



Metaheuristic-based stochastic models for GNSS relative positioning planning

Ismael Érique Koch¹ · Ivandro Klein² · Luiz Gonzaga Jr.¹ · Vinicius Francisco Rofatto³ · Marcelo Tomio Matsuoka³ · João Francisco Galera Monico⁴ · Maurício Roberto Veronez¹

Received: 23 January 2023 / Accepted: 7 October 2023 / Published online: 29 October 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

A priori variance–covariance matrix (VCM) estimation of global navigation satellite systems (GNSS) double difference observations in relative positioning is challenging. Existing methods have been limited to estimate variances only or present variables challenging to acquire a priori, unfeasible for observation planning. Ignoring the covariances produces misleading results and compromises reliance on GNSS positioning design. In this study, we propose models to estimate the VCM a priori for planning, based on simple variables accessible to any professional: observation time span, vector length, and ephemeris type. Using a database of over 140,000 GNSS vectors with double difference (DD) observations, we group the data by time span and length range and extract standard deviations and covariances for the linear regression process. The Isolation Forest algorithm is employed to filter outlying observations. Our models provide standard deviations and root square covariances in a local coordinate system, requiring only vector length, observation time span, and ephemeris type as input. Additionally, the equations can be easily implemented in a simple spreadsheet. The results show high coefficients of determination ($R^2 > 0.8$). We tested the models in a simulated GNSS network and verified broadcast ephemeris resulted in 6.5 to 16.7 times larger error ellipsoids compared to the precise ephemeris, indicating higher uncertainty. Ellipsoids differed in flattening and orientation when compared to the null covariance (variance only) approach. Although VCM models better reflect the precision of relative positioning observations, they did not affect the number of non-controllable observations in the observation's reliability tests.

Keywords GNSS · Relative positioning · Planning · Uncertainty modeling · Variance–covariance matrix · Metaheuristics

Introduction

Relative positioning is widely used in several applications, producing three-dimensional (3D) baseline vectors to obtain the coordinates of the points of interest. The

variance–covariance matrices (VCM) of these estimated vectors, which describe the uncertainty of the baseline components, are used as a quality measure, and the confidence region of the position solution depends directly on it. Estimating the uncertainty of these vectors a priori for

✉ Ismael Érique Koch
isma.koch@gmail.com

Ivandro Klein
ivandro.klein@ifsc.edu.br

Luiz Gonzaga Jr.
lgonzaga@unisinis.br

Vinicius Francisco Rofatto
vinicius.rofatto@ufu.br

Marcelo Tomio Matsuoka
tomio@ufu.br

João Francisco Galera Monico
galera.monico@unesp.br

Maurício Roberto Veronez
veronez@unisinis.br

¹ Unisinis, Av. Unisinis, São Leopoldo, RS 95093022-000, Brazil

² Department of Civil Construction, Federal Institute of Santa Catarina, Florianópolis, SC 88020-300, Brazil

³ Graduate Program in Agriculture and Geospatial Information, Institute of Geography, Federal University of Uberlândia, Monte Carmelo, MG 38500-000, Brazil

⁴ São Paulo State University UNESP, Presidente Prudente, SP 19060-900, Brazil

planning activities can lead to better decision-making and cost reduction in the fieldwork. However, defining a priori VCM is challenging, and correlations are often neglected in the first phases. Incorrect estimation, or neglecting the covariances, leads to misleading precision and reliability values not accurately representing the real uncertainties in the field campaign (El-Rabbany 1994; Kermarrec and Schön 2016).

In the context of survey planning, it would be highly advantageous to predict covariances based on simple variables, such as observation time span and vector length. Several previous studies have successfully achieved this for variances in post-processed relative positioning (Eckl et al. 2001; Gatti 2004; Soler 2006; Schwieger 2007; Ozturk 2011; Soyacan 2011; Firuzabadi 2012; Gond 2023), as well as in other methods like PPP (Precise Point Positioning) and RTK (Real Time Kinematic) (Geng 2010; Ögütçü 2018; Gökdağ 2020). Erdogan and Dogan (2019) and Kashani et al. (2004) studied the reliability and scaling of the VCM obtained from different GPS processing software. However, to the best of our knowledge, only one study considering a priori variances and covariances (or correlations) through simple variables such as vector length and observation time span is found in Gatti (2004).

Our study aims to provide models to estimate the VCM a priori for planning, based on simple variables accessible to any professional: observation time span, vector length and ephemeris type.

We have improved upon previous works by using over 140,000 global navigation satellite system (GNSS) vectors, covering an extensive range of values for the models' variables, and processed with global positioning system (GPS) and GLONASS signals. Also, this study provides models in a local coordinate system for broadcast and precise ephemeris. Our work also benefits from machine learning techniques for data filtering and modeling, combining linear regression, metaheuristics, and penalty functions to obtain better models.

By utilizing the outcomes of the models, professionals can adjust the observation time to meet established criteria and determine whether precise final ephemeris from the International GNSS Service (IGS)—which can take from 12 to 19 days—are necessary to achieve a desired level of quality. Moreover, it is important to consider that many developing countries lack dense networks of active stations, making post-processed relative positioning a daily reality for professionals in those regions.

In this study, the processed vectors are grouped by observation time span (t), length (l), and ephemeris type to check for patterns in the discrepancies (north, east, and up) predicated on these three variables (t , l and ephemeris). First, we computed and defined benchmark values of the 3D components of each baseline used in this study. Second,

the discrepancies of each of the estimated vectors are set as input in the filtering process using the Isolation Forest algorithm (IFA) to remove outlying vectors (Liu et al. 2008, 2012). Third, the standard deviations and covariances computed for each group of vectors are fed to the proposed procedure for the models building. Each variance and covariance component are modeled separately, also distinguishing between broadcast and precise ephemeris.

A simulated GNSS network is established and both cases with null covariances (VM) and VCM are tested. Reliability measures proposed in Prószyński (2010) are computed to evaluate the observations' quality and controllability to outliers. Error ellipsoids of the stations are also drawn and compared as measures of precision.

The next section presents the data we used, the pre-processing filtering, and its organization. In the later section, the regression procedure, and the metaheuristic Independent Vortices Search (IVS) are presented, along with the strategy to compute reliability measures and their application in a simulated GNSS network. Following that, we present the models obtained to estimate the VCM (available online at https://docs.google.com/spreadsheets/d/1zvMk7J_TYmkZYdgujZVzZOKIHmcpellj4NLMbDbzI/edit?usp=sharing) and the outcomes from the implementation in the GNSS network and discuss the significance of the results. Finally, we summarize the findings and discuss final considerations.

GNSS vectors data, filtering, and organization

The investigation of variance and covariances models was based on a massive database of processed GNSS vectors, based on 43 stations of the Brazilian Network for Continuous Monitoring of GNSS Systems (RBMC) (Fig. 1). 761

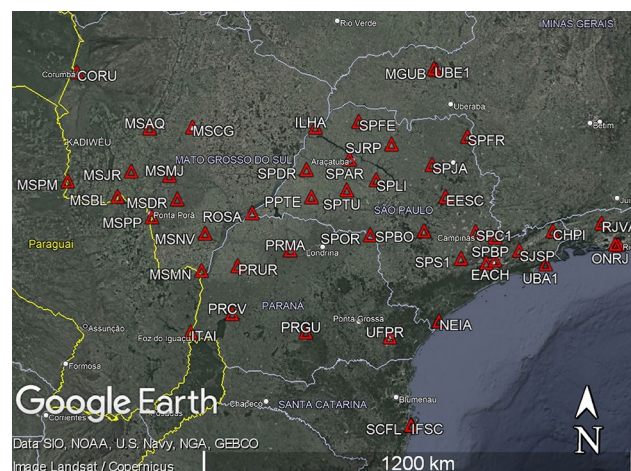


Fig. 1 Location of the RBMC stations for the GNSS relative positioning. Concentrated in the middle-west, south, and southeast of Brazil. Ranges: latitude from -18° to -28° ; longitude from -43° to -58°

three-dimensional (3D) baselines ranging from 2.3 to 950 km had their data collected on 32 different days, from January to May 2019. This period avoids data collection on high ionospheric activity. Each vector for each day of observation was processed with eight (8) unique time spans (1/2, 1, 2, 4, 6, 8, 12, and 24 hours) and separately for broadcast and precise final ephemeris. The vector processing was carried out using the software Trimble Business Center (TBC).

The TBC processing configuration was mostly kept unchanged from the default settings, as follows: the accepted solution types were fixed, and in case of failure, float; all available frequencies of satellite signals were used; the epoch processing time interval was set to automatic; the elevation mask was 10 degrees, and the enabled GNSS constellations were GPS and GLONASS. The method used to solve ambiguity varies according to the length of the vector. For vectors considered very short, less than two kilometers, TBC uses the double differences processing mode for all signal frequencies. For vectors between 2 and 20 km, double differences for all frequencies, but with ionosphere modeling. Vectors with medium lengths, from 20 to 200 km, are processed with the combination of ionosphere-free and wide lane carrier phase measurements without ionosphere modeling. For baselines from 200 to 1000 km, the software generates a mixed solution of the previous method and the Melbourne-Wübbena mode (Trimble Inc. 2018). The epoch processing time interval varies automatically according to the observation duration. While there are various alternative processing strategies, it is important to emphasize that the proposed models are applicable for processing configurations similar to the one utilized here (whether in TBC or in other software). Investigating the effect of different processing strategies on the model could be the subject of future works. Additionally, the methodology proposed for obtaining VCM models in this study is general and can be applied to different processing strategies.

All 3D vectors were transformed into a local 3D coordinate system by using the midpoint of each baseline as the reference for the transformation, applying the following (Leick et al. 2015),

$$X_{enu} = R\Delta X_{XYZ} \tag{1}$$

where X_{enu} is the vector with the transformed coordinates in the local 3D coordinate system, ΔX_{XYZ} is the vector in the Earth-centered, Earth-fixed (ECEF) coordinate system, given by

$$\Delta X_{XYZ} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = \begin{bmatrix} X_p - X_0 \\ Y_p - Y_0 \\ Z_p - Z_0 \end{bmatrix} \tag{2}$$

where $(\cdot)_p$ are the coordinates to be transformed and $(\cdot)_0$ the origin coordinates. R is the rotation matrix given by (Leick et al. 2015),

$$R = \begin{bmatrix} -\sin\lambda_0 & \cos\lambda_0 & 0 \\ -\sin\phi_0\cos\lambda_0 & -\sin\phi_0\sin\lambda_0 & \cos\phi_0 \\ \cos\phi_0\cos\lambda_0 & \cos\phi_0\sin\lambda_0 & \sin\phi_0 \end{bmatrix} \tag{3}$$

Here, λ_0 and ϕ_0 are the longitude and latitude at the origin of the local coordinate system.

We obtained benchmark values for 761 baselines by averaging vectors processed using twenty-four (24) hours of relative positioning with double differences observations and precise ephemeris. We then compared the standard deviation of the length variability with the residual of each 24-h vector. Any vector with a residual higher than three standard deviations ($\sigma > 3$) was removed, and we recalculated the average length to obtain the best benchmark values. The residuals higher than 3σ were eliminated as they may indicate problems with the baseline observation. We defined the error-free benchmark values of each baseline as the 3D components of the averaged vector.

Next, all processed vectors had their length components (e, n, u) subtracted from their benchmark pair to get the discrepancy values. Vectors with minor timespan observations or longer baselines are likely to present higher differences (Koch et al. 2022). Nevertheless, these discrepancies may indicate outlying values as GNSS vectors are also susceptible to errors. The most common causes are multipath errors, in which the satellite signal is dispersed over a surface and perceived by the receiver from different directions, and equipment errors, among others (Hofmann-Wellenhof et al. 2008).

A filtering process was conducted to identify outliers using the Isolation Forest algorithm (IFA). IFA uses a relatively simple concept, isolating the anomalies through a binary tree structure called the Isolation Tree. It contrasts with most approaches, where a model is built on the data followed by outlier detection techniques. IFA allows us to clean the database beforehand, not depending on a previous model which could have already suffered bias in the regression. Another feature that draws attention is that IFA has only two training parameters and one evaluation parameter. The number of trees (n_T) to be built and the sample size (ψ) belong to the training parameters. The tree height limit (h_{lim}) for the sample scoring makes the evaluation parameter. More details on IFA can be found at Liu et al. (2008, 2012).

Although it is possible to use a threshold-based approach to remove outliers by analyzing residuals that exceed a certain limit, IFA is capable of handling non-Gaussian distributions and correlations between variables, making it more suitable for identifying outliers for diverse data sets. IFA can detect outliers that may not be easily identified by just

looking at the residuals. Overall, IFA provides a more robust and automated approach to outlier detection than a threshold-based approach.

For the vector database, n_T was set at 2,000 and ψ at 25%, defined by a trial-and-error process while checking the results and behavior of the algorithm. Based on the heterogeneity of the data, the sample size must retain a good amount of normal data, while a higher number of trees assures thousands of sample combinations are tested, pursuing better isolations.

The remaining data contained 141.391 vectors. Approximately 4.1% and 3.9% were detected as anomalies for relative positioning processed with broadcast and precise ephemeris. All outlying vectors were removed from later experiments.

The filtered data were organized in intervals of baseline lengths, categories of eight (8) distinct time spans, and vectors processed with broadcast or precise orbits. Table 1 presents the north component standard deviations for each group of vectors. As the observation time span increases the standard deviations lower, while for longer ranges these values increase. The other two components (east and up) show a similar pattern, i.e., the shorter the time span and vector length, the lower the standard deviation. IFA identified vectors up to 50 km in length with 24 h of observation as anomalies. This could be explained by the limited number of short vectors in the database, causing them to be wrongly identified as anomalous due to their relative position (away) from other data points in the set space. However, this did not impede the construction of the models since other intervals

Table 1 Standard deviations from reference baselines (mm) for the North component. Values shaded in orange represent more significant deviations, while values in blue indicate smaller deviations

Orbits	Vector length (km)		Time span (h)								
	Range	Mean	0,5	1	2	4	6	8	12	24	
Broadcast	0 - 50	15	16,9	10,9	10,2	9,1	3,9	3,7	2,0		
	50 - 200	132	32,6	19,3	10,8	6,1	4,9	4,2	3,5	2,7	
	200 - 350	264	39,4	26,3	20,1	10,3	7,6	6,5	5,4	4,5	
	350 - 500	424	41,1	34,1	30,8	15,6	11,1	9,2	8,2	7,0	
	500 - 650	564	71,8	41,7	35,2	20,7	14,7	12,5	11,5	9,8	
	650 - 800	717	78,2	50,1	43,6	24,9	19,2	16,5	16,0	13,5	
	800 - 950	870	103,1	56,6	50,3	32,1	23,9	20,5	20,3	17,2	
	Precise	0 - 50	19	7,4	8,8	7,7	5,2	2,6	2,3	1,6	
50 - 200		132	16,1	9,9	6,7	4,6	3,7	3,2	2,6	2,0	
200 - 350		271	17,4	11,7	7,9	5,2	4,4	3,8	3,4	2,7	
350 - 500		422	19,0	12,1	8,2	5,6	5,0	4,2	3,7	2,9	
500 - 650		562	19,1	12,1	8,2	5,9	5,2	4,6	4,2	3,2	
650 - 800		717	17,2	11,7	8,4	6,1	5,4	5,0	4,3	3,2	
800 - 950		871	22,4	13,6	8,5	6,4	5,5	5,0	4,4	3,4	

were present. Moreover, in most real-life applications, 24-h observations on short length vectors are rare.

The covariances were also computed from each group of vectors, assembled by the same variables as in Table 1. Here we chose to work with the squared root of the covariances values to produce better models. Table 2 presents the covariances for X and Y axes ($\sigma(X, Y)$) in absolute squared values (mm).

A small GNSS network was also designed to simulate the application of the models we produced. Four stations of the RBMC (MSMN, MSAQ, SPBO, SPFE) forming six 3D vectors with a single control point (MSMN). This resulted in nine (9) parameters to estimate, 18 observations and nine (9) degrees of freedom. The observation time simulated was six hours and the vectors ranged from 340 to 610 km long.

Estimation approach

To build the models, a linear regression was set up using the data from the three variables we investigated: time span (t), length (l), and ephemeris type. The two first variables were repeated in the equation by adding an unknown exponent to each, plus a relation within the two (l/t) and a constant (b_0). The ephemeris type was treated in different regressions to

Table 2 Absolute square XY covariances (mm). Values shaded in orange represent more significant covariances, while blue shades indicate smaller values

Orbits	Vector length (km)		Time span (h)								
	Range	Mean	0,5	1	2	4	6	8	12	24	
Broadcast	0 - 50	15	296	271	96	54	44	25	15		
	50 - 200	132	1076	508	133	68	54	36	27	18	
	200 - 350	264	1498	638	122	145	125	83	67	39	
	350 - 500	424	1583	470	148	156	173	125	94	55	
	500 - 650	564	1162	888	68	156	277	202	136	67	
	650 - 800	717	690	1151	157	287	427	266	173	70	
	800 - 950	870	578	669	398	125	649	370	187	65	
	Precise	0 - 50	19	103	95	68	33	19	10	5	
50 - 200		132	572	300	130	54	40	30	20	12	
200 - 350		271	645	391	218	84	66	48	41	22	
350 - 500		422	573	406	226	106	82	67	50	26	
500 - 650		562	625	412	261	116	104	82	56	27	
650 - 800		717	402	417	246	119	104	87	46	23	
800 - 950		871	348	372	235	131	100	85	56	25	

produce models for each type of orbit. This equation was set for all three standard deviation and covariance components in a local coordinate system (Koch et al. 2022),

$$\sigma(\cdot)(t, l) = b_0 \frac{1}{t} + b_1 l + b_2 \frac{l}{t} + b_3 \left(\frac{1}{t}\right)^x + b_4 l^y + b_5 \left(\frac{l}{t}\right)^z + b_6 \quad (4)$$

where $\sigma(\cdot)$ stands for the standard deviation of the respective component (e , n , u , X , Y or Z), while t and l represent the observation time span in hours and the length of the vector in kilometers, respectively. The exponents x , y and z are determined during the regression process by a metaheuristic algorithm (MH). The same equation is reproduced for the covariances models where $\sigma(\cdot)$ becomes $\sigma(\cdot, \cdot)$ for en , eu , nu , XY , XZ , and YZ covariances.

Independent vortices search

To define the exponents of (4) and evaluate the penalty function result, the metaheuristic Independent Vortices Search algorithm (IVS) was implemented (Koch et al. 2019). IVS is applied for optimizing numerical functions with a unique solution. It is based on the idea of vortices that move through the search space, narrowing the production of new candidate solutions around the best solution for each vortex. The center of each vortex coincides with the best solution at that moment, and all vortices behave independently.

The candidates are produced using a normal distribution around the center where the extent (radius) of the vortex equals 1σ . Because the radius size is decreased over the cycles, the search shifts from a global to a local search (exploration to investigation) (Koch et al. 2019).

A fitness function evaluates the result for each candidate solution produced by any vortex and normalizes the values of f_i within the range of $[0;1]$.

$$\text{fit}_i = \frac{1}{1 + |f_i|} \quad (5)$$

where f_i results from the objective function of the i -th candidate.

IVS only needs a few parameters to be determined: the size of the search space common to all metaheuristics, the total count of cycles; the number of vortices; and the number of candidate solutions generated by each vortex in each cycle. For more details on IVS, the reader can refer to Koch et al. (2019) and Dogan and Olmez (2015).

Penalizing and approach overview

In the regression process, a penalty function was introduced based on Koch et al. (2022) study.

$$\min \sum_0^n \left| \frac{r_n^p}{r_n} - 1 \right| \quad (6)$$

For the regressions, Eq. (6) provides better adjustment to the more precise data (less discrepancies), where n is the number of data points, r_n^p is the n -th estimated value from the regression, and r_n is the n -th value in the dataset. As the proportion of the i -th estimated discrepancy (r_i^p) and i -th discrepancy in the dataset (r_i) approaches 1, the sum will tend to 0, reducing the penalty value applied.

Once the penalized result is computed, it returns to IVS, and the fitness of the result is evaluated. If IVS has reached a stopping criterion (the number of fitness evaluations), the algorithm verifies the null hypothesis at 5%. Any variable that presents a p -value higher than 5% is removed from the model. In case two or more variables presented p -values higher than 5%, the variable associated with the highest p -value is removed. The algorithm restarts with the remaining variables and adjusts the number of fitness evaluations in IVS when necessary.

If IVS reaches the stopping condition and all variables pass the null hypothesis test at 5%, the linear regression coefficients are returned along with the exponents estimated by the metaheuristic. The equation can then be fulfilled based on the initial model (4). A diagram detailing the procedure is presented in Fig. 2.

Occasionally, none of the exponent variables might remain in the expression (4) after continuously removing them based on their p -value. If so, the MH execution can be skipped as no exponents remain in the model. The null hypothesis test is still applied directly after the linear regression. The disadvantage of such a situation is that the penalty function is no longer activated as it only influenced the estimation of the exponents by the MH.

Validating models results in a network

To validate the VM and VCM models, they were applied for the network we presented and considered for both orbits, precise and broadcast. The network adjustment was computed using the Gauss–Markov model for linear or linearized problems, computing the residuals and comparing error ellipsoids of the adjusted vertices. Reliability measures were also evaluated for the observations.

Reliability measures

For the VM model, the local redundancy number indicates the fraction of a possible (non-random) error in an observation that is reflected in the respective residue of that observation. It is obtained by computing the redundancy matrix R_L given by

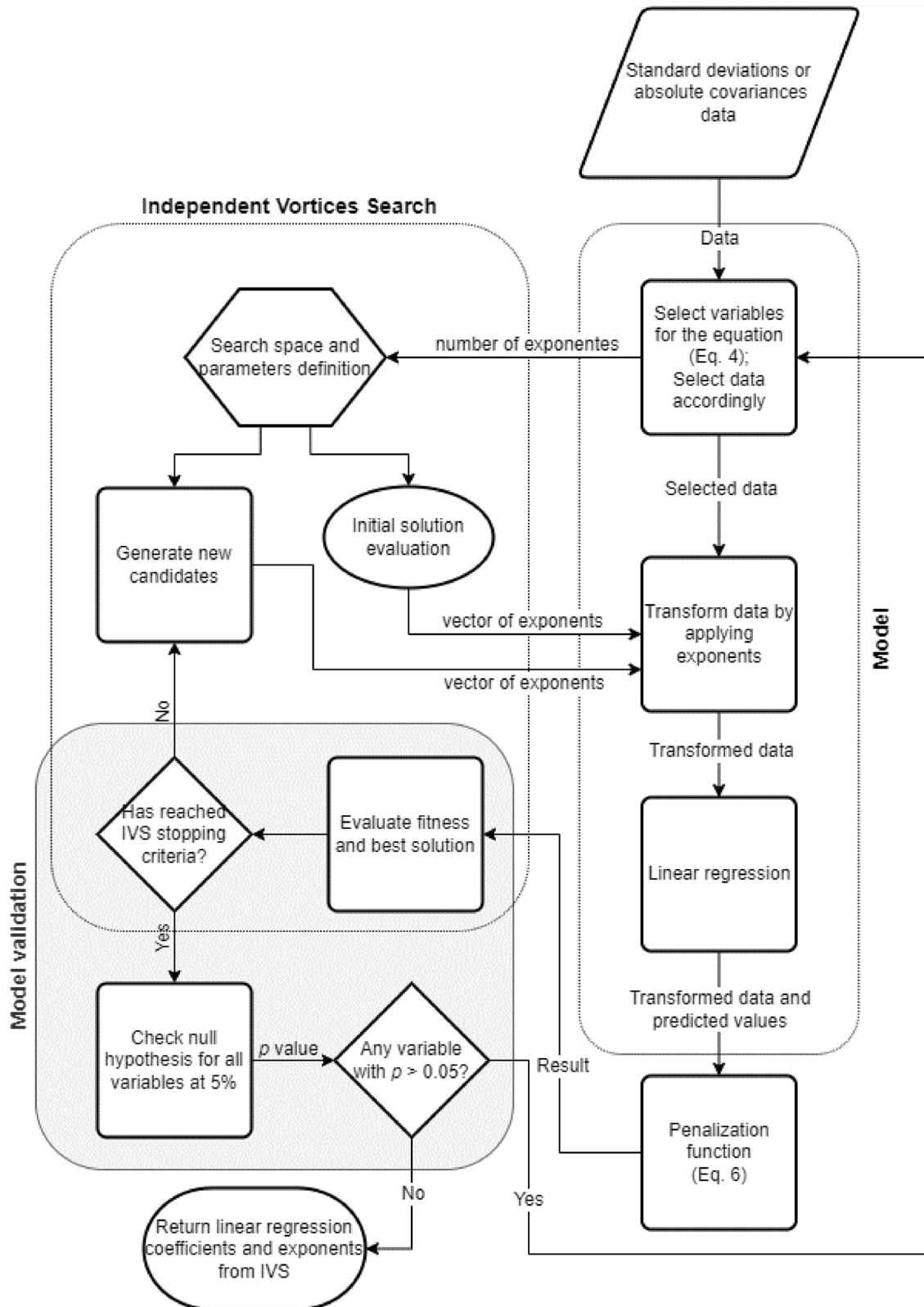


Fig. 2 Procedure for linear regression with the application of IVS and penalty function

$$R_L = I - A(A^T W_d A)^{-1} A^T W_d \tag{7}$$

where I is an identity matrix. Here, $W_d = \sigma^2 \Sigma_d^{-1}$ with Σ_d as a diagonal matrix composed by the VM of each observation. If $0.5 < r_i \leq 1$ for r_i representing the i -th diagonal element of R_L , i.e., $r_i = \text{diag}(R_L)_{ii}$, than the i -th related observation will have at least half possible (non-random) error reflected in the respective residual. The observation has good controllability.

Similarly, for the VCM model, a two-parameter reliability measure was adopted as proposed in Prószyński (2010) for the tests with covariances. It transforms the observations into correlated dimensionless variables of equal accuracy by dividing the system by $\sigma_s = \sqrt{\text{diag}(\Sigma)}$.

The system chosen to obtain the network LS solutions is also transformed, resulting in (Prószyński 2010):

$$v_s = A_s x - l_s \text{ with } \Sigma_s = \sigma_s^{-1} \Sigma \sigma_s^{-1} \tag{8}$$

where l_s , A_s , and v_s are the transformed vector of observations, design matrix, and vector of residuals, respectively, obtained by $l_s = \sigma_s^{-1} l$, $A_s = \sigma_s^{-1} A$, $v_s = \sigma_s^{-1} v$ (Prószyński 2010).

The two parameters that measure the reliability of each observation in the network are then evaluated. h_{ii} is the i -th diagonal element of H (Prószyński 2010):

$$H = I - A_s (A_s^T \Sigma_s^{-1} A_s)^{-1} A_s^T \Sigma_s^{-1} \tag{9}$$

if $0.5 < h_{ii} \leq 1$ then the first criteria is met, here meaning that at least half the error will be reflected on the i -th residual, similar to r_i from (7) which in this study yields the same result (Prószyński 2010).

The second parameter is computed as

$$k_i = \frac{1}{h_{ii}} - \frac{w_{ii}}{h_{ii}^2} - 1 \tag{10}$$

which gives the relation to observations disturbance on local and quasi-global errors. If $0 < k_i < 1$ then the local error effect of the i -th observation is greater on the i -th residual than on the quasi-global effect on all other residuals (Prószyński 2010). w_{ii} is defined by

$$w_{ii} = h_{ii} - (H^T H)_{ii} \tag{11}$$

where w_{ii} is the effect of the asymmetry of the i -th row and the i -th column of H .

Since our VCM models are designed for GNSS observations and geodetic network planning, establishing readability measures is very important. Regarding network responses to outlying observations, these reliability measures produce interpretable results. The reader can find more details in Prószyński (2010).

Results and discussion

We applied the proposed procedure to obtain standard deviations and covariances estimation equations according to the ephemeris chosen. The equations were used in a GNSS network, and null covariances and a complete VCM were compared.

Models

By implementing the procedure detailed in Fig. 2, 24 equations were obtained that model the standard deviations and covariances of the GNSS DD relative positioning as a function of the time span (t in hours) and vector length (l in kilometers). Based on the collected data, all components have different models for broadcast and precise ephemeris. Standard deviations are given in meters (m) and covariances were squared in the estimation process, meaning they are also represented in meters (m) (Table 3).

All equations kept at least one coefficient related to l and t , for both orbit products. It is essential to note that using values of l and t outside the studied intervals may produce undesirable and misleading results. Nevertheless, this study covers most practical applications since quantities beyond these intervals are rarely employed, allowing for widespread use of the models.

Regressions were also run with Eckl et al. (2001) equation to compare the models produced for standard deviations and root squared covariances. The obtained models were based on the same data, but without IVS and penalty functions, using only the four coefficients proposed in Eckl et al. (2001). The models' results were analyzed and summarized in Table 3.

The procedure to obtain the models presented in Table 4 showed better results when comparing the R^2 and RMSE values. Models with more coefficients and IVS better represented standard deviations and root squared covariances in all components (Figs. 3 and 4). For vectors with 8 to 24h of observation time span and up to 650 km long, where smaller deviation values are expected, our models reduced RMSE by 75.5% and 22% for standard deviation and root squared covariances models, respectively. This is mostly accomplished by the penalization function presented in Eq. 6, which provides better adjustment to the more precise data.

The resulting equations can be easily applied to planning GNSS relative positioning with double differences applications and require only two variables to be known: the time span of the observation, which is easily managed, and the vector length that can be planned without difficulty.

Table 3 Comparative analysis: quality metrics of obtained models vs. Eckl et al. (2001) regression models

	For standard deviation models	For root squared covariances models	Comments
Average R ²	+ 5.2%	+ 8.6%	Higher R ² on both standard deviation and root squared covariances models
Average RMSE	-48.4%	-7.0%	Reduced RMSE on both standard deviation and root squared covariances models

Table 4 Equations resulted from the proposed procedure for the standard deviations and covariances. $\sigma(\cdot)$ stands for the standard deviation of (\cdot) given in meters (m), and $\sqrt{\sigma(\cdot)}$ indicates the square root

of the covariance of (\cdot, \cdot) also given in meters (m). R² is the coefficient of determination of the statistical model, and RMSE is the root-mean-square error

Ephemeris	Component	Equation	R ²	RSME (m)
Broadcast	$\sigma(e)$	$-\frac{6.1795 \times 10^{-3}}{t} - 3.8831 \times 10^{-4} \frac{l}{t} + \frac{6.8333 \times 10^{-5}}{t^{-1.4093}} + 524.62325l^{-3.9585} + 1.4557 \times 10^{-3} \left(\frac{l}{t}\right)^{0.8551} - 0.0066$	0.990	4.20E-03
Broadcast	$\sigma(n)$	$\frac{0.02517}{t} + 3.7804 \times 10^{-5} \frac{l}{t} - \frac{0.01004}{t^{1.7989}} + 2.4879 \times 10^{-7} l^{1.6408}$	0.987	3.48E-03
Broadcast	$\sigma(u)$	$\frac{0.01364}{t} + 1.0713 \times 10^{-5} l - \frac{8.5144 \times 10^{-4}}{t^{3.9658}} + 174.45336l^{-3.7295} + 2.3845 \times 10^{-3} \left(\frac{l}{t}\right)^{0.5518}$	0.998	2.19E-03
Broadcast	$\sqrt{\sigma(e, n)}$	$\frac{3.6617 \times 10^{-3}}{t} + 1.0544 \times 10^{-5} l + 5.0505 \times 10^{-5} \frac{l}{t} - 9.2943 \times 10^{-7} \left(\frac{l}{t}\right)^{1.5257}$	0.880	4.83E-03
Broadcast	$\sqrt{\sigma(e, u)}$	$\frac{4.2683 \times 10^{-3}}{t} + 1.1095 \times 10^{-5} l + 2.9949 \times 10^{-5} \frac{l}{t}$	0.917	5.92E-03
Broadcast	$\sqrt{\sigma(n, u)}$	$1.0994 \times 10^{-5} l + 2.4587 \times 10^{-4} \left(\frac{l}{t}\right)^{0.6943}$	0.945	4.16E-03
Precise	$\sigma(e)$	$\frac{7.5293 \times 10^{-3}}{t} + 1.2504 \times 10^{-5} l + 2.0717 \times 10^{-5} \frac{l}{t} - 4.8669 \times 10^{-13} l^{3.4713}$	0.969	2.38E-03
Precise	$\sigma(n)$	$-\frac{0.03037}{t} + 1.2166 \times 10^{-5} l + 5.3044 \times 10^{-6} \frac{l}{t} + \frac{0.03836}{t^{0.908}} - 3.175 \times 10^{-7} l^{1.5121}$	0.988	8.78E-04
Precise	$\sigma(u)$	$\frac{3.0476 \times 10^{-3}}{t} + 1.3162 \times 10^{-4} l + \frac{0.03174}{t^{0.1381}} - 5.1498 \times 10^{-5} l^{1.1361} + 1.4813 \times 10^{-3} \left(\frac{l}{t}\right)^{0.4504} - 0.0203$	0.972	2.18E-03
Precise	$\sqrt{\sigma(e, n)}$	$\frac{2.7479 \times 10^{-3}}{t} + 1.2758 \times 10^{-6} l + 2.216 \times 10^{-9} \left(\frac{l}{t}\right)^{2.0771}$	0.921	1.06E-03
Precise	$\sqrt{\sigma(e, u)}$	$\frac{8.1522 \times 10^{-3}}{t} + 2.886 \times 10^{-6} l - \frac{5.2927 \times 10^{-4}}{t^{4.0}}$	0.866	1.98E-03
Precise	$\sqrt{\sigma(n, u)}$	$-\frac{3.5522 \times 10^{-3}}{t} + \frac{3.6728 \times 10^{-4}}{t^{3.9952}} + 0.01807l^{-0.1863} + 0.06403 \left(\frac{l}{t}\right)^{0.0276} - 0.0736$	0.858	7.59E-04

Application

All models were tested on the simulated network presented earlier, which consisted of four RBMC stations, one control point, and six 3D observation vectors. Adjustments were computed using both VCM and VM (null covariances), and the error ellipsoids and reliability measures were analyzed for each station. The volume values of the ellipsoids for each network station are presented in Fig. 5.

The error ellipsoids for the stations computed using the broadcast ephemeris were found to have larger volumes, indicating that the uncertainty of these points is substantially higher compared to using precise final orbits. For the studied network, the error ellipsoids from the broadcast models were found to have volumes ranging from 6.5 to 16.7 times greater

than with the final ephemeris, with a mean value of approximately 12.4 times greater. Comparing the error ellipsoids computed using the VCM and VM approaches showed no significant difference in the volumes. Figure 6 illustrates the error ellipsoids for station MSAQ using the local models.

Reliability measures were used to evaluate how controllable the observations in the network is. Figure 7 illustrates the percentage of rejected vectors in the tests. In the case of VCM models, both h_{ii} and k_i are considered for controllability, while only r_i is used for VM models.

The VCM models consider more uncertainty in the observations and therefore better reflect the precision involved in relative positioning observations, although no difference is detected in the number of non-controllable observations. The reliability metrics of Prószyński (2010) for individual

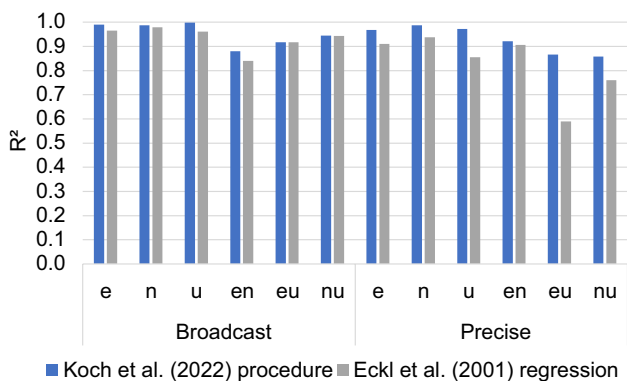


Fig. 3 Comparison of the R^2 values obtained for each component of standard deviations and root squared covariances. In blue color, values for the models produced from Koch et al. (2022) procedure with IVS and penalizing function. In gray color, values from the regressions with four components (Eckl et al. 2001)

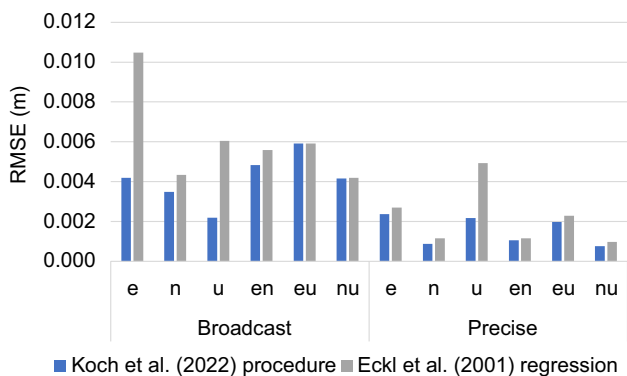


Fig. 4 Comparison of the RMSE values obtained for each component of standard deviations and root squared covariances. In blue color, RMSE values produced from Koch et al. (2022) procedure with IVS and penalizing function. In gray color, RMSE with four components models (Eckl et al. 2001)

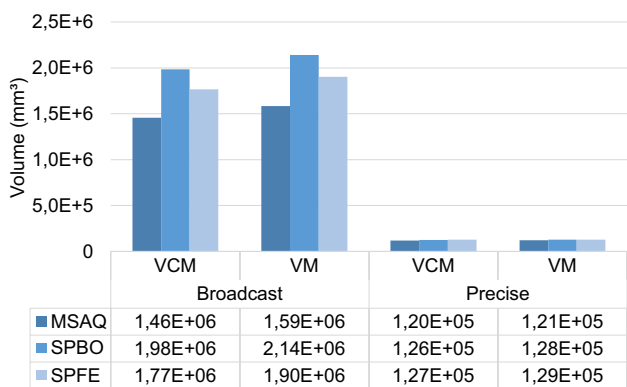


Fig. 5 Vertices volumes in the network, for broadcast and precise ephemeris and both models of variance and covariance matrices. Broadcast ephemeris present larger volumes meaning more uncertainty

observations generally vary with the type of ephemeris, with the broadcast ephemeris showing greater variation. However, the number of controlled observations within the outlier exposing area did not change with the type of ephemeris or the model (VM or VCM). This is the first study to make this type of comparison, and therefore further investigations with other networks would be necessary. Figure 8 illustrates the two-parameter reliability measure for the observations computed with the VCM models. The precise orbits resulted in less dispersion on h , but the number of observations that fit within the reliability area does not appear to improve.

The final summation of the squared residuals from the network’s adjustments is also assessed. Figure 9 shows the absolute difference between the sum of the squared residuals computed with VCM and VM.

The differences in the final sum of squared residuals between the VCM and VM models were found to be negligible, with values under 100 mm^2 . Precise orbits yielded even lower values, under 1 mm^2 . This confirms that the weight matrix, which is built from variance and covariance models, has a low impact on the estimator but more on the VCM and reliability measures.

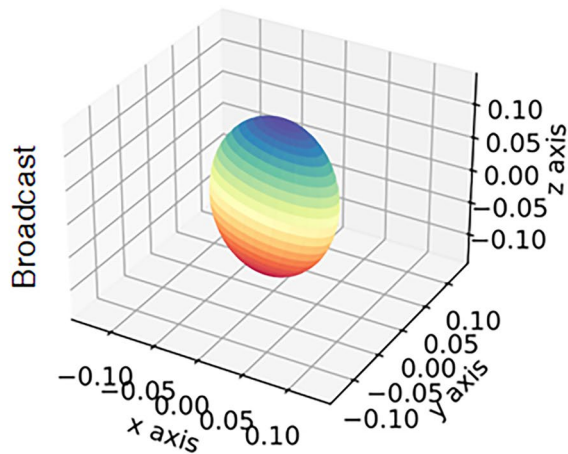
When adjusting networks, it is always advisable to use the VCM models, even if they may not improve the final solution. The VCM provides a more realistic error ellipsoid in terms of its flattening and orientation, which helps prevent the overestimation of the final precision of the computed variables. Furthermore, even though the total number of controlled observations did not change (in the global analysis), the reliability of each individual observation (local) did change. Therefore, the VCM model is also recommended in terms of observation reliability.

Conclusion

This study proposed to provide models to estimate the VCM of GNSS DD vectors a priori for planning, based on observation time span, vector length and ephemeris type. These values are critical for decision-making in structural displacement monitoring, GNSS networks, and several civil infrastructures that use GNSS observations. The study expanded the results of previous research by applying data from various vector lengths and time spans for both broadcast and precise orbits based on 140,000 GNSS relative positioning with double difference vectors. The models obtained are easy to implement with low computational cost and were tested on a simulated GNSS network.

The proposed procedure combines the IVS metaheuristic algorithm with a penalty function to better fit lower discrepancy values. It resulted in 24 models for standard deviations and covariances using broadcast and precise ephemeris for a local coordinate system.

Station MSAQ (Local)
VCM



VM

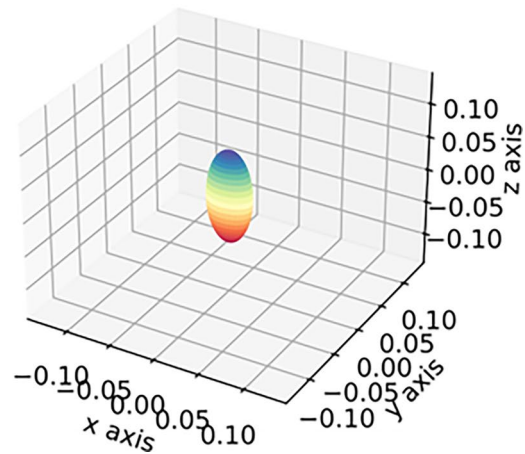
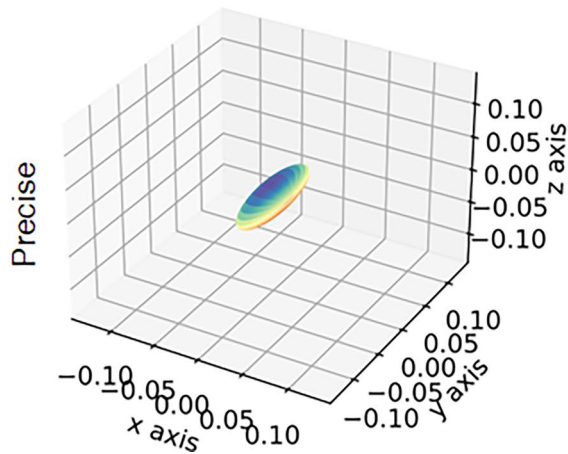
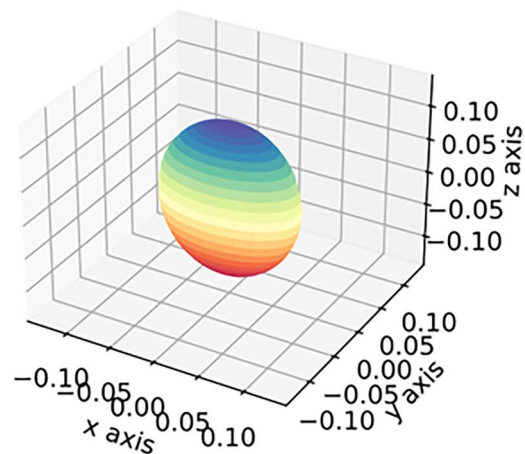


Fig. 6 Ellipsoids of station MSAQ. On the left, the error ellipsoids using the VCM for both broadcast and precise ephemeris. On the right, the ellipsoids produced when computing the errors only with

the VM. Different volumes, orientations and flattening reinforce the need to work with a full VCM

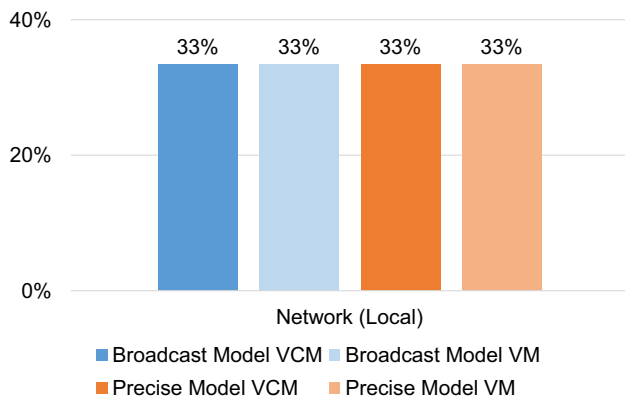


Fig. 7 Percentage of non-controllable observations. No change is detected for VCM or VM or broadcast and precise ephemeris

The obtained models require only two variables to be known: the time span of the observation and vector length, which can be planned in any project’s design phase. The RMSE ranged from 7.59×10^{-4} to 5.92×10^{-3} m for the models, and the high coefficients of determination ($R^2 > 0.8$).

When tested on a simulated GNSS network, error ellipsoids obtained with the broadcast ephemeris presented larger volumes, indicating that the uncertainty of these points is substantially higher than with precise ephemeris. Despite not presenting big changes in the volumes, in terms of flattening and orientation the use of a complete VCM brings differences when compared to the VM (null covariances) approach.

Although the VCM models consider more uncertainty in the observations and therefore better reflect the precision involved in relative positioning observations, no difference

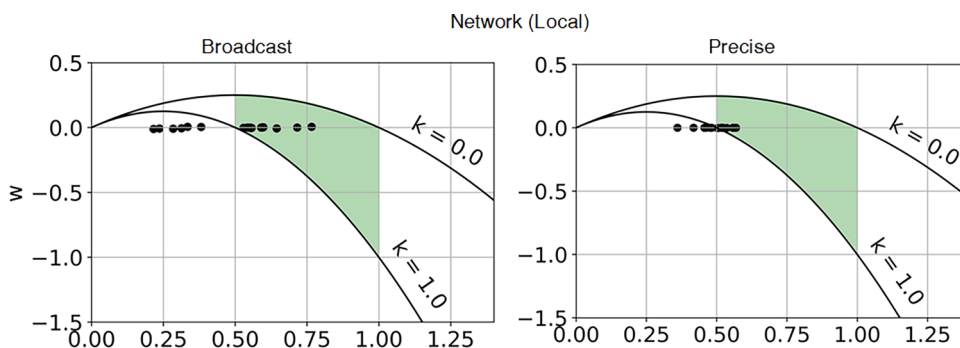


Fig. 8 Reliability measures of the observations (points) computed with the VCM models, divided into broadcast (right) and precise (left) ephemeris. The green area represents the space where the reli-

ability measures are both met. Although precise orbits presented less dispersion on h , they did not seem to improve the controllability of the observations on the network

is detected in the number of non-controllable observations by reliability measures applied in this study.

Final residuals produced by adjusting the networks with VCM and VM presented negligible differences in terms of the residual magnitude. In general, the variance and covariance models have a low effect on the estimator and more on the VCM and reliability measures. When comparing orbits, precise ephemeris reduced the differences in residuals.

In conclusion, this study demonstrated that using the VCM is essential for obtaining more reasonable a priori error ellipsoids, avoiding misleading predictions in planning observations. The linear regression-based models led to uncomplicated implementation of the models for simulations and practical use. Future research could extend the experiments with tests on high ionospheric activity and high variations in height within the stations. Also, we recommend comparing our method with expert models based on the physical properties involved, considering factors such as

the elevation angles of the satellites and the temporal correlation in the process of constructing the double differences.

Acknowledgements Partial financial support was received from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. Also, a CNPq productivity grant, process 313699/2021-6, for Ivandro Klein.

Author contributions I.E.K, I.K. and M.R.V. wrote the main manuscript text. I.E.K prepared all figures and tables. All authors reviewed the manuscript and participated in the research.

Data availability The datasets generated during and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

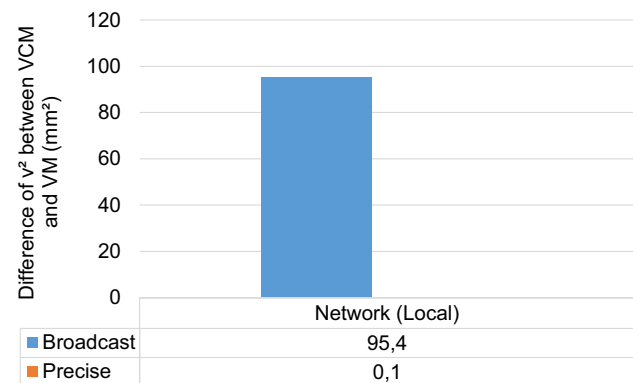


Fig. 9 Absolute difference of the sum of the squared residuals. Discrepancies between VCM and VM models resulted in negligible residuals in terms of magnitude; however, working with precise ephemeris ensures less difference

References

Dogan B, Ölmez T (2015) A new metaheuristic for numerical function optimization: vortex search algorithm. *Inf Sci* 293(125–145):S0020025514008585. <https://doi.org/10.1016/j.ins.2014.08.053>

Eckl M, Snay R, Soler T (2001) Accuracy of GPS-derived relative positions as a function of interstation distance and observing-session duration. *J Geodesy* 75:633–640. <https://doi.org/10.1007/s001900100204>

El-Rabbany A (1994) The effect of physical correlations on the ambiguity resolution and accuracy estimation in GPS differential positioning. Ph.D. Thesis, University of New Brunswick, Department of Geodesy and Geomatics Engineering, Fredericton, NB

Erdogan B, Dogan AH (2019) Scaling of the variance covariance matrix obtained from Bernese software. *Acta Geod Geoph* 54:197–211. <https://doi.org/10.1007/s40328-019-00252-w>

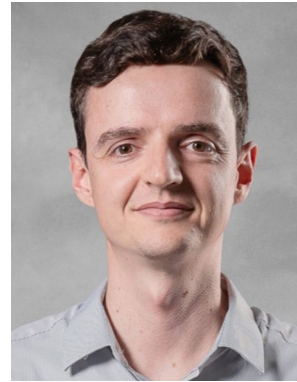
Firuzabadi D, King RW (2012) GPS precision as a function of session duration and reference frame using multi-point software. *GPS Solutions* 16:191–196. <https://doi.org/10.1007/s10291-011-0218-8>

- Gatti M (2004) An empirical method of estimation of the variance—Covariance matrix in GPS network design. *Surv Rev* 37:531–541. <https://doi.org/10.1179/sre.2004.37.293.531>
- Geng J, Meng X, Teferle FN, Dodson AH (2010) Performance of precise point positioning with ambiguity resolution for 1- to 4-hour observation periods. *Surv Rev* 42(316):155–165. <https://doi.org/10.1179/003962610X12572516251682>
- Gökdaş Ö, Özlüdemir MT (2020) A Variance model in NRTK-based geodetic positioning as a function of baseline length. *Geosciences* 10:262. <https://doi.org/10.3390/geosciences10070262>
- Gond AK, Ohri A, Maurya SP, Gaur S (2023) Accuracy assessment of relative GPS as a function of distance and duration for CORS network. *J Indian Soc Remote Sens* 51:1267–1277. <https://doi.org/10.1007/s12524-023-01701-4>
- Kashani I, Wielgosz P, Grejner-Brzezinska DA (2004) On the reliability of the VCV Matrix: A case study based on GAMIT and Bernese GPS Software. *GPS Solutions* 8:193–199. <https://doi.org/10.1007/s10291-004-0103-9>
- Kermarrec G, Schön S (2016) Taking correlations in GPS least squares adjustments into account with a diagonal covariance matrix. *J Geodesy* 90:793–805. <https://doi.org/10.1007/s00190-016-0911-z>
- Koch IÉ, Klein I, Gonzaga L, Matsuoka MT, Rofatto VF, Veronez MR (2019) Robust estimators in geodetic networks based on a new metaheuristic: independent vortices search. *Sensors* 19:4535. <https://doi.org/10.3390/s19204535>
- Koch IÉ, Klein I, Gonzaga L, Rofatto VF, Matsuoka MT, Monico JF, Veronez MR (2022) GNSS vector quality modelling combining Isolation Forest and Independent Vortices Search. *Measurement* 189:110455. <https://doi.org/10.1016/j.measurement.2021.110455>
- Leick A, Rapoport L, Tatarnikov D (2015) *GPS satellite surveying*. John Wiley & Sons Inc, Hoboken, NJ, USA. <https://doi.org/10.1002/9781119018612>
- Liu FT, Ting KM, Zhou ZH (2008) Isolation forest. In: *Proceedings—IEEE International Conference on Data Mining, ICDM*, pp. 413–422. <https://doi.org/10.1109/ICDM.2008.17>
- Liu FT, Ting KM, Zhou ZH (2012) Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data* 6. <https://doi.org/10.1145/2133360.2133363>
- Öğütücü S, Kalaycı İ (2018) Accuracy and precision of network-based RTK techniques as a function of baseline distance and occupation time. *Arab J Geosci* 11:354. <https://doi.org/10.1007/s12517-018-3712-2>
- Ozturk D, Sanli DU (2011) Accuracy of GPS positioning from local to regional scales: a unified prediction model. *Surv Rev* 43(323):579–589. <https://doi.org/10.1179/003962611X13117748892191>
- Prószyński W (2010) Another approach to reliability measures for systems with correlated observations. *J Geodesy* 84:547–556. <https://doi.org/10.1007/s00190-010-0394-2>
- Schwieger V (2007) Determination of synthetic covariance matrices—An application to GPS monitoring measurements. In *2007 15th European Signal Processing Conference* (pp. 1161–1165). IEEE
- Soler T, Michalak P, Weston N, Snay RA, Foote RH (2006) Accuracy of OPUS solutions for 1- to 4-h observing sessions. *GPS Solutions* 10:45–55. <https://doi.org/10.1007/s10291-005-0007-3>
- Soycan M, Ocalan T (2011) A regression study on relative GPS accuracy for different variables. *Surv Rev* 43(320):137–149. <https://doi.org/10.1179/003962611X12894696204867>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted

manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Ismael Érique Koch received a B.S. degree in cartographic and surveying engineering in 2017, and his M.Sc. and PhD degrees in applied computing in 2019 and 2023, respectively, from Universidade do Vale do Rio dos Sinos (Unisinos), São Leopoldo, Brazil. He is currently pursuing a PhD degree in applied computing. His current research is focused on AI techniques for application in geodesy.



Ivandro Klein received his M.Sc. and PhD degrees in Remote Sensing from the Federal University of Rio Grande do Sul state, Brazil, in 2012 and 2014, respectively. He has held lecturer and researcher positions at the Federal Institute of Santa Catarina state, Florianópolis, Brazil. His research interests include design, adjustment and quality control of geodetic networks.



Luiz Gonzaga Jr. received his M.Sc. and PhD degrees in electrical engineering from the State University of Campinas, Brazil, in 1996 and 2005, respectively. He has held lecturing and professor positions at Universidade do Vale do Rio dos Sinos (Unisinos), São Leopoldo, Brazil, where he co-created and co-heads Vizlab—XReality and GeoInformatics Lab.



Vinicius Francisco Rofatto received his PhD degree in Remote Sensing from the Federal University of Rio Grande do Sul, Brazil, in 2020. He holds lecturer and researcher positions at the Federal University of Uberlândia, Brazil. His research interests include Computational Methods for solving linear and nonlinear geodetic problems and Deep Learning in Computer Vision for geospatial applications.



Marcelo Tomio Matsuoka received his M.Sc. and PhD in Cartographic Sciences from the São Paulo State University, Brazil, in 2003 and 2007, respectively. He has held lecturer and researcher positions at the Federal University of Uberlândia, Brazil. His research interests include Quality Control of Geodetic Networks and Intelligent Computational Methods Applications in Geodesy.



Maurício Roberto Veronez received his PhD in transportation engineering from the University of São Paulo at the São Carlos School of Engineering State University, Brazil, in 2004. He has held professor positions at Unisinos, Brazil, where he co-created and co-heads the Vizlab—X-Reality and GeoInformatics Lab. His research interests include GNSS, remote sensing, and immersive visualization.



João Francisco Galera Monico received his PhD degree in Survey Engineering and Spatial Geodesy at IESSG University of Nottingham (1995). He leads the GEGE research group (Spatial Geodesy Study Group). He is currently Researcher I at UNESP and 1-A at CNPq.