrarian Reform (INCRA).

Keywords*:* GNSS reliability, Decision tree, Adjustment of observations.

Resumo:

O correto processamento das medições GNSS, bem como a correta interpretação dos seus resultados, são fatores fundamentais para a análise da qualidade dos trabalhos de medições agrárias. Neste sentido, é importante ter em mente que, embora, os valores estatísticos indicados pela maioria dos programas de computador comerciais de processamento GNSS descrevam a credibilidade do trabalho realizado, eles não são consistentes para indicarem a confiabilidade dos valores das coordenadas processadas. Com base nessa suposição, este artigo propõe uma tabela para classificar a confiabilidade de linhas de base obtidas por meio de processamento de dados GNSS. Como dados de entrada foram utilizadas medições realizadas entre os anos 2006 e 2008,

considerando diferentes estações do ano, diferentes configurações das estações RBMC e diferentes comprimentos de linhas de base. Como demonstrado, verifica-se que parâmetros, tais como, comprimentos da linha de base, solução da ambiguidade, valores PDOP e valores da precisão das coordenadas obtidas, podem ser usados como parâmetros de confiabilidade. A classificação proposta garante os requerimentos da Lei N° 10.267/2001 do Instituto Nacional de Colonização e Reforma Agrária (INCRA).

Palavras-chave: Confiabilidade GNSS, Árvore de decisão, Ajustamento de observações.

1. Introduction

With the advent of the artificial satellites technology, the activities that involve geodetic measurements have experienced a revolution in recent years, due to the ability that the Global Navigation Satellite System (GNSS) has to produce accurate point positioning, combining promptness and accuracy. These facts have motivated the necessity of revision of the characteristics of the Brazilian Geodetic System, resulting in the implementation of the Brazilian Network for Continuous Monitoring - RBMC and therefore the establishment of SIRGAS2000.

Based on this approach, this paper analyzes the coordinates obtained from GNSS data processing, as well as the related statistical values from the adjustment procedures applied to obtain the final coordinates to classify the GNSS measurements in terms of reliability. The accuracy values were determined by comparing the differences in values of the computed coordinates and the official ones made available by IBGE. The data processing and adjustments have been conducted based on GNSS measurements collected in some periods of the years 2006, 2007 and 2008, considering the baseline geometries, distance between stations, seasons of the year, occurrence of solar activities and the use of control points (horizontal and vertical). The results were classified into four levels using the Decision Tree technique. As a result, it is proposed a classification table of reliability, so that the information can be used in different land surveying projects. In addition, this paper proposes a methodology to evaluate the quality and accuracies of short and long baselines to comply with the requirements of the Brazilian´s Law N° 10.267/2001 (LEI N° 10.267, 2001), which is related to the rural lands georreferencing. According to this law, the owners of rural areas are required to submit the georreferencing plant of their rural properties to INCRA, attesting that their property does not overlap other rural lands or environmental conservation areas, as for instance, Natives reservations or Maroons.

2. GNSS

The basic principle of GNSS technology is the determination of the travel time of the signal transmitted by artificial satellites, which reach a terrestrial receiver antenna and, consequently, generates the vectorial distance between the receiver antenna and the satellite antennas. Based on baseline computations, it is possible to calculate the coordinates of the receiver antenna in the

same coordinate system of a terrestrial known point, considered as reference in the data processing.

The coordinates obtained from the data collected by GNSS receivers undergo a series of influences, due to the environment in which the electromagnetic waves pass through before reaching the receiver antenna. As a result, the processed coordinates are degraded in precision and accuracy. In this work, these influences are classified into three groups: (1) uncertainties arising in signal emission, (2) uncertainties arising in signal propagation and reception and (3) uncertainties arising in data processing and adjustment.

2.1 Uncertainties arising in signal emission

Uncertainties are related to the geometry of distribution of the satellites on the observer's horizon (geometry) and to the quality of the transmitted ephemerides by the satellites. The geometry can be evaluated by named DOP factors (Dilution of Precision), being the PDOP factor (Precision of DOP) considered in this study.

Regarding the quality of ephemerides, there are two different classes: broadcasted ephemerides by the satellites and precise ephemerides provided by various civil organizations. The broadcasted ephemerides emitted by the GNSS satellites provides, by navigation message, the orbital elements necessary for the coordinates computation of each satellite as well as the coefficients for the satellite clock corrections. The precise ephemerides are derived from calculations performed by research centers, under contracts with the International GNSS Service - IGS, from orbital elements collected. Based on this fact, one can say that the experiments that use precise ephemerides are more accurate than those processed by broadcasted ones. It is recommended that broadcast ephemerides shall be used only for computation of coordinates in short baselines, while precise ephemerides are recommended for long baselines. In this study, it has been used only IGS precise ephemerides downloaded from http://igscb.jpl.nasa.gov/components/prods.html (accessed on 01/31/2011), which are a result of the combination of GNSS data collected from several research centers. These data are normally available on websites, at least a week after data collection.

2.2 Uncertainties arising in signal propagation and reception

This topic is related to all disturbances that act on the wave front during its propagation from the satellite antennas to the receiver antenna.

Tropospheric refraction

Troposphere is a layer of the atmosphere, situated above the earth's surface up to 50 km altitude, formed mostly by the concentration of gases, primarily containing Nitrogen (N2), Oxygen (O2), Dioxide of Carbon (CO2), Argon (Ar), water vapor and neutral particles. The propagated signal is influenced directly by water vapor, atmospheric pressure and temperature that exist in this layer. The effects of the troposphere are considerable in the GNSS positioning, especially in the geometric height determination. Some GNSS software have models for computing such effects, by estimating the tropospheric zenith delay.

Ionospheric refraction

Ionosphere is a layer of the atmosphere that extends in various layers from about 50 km to 1,000 km above earth. These layers are composed of ionized gases that interacts with their constituents elements. The ionization depends on the molecules of the gases, on their physical and on the intensity of radiation. As these radiations are close to the earth´s surface, they find dense layers that increase the density of the free electrons by means of a photochemical process. Simultaneously, there is a process of particles transportation forming a distribution of layers with different densities. Thus, the ionosphere causes a delay in the pseudodistance measurements and an equivalent advance of carrier phase due to the presence of free electrons in this layer. The ionosphere behavior is a function of time of day, time of year, latitude and longitude of the observation points and the solar activities.

Solar activities

The uncertainty due to the ionosphere is directly proportional to the Total Electron Content (TEC). The TEC varies regularly in time and space in relation to the sunspot cycle, epoch of year, time of day, geographic location, among others. However, the TEC can undergo abrupt changes in their behavior, due to, for instance, the occurrence of intense solar flares. This occurrence can happen in two ways: (1) unexpectedly, considering that the register are due to the systematic mapping of the solar surface realized by terrestrial observatories, and (2) expectedly, provided by the cycle of solar activity that is monitored by official research centers, for more than a century. The monitoring has found that they present a regular and repetitive cycle, lasting approximately 11 years. See figure 1.
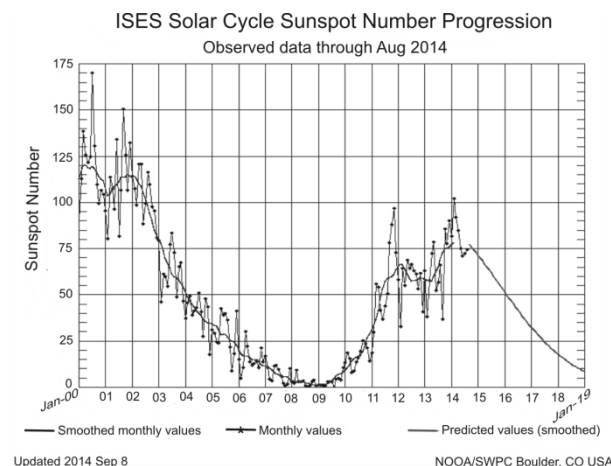

Figure 1: Solar cycle progression from January 2000 to January 2019.

Source: Adapted from www.swpc.noaa.gov/SolarCycle (accessed at 10/10/2014).

Cycle slips

Cycle slips can occur due to existing obstructions near to the receiver antenna, such as buildings, trees, bridges, mountains and others. Furthermore, they can also be a result of abrupt changes in the atmosphere conditions, interference from other radio sources and problems in the software/hardware of the receiver. In case of cycle slips detection, the observations can be corrected, or disregarded during the data processing.

Multipath

Multipath, can be termed as "signals reflection", is a process that occurs when the front of wave that propagates between the satellite antennas and the receiver antenna find a discontinuity in the refractive index. Thus, the receiver antenna can receive, in some circumstances, besides the

signal arriving directly to it from the satellites, signals reflected from surfaces close to it, such as buildings, cars, trees, water masses, fences, etc. The effect of multipath in the code observables is almost double of the carrier phase, and to reduce its effects, the users must take care of the places where the receiver antennas are installed.

Ambiguity

The GNSS receivers measures the fractional part of the phase of the carrier wave starting the counting of cycles as soon as the receiver is turned on. The ambiguity can be understood as the number of cycles since of the first observable, and is a parameter determined by adjustment, together with other parameters such as, the clock errors. It is inherent in the carrier phase measurement, and depends on the quality of the receiver, receiver antenna and the satellite signal. It can be said that in all data collection for a long period, there is always occurrence of cycle slips. The estimated values of the ambiguity can be real numbers (float solution or free solution) or can be set as integer numbers (fix solution). However, the integer ambiguities is not a simple task, and the solution can be affected by ionospheric effects, tropospheric refraction, multipath, geometry and number of satellites available on the horizon of the receiver antenna, as well as the collection time of observables.

2.3 Uncertainties originated in the data processing and adjustment

The factors that influence the accuracy of coordinate of a point considered in this paper are: (a) distance and geometry among the RBMC stations, (b) quality of the data and (c) horizontal and vertical accuracies obtained from the adjustment of the processing data.

(a) Distance and geometry between RBMC stations

In this study all analysis are based on baselines from 300 km up to 1,400 km length. Distances shorter than 300 km were considered short baselines due to the high quality of the RBMC data. Regarding the geometry of the RBMC stations, it have been used two different configurations: one using equivalent distances (approximately equal) and another one with different distances.

(b) Adjustment of observations by the method of least squares and quality analysis using the chi-square test ($\chi2$)

Adjustment of observations is a mathematical model for the solution of an overdetermined system of equations based on the principle of least squares (Amiri-Simkooei et all, 2009). Beside the unique solution, the adjustment is also an important tool for statistical analysis and quality of the results including precisions and statistical tests for quality control. In this paper, the quality control of the adjustment have been done using the Chi-Square Test ($\chi2$).

3. Brazilian continuous monitoring network (RBMC)

The RBMC aims to provide an infrastructure of geodetic reference using GNSS technology, facilitating the use of the system by end-users and ensuring the quality of the results. It is worth mentioning that the RBMC is also the main connection with the global reference systems. It is

composed by GNSS reference stations with known coordinates, eliminating the need to immobilize a receiver over a geodetic point during measurement campaigns. Moreover, the RBMC stations are fitted with receivers to high performance, providing high-quality of the observables and consequently providing reliability to the data processing results.

4. Decision tree learning (DTL)

DTL are powerful and popular tools for classification and prediction of cases (Moussas, 2006). The attractiveness of the use of decision trees is due to the fact that, in contrast to neural networks, decision trees represent criteria, that can readily be expressed so that humans can understand them or even directly used in a database access language like SQL so that records falling into a particular category may be retrieved.

In some applications, the accuracy of a classification or prediction is the only thing that matters. In such situations, we do not necessarily care how or why the model works. In other situations, the ability to explain the reason for a decision is crucial. There are varieties of algorithms for building decision trees that share the desirable quality of interpretability.

A decision tree can be used as a model for sequential decision of problems under uncertainties. It describes graphically the decisions that can be made, the events that may occur, and the outcomes associated with combinations of decisions and events. Probabilities are assigned to the events, and values are determined for each outcome. A major goal of the analysis is to determine the best decisions. The idea of a DTL is that the learning preceptors are not only used to act, but also to enhance the ability of the agent to act in the future. Learning takes place to the extent that the agent observes their interactions with the world and its internal decision-making process.

DTL is an example of inductive learning, i.e. through a hypothesis based in a particular case to produce general conclusions. It is very useful to implement expert systems and classify problems. The DTL takes as input data described by a set of attributes and returns a decision, which is the predicted value for the input value. This paper used this technique to create the reliability classification table (see section 7).

5. Material and methods

As already mentioned, the data processing has been performed considering the geometry of the baselines, the distances between RBMC stations, the season of the year, the significant occurrence of solar activities and the use of horizontal and vertical control points. To analyse the influence of the network geometry, the data processing has been performed taking into account two cases: (1) stations with similar distances and (2) stations with different distances.

Regarding the case of good geometry and long baselines, the stations used were CUIB, BRAZ and PPTE. For the case of long distances and bad geometry, the chosen stations were CUIB,

BRAZ and MCLA. In both cases, the considered "unknown" station in the processing was CUIB. Figure 2 illustrates these two situations.
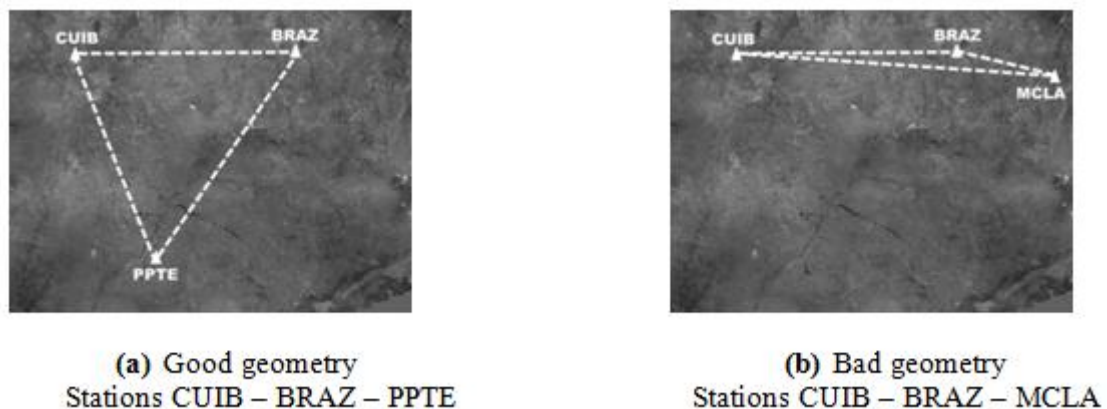

(a) Good geometry
Stations CUIB – BRAZ – PPTE

(b) Bad geometry
Stations CUIB – BRAZ – MCLA

Figure 2: Network geometry for long baselines.
Source: IBGE

The stations used for shorter and similar distances were VARG, VICO and RIOD and for shorter (and different) distances were VARG, RIOD and UBER. The station VARG was considered as "unknown". The choice of this station has been done in order to get the required geometry and due to the greater quantity and quality of data for these stations. Figure 3 shows these two situations:
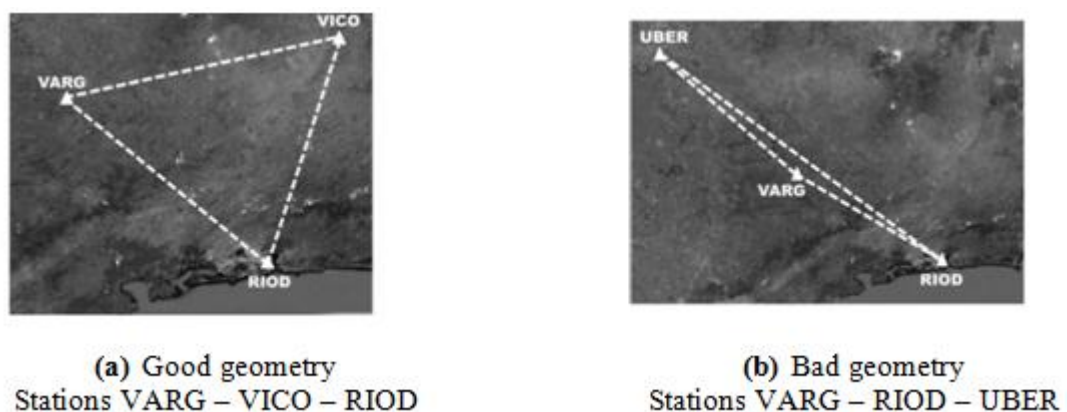

(a) Good geometry
Stations VARG – VICO – RIOD

(b) Bad geometry
Stations VARG – RIOD – UBER

Figure 3: Network geometry for short baselines.
Source: IBGE.

Regarding the period of time on this research, it was considered four days of data referring to four different seasons of the year (approximately 20 days after initiation of season). The days have been selected according to data availability from RBMC in the IBGE site, and the data processing have been carried out using the period of data provided by IBGE for each day (in most cases, these periods were 24 hours). The days selected were (considering south hemisphere):

- FALL (03/21 to 06/21): April, 11 to 14 (Julian days 101 to 104 in the years 2006 and 2007; 102 to 105 in the year 2008 (leap year));

- WINTER (06/21 to 09/23): July, 12 to 15 (Julian days 193 to 196 in the years 2006 and 2007; 194 to 197 in the year 2008 (leap year));

-

- SPRING (09/23 to 12/21): October, 13 to 16 (Julian days 286 to 289 in the years 2006 and 2007; 287 to 290 in the year 2008 (leap year));

- SUMMER (12/21 to 03/21): January, 11 to 14 (Julian days 11 to 14 in the year 2006).

The software used in the data processing was Topcon Tools, version 7.1, considering the SIRGAS2000 parameters and precise ephemeris obtained on the website http://igscb.jpl.nasa.gov/components/prods.html, accessed in 1/31/2011. After the data processing, the results of the processed coordinates have been separated and sorted by DTL in the software WEKA (Waikato Environment Knowledge Analysis), version 3.5, according to the previously established criteria.

For the adjustment of the observations, two known stations have been used for each case: VICO and RIOD for the first, UBER and RIOD for the second, BRAZ and PPTE for the third and BRAZ and MCLA for the fourth. The website used to obtain information regarding solar activities was: www.sec.noaa.gov (accessed 1/31/2011). The ionospheric conditions have been evaluated by the Klobuchar model.

To evaluate the correlation between the precision and accuracy of the coordinates obtained in the data processing with a higher degree of confidence, a relevant fact considered in this paper was the accuracy of the equipment used in the RBMC stations. The devices have different accuracies among themselves, thus for each classification this factor has been taking into account. The related information about the equipment and stations are listed below. See Table 1:

Table 1: List of equipment used in some stations of the RBMC for collecting GNSS data. Source: IBGE.

Station	Equipment	Precision
VARG	Receiver: Astech Z-FX	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $5mm+1ppm$
VICO	Receiver: Trimble 4000 SSI Receiver: Trimble NetR5	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$ Horizontal accuracy for the static method: $3mm+0.1ppm$ Vertical accuracy for the static method: $3.5mm+0.4ppm$
RIOD	Receiver: Trimble 4000 SSI Receiver: Trimble NetRS	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$ Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$
UBER	Receiver: Astech Z-FX	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $5mm+1ppm$
CUIB	Receiver: Trimble 4000 SSI Receiver: Trimble NetRS	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$ Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$
PPTE	Receiver: Trimble NetRS	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$
BRAZ	Receiver: Trimble 4000 SSI Receiver: Trimble NetRS	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$ Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $10mm+1ppm$
MCLA	Receiver: Astech Z-FX	Horizontal accuracy for the static method: $5mm+1ppm$ Vertical accuracy for the static method: $5mm+1ppm;$

6. Processing results

The data processing have been performed using the following parameters:

- Reference System: SIRGAS2000;
- Type of coordinates: latitude, longitude and ellipsoidal height;
- Coefficient of Refraction 0,14;
- Elevation Mask: 15º;
- Adjustment by Least Squares Method;
- Confidence level of Adjustment: 68%;
- Number of set points: 3
- Number of control points: 2;
- Number of vectors: 2
- Rejection Criteria Adjustment: Chi-square (UWE).

For each chosen day of the year, it has been done four distinct data processing, taking into account the distance and the geometry among stations.

All results of coordinates have been transformed to UTM coordinates on different projection zones, depending on the station position. All coordinate values are related to SIRGAS2000.

Case 1: Similar distances among stations considering baselines shorter than 300 km

For this case, the following RBMC stations have been used: VARG, VICO and RIOD. VICO and RIOD stations have been considered as control points and VARG station was the unknown point (see Table 2).

Table 2: UTM coordinates of stations RIOD, VICO and VARG (SIRGAS2000 Reference System), approved by IBGE.
Source: IBGE.

Stations	E	N
RIOD	673.825,217 m	7.475.648,024 m
VICO	721.757,711 m	7.702.785,751 m
VARG	454.968,971 m	7.617.727,294 m

Case 2: Different distances among stations considering at least one baseline shorter than 300 km

For this case, the following RBMC stations have been used: RIOD, UBER and VARG. RIOD and UBER stations have been considered as control points and VARG station was the unknown point, (see Table 3).

Table 3: UTM coordinates of stations RIOD, UBER and VARG (SIRGAS2000 Reference System), approved by IBGE.
Source: IBGE.

Station	E	N
RIOD	673.825,217 m	7.475.648,024 m
UBER	782.656,488 m	7.909.251,415 m
VARG	454.968,971 m	7.617.727,294 m

Case 3: Similar distances among stations and baselines greater than 700 km

For this case, the following RBMC stations have been used: BRAZ, PPTE and CUIB. BRAZ and PPTE stations have been considered as control points and CUIB station was the unknown point (see Table 4).

Table 4: UTM coordinates of stations BRAZ, PPTE and CUIB (SIRGAS2000 Reference System) approved by IBGE.
Source: IBGE.

Station	E	N
BRAZ	191.901,220 m	8.234.747,341 m
PPTE	457.866,057 m	7.553.844,608 m
CUIB	599.737,357 m	8.280.040,831 m

Case 4: Large distances and different baselines length, considering at least one baseline greater than 700 km.

For this case, the following RBMC stations have been used: BRAZ, MCLA and CUIB. BRAZ and MCLA stations have been considered as control points and CUIB station was the unknown point (see Table 5).

Table 5: UTM coordinates of stations BRAZ, MCLA and CUIB (SIRGAS2000 Reference System), approved by IBGE.
Source: IBGE.

Station	E	N
BRAZ	191.901,220 m	8.234.747,341 m
MCLA	619.257,849 m	8.151.040,841 m
CUIB	599.737,357 m	8.280.040,831 m

7. Classification and results analysis

The parameters used to classify the results were:

- When the discrepancy between the coordinates approved by IBGE and the estimated coordinates found in data processing was exceeded in 10 cm, the data would be immediately rejected. This value has been adopted based on the Technical Standard for Georreferencing of Rural Properties - 2ª edition (INCRA, 2010.);

- The solution of ambiguity as "float solution" has been accepted only for baselines greater than 300 km. The classification has been carried out taking into account the following considerations: if only one baseline does not produce the expected ambiguity solution, the process would be of a lower class, but if two or three baselines do not produce the expected ambiguity solution, the process would be rejected. This decision was based on previous research (MENZORI, 2005);

- For the classification "A" the PDOP value should be smaller than 3. If PDOP values were between 3 and 6 the related baseline was automatically labelled as a lower class. With two or three PDOP values between 3 and 6, or with only one value above 6, the data processing was rejected;

- To estimate the limit value of the horizontal and vertical accuracies, it has been taken into account the accuracy of each receiver in each baseline and its time of operation during the data collection.

The results of all data processing have been divided according to the criteria presented in Figure 4.

Parameters	2006	2007	2008
Processing and Adjustment	-	-	-
Discrepancy	-	-	-
Ambiguity	4	-	-
PDOP	2	-	-
Horizontal precision	-	-	-
Vertical precision	-	-	-

Results of VARG-VICO-RIOD configuration.

Parameters	2006	2007	2008
Processing and Adjustment	2	1	-
Discrepancy	-	-	-
Ambiguity	-	-	-
PDOP	2	-	-
Horizontal precision	-	-	-
Vertical precision	-	-	-

Results of VARG-RIOD-UBER configuration.

Parameters	2006	2007	2008
Processing and Adjustment	5	1	-
Discrepancy	-	-	-
Ambiguity	-	-	-
PDOP	2	-	-
Horizontal precision	-	-	-
Vertical precision	-	-	-

Results of CUIB-BRAZ-PPTE configuration.

Parameters	2006	2007	2008
Processing and Adjustment	4	1	1
Discrepancy	-	1	2
Ambiguity	-	-	-
PDOP	12	12	14
Horizontal precision	-	-	-
Vertical precision	-	-	-

Results of CUIB-BRAZ-MCLA configuration.

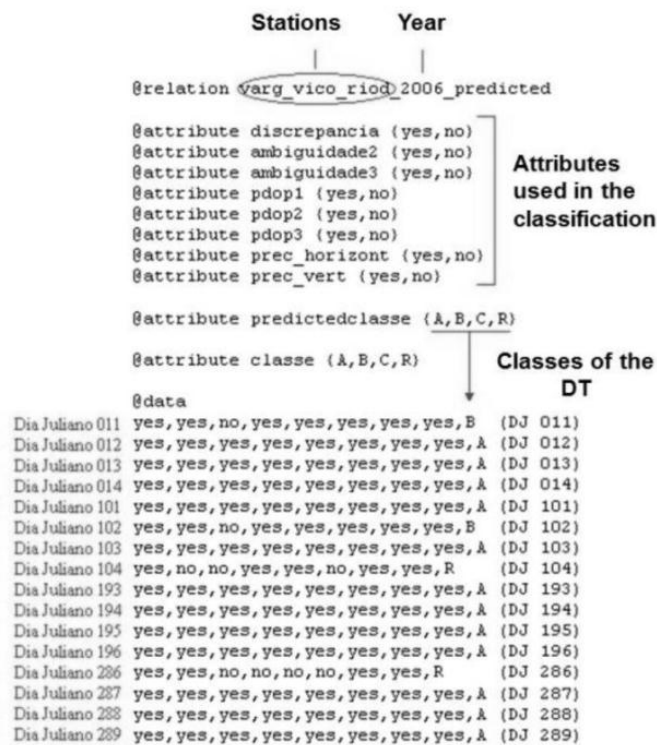**Figure 4:** Results of different configurations.

Table 6 presents the proposed classification, which has been considered in the DTL for the analysis of the adjusted results of the coordinates.

Table 6: Proposal for classification of baselines.

Classes	Conditions
A	All conditions were met satisfactorily according to the criteria established.
B	One of the conditions was not met satisfactorily. For example, the PDOP of a particular station was greater than 3, or the ambiguity of the baseline was not solved as it was expected, etc.
C	Two conditions were not met satisfactorily.
R	Three or more conditions were not met satisfactorily.

Classification by WEKA

Figure 5 shows the database structure generated for classification proposal in this paper, using the WEKA software.


```
                    Stations      Year
                       |            |
        @relation varg_vico_riod 2006_predicted

        @attribute discrepancia (yes,no)
        @attribute ambiguidade2 (yes,no)
        @attribute ambiguidade3 (yes,no)
        @attribute pdop1 (yes,no)           Attributes
        @attribute pdop2 (yes,no)           used in the
        @attribute pdop3 (yes,no)           classification
        @attribute prec_horizont (yes,no)
        @attribute prec_vert (yes,no)

        @attribute predictedclasse (A,B,C,R)

        @attribute classe (A,B,C,R)          Classes of the
                                                   DT
        @data
Dia Juliano 011  yes,yes,no,yes,yes,yes,yes,yes,B   (DJ 011)
Dia Juliano 012  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 012)
Dia Juliano 013  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 013)
Dia Juliano 014  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 014)
Dia Juliano 101  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 101)
Dia Juliano 102  yes,yes,no,yes,yes,yes,yes,yes,B   (DJ 102)
Dia Juliano 103  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 103)
Dia Juliano 104  yes,no,no,yes,yes,no,yes,yes,R     (DJ 104)
Dia Juliano 193  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 193)
Dia Juliano 194  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 194)
Dia Juliano 195  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 195)
Dia Juliano 196  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 196)
Dia Juliano 286  yes,yes,no,no,no,no,yes,yes,R      (DJ 286)
Dia Juliano 287  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 287)
Dia Juliano 288  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 288)
Dia Juliano 289  yes,yes,yes,yes,yes,yes,yes,yes,A  (DJ 289)
```

Figure 5: Example of a structure of database used by WEKA.

In the header of Figure 5 it is shown the stations (VARG, VICO and RIOD) and the processing year, that in this experiment was 2006. The following lines contain the qualifying attributes used for classification. The software interface, as well as the graph generated for this experiment can be visualized in the Figure 6.
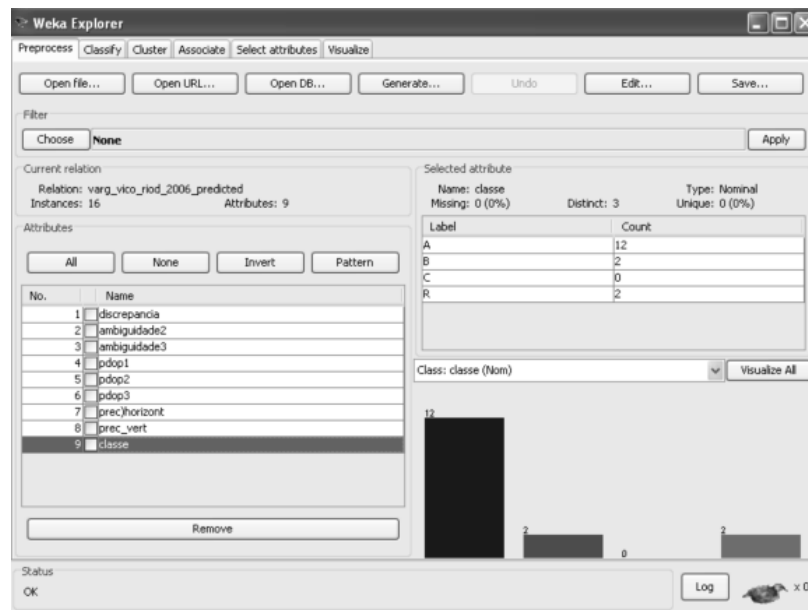
Figure 6: The interface of WEKA and the graphic.

The classification results generated by WEKA software can be viewed in Figure 7 (referring to the four cases described previously). In the figure, the reader can see the amount of data processing classified as A, B, C or R, for each year.

Class	2006	2007	2008
A	12	16	16
B	2	-	-
C	-	-	-
R	2	-	-

Classification of the
VARG-VICO-RIOD configuration

Class	2006	2007	2008
A	12	15	16
B	1	-	-
C	-	-	-
R	3	1	-

Classification of the
VARG-RIOD-UBER configuration

Class	2006	2007	2008
A	9	15	16
B	2	-	-
C	-	-	-
R	5	1	-

Classification of the
CUIB-BRAZ-PPTE configuration

Class	2006	2007	2008
A	-	1	-
B	4	12	14
C	7	-	-
R	5	3	2

Classification of the
CUIB-BRAZ-MCLA configuration

Figure 7: Results of the classification under different configurations.

8. Analysis and results

The results have been classified accordingly to the behavior of the factors generated by the data processing and adjustment, considering the direct influence of solar activities, baseline lengths and geometric configuration between RBMC stations.

The behavior of the results, for the months in which the solar activity was high, was very similar to the behavior of the results for the months with no atypical sunspot activity in all baseline lengths and all geometric configurations of the stations used in this study. For baselines shorter than 100 km, the behavior of the results of both cases is also very similar, according to indications of Menzori (2005).

Regarding the baselines lengths, a large number of data processing for baselines greater than 700 km has been rejected and classified as classes B and C, while in baselines shorter than 300 km, they have been classified as class A.

The solution of the ambiguity behaved as expected, except for some baseline lengths shorter than 300 km, where the solutions of the ambiguity were not fixed. For data processing with baseline lengths up to 800 km, for which it was expected that all solutions would be of floating type, in some cases, the solution was of fixed ambiguity. These cases require further analysis, since the solution may be good from a statistical point of view, however, the coordinates of the points may have significant errors and hence the computed distances of the baselines may have significant errors.

The baseline length does not influence directly the results of the ambiguity solution, but on the other hand, in conjunction with the geometric configuration of the RBMC stations, it can be critical for the accuracies of the horizontal and vertical coordinate values. For long baselines and bad network geometry, the accuracies were within the thresholds set by the equipment manufacturers. The accuracies almost doubled in cases where the baselines had a maximum length of 300 km and good geometric configuration (the distances between stations were similar - equilateral triangle), making it clear that the latter situation allows a better quality of the processed results.

Another relevant factor is related to the PDOP values. For baselines below 300 km, almost all PDOP values were less than 3. For baselines greater than 1,000 km, and considering triangles with different lengths, the vast majority of PDOP values were greater than 3, which degrades the quality of the results.

It is noteworthy that a few numbers of results were rejected due to discrepancies between the coordinates processed and the ones approved by the IBGE. All rejected cases occurred in situations where the distance between the RBMC stations were very different. In situations of similar baselines lengths, even for long baselines, there was no problem regarding the different baseline lengths. Therefore, it is apparent the importance of the reference station network geometry.

9. Analysis of the classification

The data processing have been classified according to five situations: (1) discrepancy between the calculated coordinates and those approved by the IBGE, (2) solution of the ambiguity, (3) PDOP value, (4) horizontal accuracies and (5) vertical accuracies. Thus, it can be affirmed that:

- Regarding the coordinate discrepancies, the rejected data processing occurred only for those that had different distances between RBMC stations, i.e., they had a bad geometry. Usually these discrepancies were evident in the East-coordinate;

- Regarding the PDOP, it was noted a trend that most of the bad results for the PDOP values were for baselines bigger than 1,300 km and with different distances between RBMC stations (bad geometry). The variation of the PDOP for most part of data processing was small;

- Regarding the ambiguity, it was expected float solution in almost all processing. However, some surprising results were obtained, as the fixed solution of ambiguity for baselines greater than 700 km. These cases must be carefully interpreted, since the coordinates obtained from the data processing may present considerable errors when compared with the real value of the related. Fix ambiguity solution for long baselines does not ensures a good quality for the coordinates;

- Regarding to the horizontal and vertical accuracies, the values found can be accepted according to the manufacturer's specification for those distances. However, it has been found that for long

baselines and with a bad geometry among the stations, the values in several cases were worse than specified. It is important to highlight that, even with baselines greater than 1,000 km; in some cases, the accuracies were found to be smaller than 20 cm.

10. Accuracy classification table

The considerations adopted in this paper to classify the results are reliable. After result analysis, the authors propose the Table 7 for future investigation on this matter.

Table 7: Classification table of accuracy

Parameters	Classes
1. Good geometric configuration between the control stations and the unknown point, and baselines smaller than 700 km; 2. Satisfactory solution of the ambiguity in all baselines (fixed ambiguity solution for baselines up to 300 km); 3. PDOP smaller than 3 for all baselines; 4. Horizontal and vertical accuracy lower than 20 cm on all baselines.	A
One of the previous conditions not fulfilled	B
Two of the previous conditions not fulfilled	C
Three or more of the previous conditions not fulfilled	R

11. Conclusion

The coordinate accuracy values resulting from GNSS data processing and adjustments, is a hard factor to be determined in a day-by-day basis. In this article, it is proposed a reliability classification table based on different factors as baseline lengths, PDOP values and geometric configuration of the reference stations. As it was stated in the article, the statistical factors generated in the data processing and adjustment cannot be individually considered as a reliability indicator. Even the ambiguity solution cannot always determine the measurement quality once it is possible to reach fixed solution even for baseline length greater than 300 km. However, the reliability of this result is doubtful, demonstrating once more the importance of a proper choice of control points that will compose the coordinate determination.

The article demonstrates that the modeling of the results of GNSS data processing allows the classification of them in the form of a table of reliability. As a result, it is recommended to use the presented classification table as a parameter to analyze the quality of planned land surveying works. Particularly this recommendation may be applied for planning land survey works in rural properties in Brazil aiming to comply with the denominated INCRA's Law 10.267/2001. Based on the results obtained in this research, a level of reliability classified as class A or B may be qualified to obtain INCRA´s certification.

REFERENCES

Amiri-Simkooei, A., Rimkooel, A. R., Teunissen, P. J. G. and Tiberius, C. C. J. M. "Application of Least-Squares Variance Component Estimation to GPS Observables." *Journal of Surveying Engineering*, Vol. 135, No. 4, 149-160. Online Publication Date: 15 Oct 2009.

Bakula, M. "Study of Reliable Rapid and Ultrarapid Static GNSS Surveying for Determination of the Coordinates of Control Points in Obstructed Conditions." *Journal of Surveying Engineering*, Vol. 139, No. 4, 188-193. Online Publication Date: 3 Apr 2013.

Baselga, S. and García-A, L. "GNSS Differential Positioning by Robust Estimation." *Journal of Surveying Engineering*, Vol. 134, No. 1, 21-25. Online Publication Date: 1 Feb 2008.

Domingues, Sarubi, M. K. "Proposta de classificação de linhas de base obtidas com dados GPS, à luz de árvore de decisão." *Dissertação de Mestrado*. Escola de Engenharia de São Carlos, USP, 2010.

IBGE. http:// www.ibge.gov.br. Instituto Brasileiro de Estatística. Last access in 30/03/2015.

IGS. IGS. http: igscb.jpl.nasa.gov/components/prods.html. Last access in 29/01/2010.

INCRA (Instituto Nacional de Colonização e Reforma Agrária). *Norma Técnica para Georreferenciamento de Imóveis Rurais.*" 2ª Edição; 2013.

LEI N° 10.267, de 28 de agosto de 2001. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. Brasília, DF.

Matsuoka, M. T., Camargo, P. O. and Batista, I. S. "Impacto de Explosões Solares no Comportamento da Ionosfera e no Posicionamento com GPS na Região Brasileira: Estudo de caso para o dia 28 de outubro de 2003." *Boletim Ciências Geodésicas*, v. 12, n0 2, 315-334, julho-dezembro, 2006.

Menzori, M. "Classificação da exatidão de coordenadas obtidas com a fase da portadora L1 do GPS." *PhD Thesis*, Escola de Engenharia de São Carlos, USP, 2005.

Miller, C., O'Keefe, K. and Gao, Y. "Time Correlation in GNSS Positioning over Short Baselines." *Journal of Surveying Engineering*, Vol. 138, No. 1, 17-24. Online Publication Date: 19 Feb 2011.

Moussa, M., Ruwanpura, J. and Jergeas, G. "Decision Tree Modeling Using Integrated Multilevel Stochastic Networks." *Journal of Construction Engineering and Management*, Vol. 132, No. 12, 1254-1266. Online Publication Date: 1 Dec 2006.

NOAA. National Oceanic and Atmospheric Administration. http://www.sec.noaa.gov/. Last access in 31/01/2011.

Segantine, P.C.L. "GPS – Sistema de Posicionamento Global." Edited by Escola de Engenharia de São Carlos, USP, 2005.

Yeh, Ta-K., Chung, Yi-D., Wu, Chien-T., Wang, Chuan-S., Zhang, K. and Chen, Chieh-H. "Identifying the Relationship between GPS Data Quality and Positioning Precision: Case Study on IGS Tracking Stations." *Journal of Surveying Engineering*, Vol. 138, No. 3, 136-142. Online Publication Date: 21 Dec 2011..