Variance Components Analysis in GPS Measurements

Gligorije Perović
A Faculty of Civil Engineering, Belgrade, Serbia

Abstract: The inverse problem of errors, i.e. the structural error analysis, for GPS measurements through the variance component (VC) analysis is presented. The total measurement error is treated as a random process in time where the error is structurally decomposed into three groups due to different sources. The first group contains time variable errors due to tropospheric-ionospheric influences, $\alpha$, the second one time variable errors due to quasi-stationary refraction blocks, $\beta$, whereas the third group contains purely random errors – random process noise. The VC analysis is based on a hierarchical error structure, because the $\beta$ errors are nested in the $\alpha$ errors, since the $\alpha$ errors within a sufficiently long time interval, 2-6 h, have an approximately constant value, whereas the $\beta$ errors within a sufficiently short time interval, from about 5-10 minutes, to 2-3 h, have an approximately constant value. Therefore, the 2-way hierarchical model with random effects is used; where some solutions for unbalanced data are given. In the VC analysis the author’s PERG2FH ANOVA of GPS measurements is applied. The method has been confirmed on the results concerning measurements of the four base vectors: 2 km, 20 km, 40 km and 80 km.

Key words: GPS measurements, stochastic process, irregular residual effects, two-way hierarchical random effects model, inverse problem, variance components estimation, PERG2FH VC method.

1. Introduction

In using measurements in designing and optimising geodetic works one of the crucial elements is the accuracy of the measurements. GPS is a new technology, which in geodetic measurements has been generally used during last 20 years. Therefore, the errors of the GPS measurements are still being intensely studied – in principle every error is studied individually. So [7] considers various sources of errors for GPS measurements, as well as many sources causing noise. [17] present influences of various sources of errors in GPS observation equations and in positioning concepts. A detailed study of GPS measurement errors can be found in [2] and [9].

The multipath causes positive and negative errors in the measurements, it appears at many points of measurements, it acts from time to time within short time intervals, whereas the error due to it belongs to the group of errors which limit the measurement accuracy. For this reason a special attention is paid to multipath ([18] – signal multipath and scattering, [2], [13], [8], and [19]. However, I propose two efficient ways of reducing or removing multipath: the first is by applying a statistical treatment (gross error test), the second one – repeated measurements after 2-8 hours; therefore, multipath is not worth of such a great attention (see gross error test for $Y_{i,k}$ – Table 1).

All these errors (also including multipath) form one – the first group of measurement errors, the sum of which is the total error, to which the central limit theorem can be applied, so that the total error has a normal distribution.

In the second group of measurement errors we have irregular variations of the measurement results having short duration, caused by quasi-stationary refraction blocks – existing at every point of space and at any time [20].

In the third group of measurement errors are somewhat longer in duration irregular variations of measurement results due to tropospheric-ionospheric (Tropo-Iono) influences, also existing at every point of space and at any time. [5] examined cyclical refractivity variations and calculated the GPS L1 signal delay in a vertical refractivity profile. Based on this calculation I find that the Scale Factor variation maximal, equal to $5 \times 10^{-7}$, on January 1, 1990 – when the solar activity is high, and minimal, $1 \times 10^{-7}$, on January 1, 1995, when the solar activity is low [9]. Using the experiment performed by [6], wherein he realized a 7-day GPS campaign of measurements at seven points mutually distant between 20 km and 300 km, on the basis of six comparisons of diurnal (24 hours) measurements (one session) by transforming the coordinates with 7 parameters I have estimated the Scale Factor variations and obtained a value of $2 \times 10^{-7}$ [9] being the

\[\text{11\ We acquired the knowledge of quasi-stationary refraction blocks during the study in 1965. We were convinced of their effect during the ten-days’ student training of precise measurements of horizontal angles in stable atmospheric conditions (summer period from 6:00 h to 8:30 h and from 17:00 h to 19:30 h), by using optical-mechanical theodolites. In this measurement there appeared two or more continuous groups of results so that they match up within a priori stipulated narrow limits, but between the groups they differ much more than the allowed values.}\]
standard error of light speed in the air for one phase pseudorange measurement (with respect to a satellite), because all other errors in 24-hours measurements are reduced to a negligible value.

The presented research results for the second and third error groups indicate that VC analysis of GPS measurements based on the 2-way hierarchical model with random effects is to be applied.

2. Motivation

When designing geodetic measurements it is necessary to know the distribution parameters for both individual (component) errors and the total measurement error for each individual measured quantity. For (final) results of GPS measurements we have still not definitely examined either the variance or its estimate. The purpose of the present paper concerns just this matter.

The essence of the Belgrade school of measurement error theory, which is based on the Russian geodetic school of measurement error theory – the leading school of this kind in the world, but also significantly based on the European geodetic school – especially on the German geodetic school, is the following [9].

1) At first to study the measurement object thoroughly, where the object over time can be a constant or variable quantity, process, phenomenon, and the like.

2) In advance to study all influential quantities (error sources) causing errors in the measurement results and for every such error to determine the distribution and its parameters, in particular the expectation and standard deviation (error), and also for the deviation (error) the definite estimates of the measured quantity, the distribution and its the distribution parameters.

3) To define measurement conditions, including the instruments and equipment.

4) To define functional models serving for the purpose of a good (adequate) description of the corrections necessary to be included in the raw measurement results.

5) To determine the main variance components for the measurement results serving for the purpose of a good presentation of the total variance of measurement results, as well as of the main VC for the measured quantity estimate.

6) To define the measures and criteria for precision and reliability of measurement results.

7) To define the model serving for the purpose of estimating the measured quantity as well as the model for the case of the estimate accuracy.

8) To determine the upper accuracy limit for the measured quantity estimate (“upper method accuracy limit”).

The present paper is devoted to the objective mentioned in the first part of the sentence in item 5), i. e. it is devoted to determining the main variance components for measurement results yielding a good presentation of the total error variances in the measurement results.

3. Mathematical background - The measurement error analysis in the 2-way hierarchical classification with random effects

The For the purpose of analysing the VC the PERG2FH2) ANOVA method is used (see Sec. 5). It is based on the 2-way hierarchical classification with random effects. For this reason this method is first to be described.

Here the case concerning the unbalanced data will be considered because such a case is used in the PERG2FH ANOVA analysis of GPS measurements.

3.1. The 2-way hierarchical classification, random model

Let \( A \) be the true value of the measured quantity, \( \delta \) – constant systematic error (fixed parameter) in all measurement results (\( \delta \) can be also zero), \( \alpha_i \) – random effect of the first factor, \( \beta_{ij} \) – random effect of the second factor and \( e_{ijk} \) – purely random measurement error – „pure” error.

Preserving the designations used by [15] and [16] – except \( e \) for which Searle uses the designation \( e \), in the 2-way hierarchical classification the random effects \( \beta_{ij} \) of the second factor are nested within the random effects \( \alpha_i \) of the first factor, and the random errors \( e_{ijk} \) – within the random effects \( \beta_{ij} \) of the second factor. So the measurement model equation for 2-way hierarchical classification with random effects for unbalanced data is [12]:

\[
Y_{ijk} = A + \delta + \alpha_i + \beta_{ij} + e_{ijk} \tag{1}
\]

\[
i = 1, \ldots, a, \; j = 1, 2, \ldots, b_{i}, \; k = 1, 2, \ldots, n_{ij}; \tag{2}
\]

\[
N_i = \sum_{j=1}^{b_i} n_{ij} = n_i, \quad B = \sum_{i=1}^{a} b_{i} = b, \tag{2'}
\]

\[
N = \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij}; \tag{3}
\]

with conditions: \( n_{ij} \geq 2; \; b_{i} \geq 2; \; a \geq 2; \)
and with regard to what said above about the effects their distribution will be normal, so that the stochastic model will be:

\[ \alpha_i \sim N(0, \sigma^2_\alpha), \quad \forall i ; \quad \beta_{ij} \sim N(0, \sigma^2_\beta), \quad \forall i, j ; \quad \epsilon_{ijk} \sim N(0, \sigma^2) , \quad \forall i, j, k, \]

\[ C[\alpha_i, \beta_{ij}] = C[\alpha_i, \epsilon_{ijk}] = C[\beta_{ij}, \epsilon_{ijk}] = 0, \quad \forall i, j, k, \quad C[\epsilon_{ij}, \epsilon_{j'k'}] = 0, \quad \forall k \neq k'. \]

(4a)

where \( C[u, v] \) is the covariance operator between \( u \) and \( v \).

Here \( \delta + \alpha_i + \beta_{ij} + \epsilon_{ijk} = \Delta_{ijk} \) is the true error of \( Y_{ijk} \).

With the designation \( \mu = A + \delta = \text{constant} \) the expectation of observation will be \( E[Y_{ijk}] = \mu \) so that according to (1) to (4b) we find:

\[ Y_{ijk} \sim N(\mu, \sigma^2_i), \quad \forall i, j, k, \]

with expectation

\[ \mu = E[Y_{ijk}] = A + \delta, \quad \forall i, j, k, \]

and variance

\[ \sigma^2_i = V[Y_{ijk}] = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2, \quad \forall i, j, k, \]

where \( E[] \) is the expectation operator, \( V[] \) – variance operator, and \( \sigma^2_\alpha, \sigma^2_\beta \) and \( \sigma^2 \) are the components of measurement variance, \( \sigma^2_i \), which on the basis of measurements (1) should be determined.

So instead of model (1) we can consider the following model

\[ Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk} \quad \text{(1a)} \]

Now the procedure for VC determination [15], [16] and [12] will be given.

### 3.2. Expected sums of squares

In the basis of the ANOVA analysis lies the decomposition of the total sum of squares into three independent sums, SSA, SSB and SSE [15]:

\[ \text{SSA} = \sum_{i=1}^{a} N_i (\bar{Y}_i - \bar{Y})^2, \quad \text{SSB} = \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij} (\bar{Y}_{ij} - \bar{Y}_i)^2, \]

\[ \text{SSE} = \sum_{i=1}^{a} \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij})^2, \quad \text{(5)} \]

with

\[ \bar{Y}_{ij} = \frac{Y_{ij}}{n_{ij}}, \quad \forall i, j ; \quad \bar{Y}_i = \frac{Y_{i..}}{N_i}, \quad \forall i ; \quad \bar{Y} = \frac{Y_{..}}{N}, \quad \text{(6)} \]

\[ Y_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk} , \quad Y_{i..} = \sum_{j=1}^{b_i} n_{ij} Y_{ij} , \quad Y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} Y_{ijk}. \quad \text{(7)} \]

ANOVA is based on the distributions of the sums of squares, SSA, SSB and SSE. In this decomposition the expected sums of squares for SSA, SSB and SSE are respectively:

\[ M[\text{SSA}] = \left( N - \frac{1}{N} \sum_i N_i^2 \right) \sigma^2_\alpha + \left( \sum_{i} \frac{n_{ij}^2}{N_j} - \frac{\sum_{i} n_{ij}^2}{N} \right) \sigma^2_\beta + (a-1)\sigma^2 \quad \text{(8)} \]

\[ M[\text{SSB}] = \left( N - \sum_i \frac{n_{ij}^2}{N_i} \right) \sigma^2_\delta + (B-a)\sigma^2, \quad \text{(9)} \]

\[ M[\text{SSE}] = (N-B)\sigma^2. \quad \text{(10)} \]

### 3.3. Testing Hypotheses

For testing the influences of individual factors the mean squares are needed:

\[ m^2 = \frac{\text{MS} A}{a-1} \quad \text{d.f.}, \quad m^2 = \frac{\text{MS} B}{B-a} \quad \text{d.f.}, \]

and

\[ m^2 = \frac{\text{MSE}}{N-B} \quad \text{d.f.}. \]

At first the hypothesis on the nested factor influence is tested:

1) The hypothesis on the nested factor influence \( \beta \) is tested:

\[ H_{0,\beta} : \sigma^2_\beta = 0 \quad \text{versus} \quad H_{a,\beta} : \sigma^2_\beta > 0 \quad \text{(11)} \]

The variance \( m^2 \) does not depend on the general means.

The variance \( m^2 \) characterizes the variation between second order groups within the first order groups. So, if the general means for the second order groups within each first order group are equal one to another, then \( m^2 \) will have the \( \text{m}^2 \) distribution with parameters \( \{\sigma^2, (B-a)\} \quad \text{(7)} \). Therefore, the ratio \( F = m^2 / m^2 \) will have the central \( F \) distribution with parameters \( ((B-a), (N-B)) \), which enables us to test \( H_{a,\beta} \) versus \( H_{a,\beta} \).

---

3) SSB is the abbreviated form of SSB:A – Sum of Squares of deviations from Beta factor; within Alpha.

---

6) In [4] we find the mean deviation squares and tests for the factors – for balanced data, whereas [12] extends this toward unbalanced data.

7) If \( m^2 < m^2 \), the test is not carried out because the hypothesis \( H_{0,\beta} : \sigma^2_\beta = 0 \) - there are no influences of \( \beta \) factor is a priori accepted.

8) The distribution of statistics \( m^2 \), and [12] extends this to unbalanced data.

---

*Imperial Journal of Interdisciplinary Research (IJIR)*

Page 359

---
The test statistics is (Hald, 1957):

$$F_β = \frac{m_2^2}{m_1^2}, \text{ with } F_β | H_{0,β} \sim F(f_2, f_1).$$ \hfill (12)

1a) For the case:

$$F_β < F_{i-a; a,a,N-B} \; - \; \text{accepted } H_{0,β}.$$  

Then all observations within the first order group can be regarded as obtained from the same general population, so that on the basis of the two estimates, $m_1^2$ and $m_2^2$, we can obtain the pooled estimate for $σ^2$:

$$m_ε^2 = \frac{f_1 m_1^2 + f_2 m_2^2}{f}, \quad f = f_1 + f_2.$$  \hfill (13)

However, by applying the F criterion the equality of the general means is not confirmed. (In the further analysis, using then the test power, it is possible to conclude that it would be more appropriate not to pool the two variances).

If accepted $H_{0,β}$ from (11) and $H_{a,a}$ from (14), the problem is reduced to a 1-way problem with random effects $α$.

1b) For the case:

$$F_β > F_{i-a; a,a,N-B} \; - \; \text{accepted alternative } H_{a,β},$$  
i. e. the means of the general populations should be regarded as unequal, so that each mean value in the second order group is the estimate of the corresponding general mean $μ_i + β_{i,j}.$ Then $σ_β^2$ is estimated (see formula (25)).

1c) The consequences of testing hypotheses on the second factor influence.

1c0β) If accepted $H_{0,β}$ i. e. $σ_β^2 = 0$ then there is no estimation of $σ_β^2$ and all observations within the second order group belong to the same general population. Then on the basis of the two values $m_1^2$ and $m_2^2$ we can obtain an estimate for $σ^2$ according to the formula:

$$m_ε^2 = \frac{f_1 m_1^2 + f_2 m_2^2}{f}, \quad f = f_1 + f_2 = N - B.$$  \hfill (14)

1cαβ) If accepted $H_{a,β}$ i. e. $σ_β^2 > 0,$ then the means of the second order general populations should be regarded as unequal and, according to (35), $σ_β^2$ is estimated. Of course, then individual means $Y_{i,j}$ should be tested for gross errors.

2) Testing the hypothesis on the first factor $α$ influence.

After finishing the testing hypotheses on $β$ it follows the testing hypotheses on the first factor $α$ influence. It is tested:

$$H_{0,a} : σ_α^2 = 0 \; \text{ versus } H_{a,a} : σ_α^2 > 0.$$ \hfill (15)

2a) Always $m_1^2$ is distributed in $m^2$ with parameters $[σ^2, (N - B)]$ so that if $H_{0,β}$ is accepted, then $m_1^2$ under $H_{0,a}$ is distributed in $m^2$ with parameters $[σ_α^2, (a - 1)].$ Then the equality of the general means for the first order groups $− H_{0,a}$ is tested through the ratio $F = m_2^2 / m_1^2,$ which has a central $F$-distribution with parameters $[(a - 1), (N - B)]$ if $H_{0,a}$ from (14) is correct. If $H_{0,a}$ is also accepted, then model (1) is reduced to a number of measurements of one quantity only, $Y_{i,j,k} = μ_i + ε_{i,j,k},$ so that the estimate $σ_ε^2$ is obtained according to the formula:

$$m_ε^2 = \frac{1}{f} \sum_{i=1}^{a} \sum_{j=1}^{n_i} \sum_{k=1}^{m} (Y_{i,j,k} - \bar{Y})^2,$$  \hfill (16)  

with $f = N - 1.$

2b) Since the quantities $β_{i,j}$ are connected to the variation between the second order groups within the first order groups, and assumed to be normally distributed with parameters $[0, σ_β^2],$ it follows that $α$ variances:

$$m_{2i} = \frac{1}{b_i - 1} \sum_{j=1}^{b_i} a_{i,j} (\bar{Y}_{i,j} - \bar{Y})^2,$$ \hfill (17)  

$$f_{2i} = b_i - 1,$$  

$i = 1, 2, \cdots, a,$  \hfill (18)  

under $H_{p,β},$ have a theoretical value approximately $σ^2 + πσ_β^2$ (which can be verified, using for instance Bartlet’s test of variances equality), where is:

$$\bar{π} = \frac{1}{B - a} \left( N - \sum \frac{n_i^2}{N} \right)$$  

$$- \left( \text{“average” number of observations within a group for } β \right).$$

Now the variance $m_2^2,$ as a pooled estimate on the basis of (16), $= SBB / (B - a),$ will have an $m^2$ distribution with parameters $[σ^2 + πσ_β^2, (B - a)].$ Besides, if the hypothesis $μ_1 = μ_2 = \cdots = μ_a$ is correct, then the variance $m_2^2$ will have an $m^2$ distribution with parameters $[σ^2 + πσ_β^2, (a - 1)]$, with:

$$\bar{π}' = \frac{1}{a - 1} \left( \sum \frac{\Sigma n_i^2}{N} - \sum \frac{Σ n_j^2}{N} \right).$$  \hfill (19)

However, since $\bar{π}' \approx \bar{π}.$

If $m_1^2 < m_2^2$, the test is not carried out because then one a priori accepts the hypothesis $H_{0,a}: σ_α^2 = 0$ – that no influences of factor $α$ exist.
For balanced data the test quantity for testing the hypothesis $H_{0,a}$ is [4]:

$$F_a = \frac{m_2^2}{m_2^2}.$$  (20)

For unbalanced data the quantity $m_2^2$ under $H_{a,\alpha}$ has a $m^2$ distribution with parameters $(\sigma^2 + \pi \sigma_0^2, B - a)$, whereas the quantity $m_2^2$ under $H_{0,a}$ has a $m^2$ distribution with parameters $(\sigma^2 + \pi \sigma_0^2, a - 1)$

$^9)$. However, if the numbers of measurements within second order groups not mutually differ significantly, say less than 50%, then $F_a$ under $H_{0,a}$ will have approximately the central $F$ distribution with $a - 1$ and $B - a$ degrees of freedom, so that the test decision will be:

$$F_a > F_{1-a, a-1, B-a} \rightarrow \text{accepted alternative } H_{a,\alpha}$$

- then $\sigma_2^2$ is also estimated (see formula (26)).

One should note that the variation between the first order groups is due to both the variation within the second order groups and between them, therefore the hypothesis is tested by comparing $m_2^2$ and $m_2^2$.

2c) The consequences of testing hypotheses on first factor influence

2c0a) If accepted $H_{0,a}$ – i.e. $\sigma_a^2 = 0$ – it means that there is no first factor $\alpha$ influence, and then there is no $\sigma_a^2$ estimation. Then:

- if $H_{0,\beta}$ also accepted, then the model is $Y_{ijk} = \mu + v_{ij} + e_{ijk}$, so that we have an estimate for $\sigma_v^2$ only (see (15));
- if $H_{a,\beta}$ also accepted, then the model is $Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk}$, so that the model is reduced to 1-way with random effects $\beta$.

2caax) If accepted then the means of the general first order populations should be regarded as distinct and $\sigma_a^2$ is estimated. Then:

- if $H_{0,\beta}$ also accepted, then the model is $Y_{ijk} = \mu + \alpha_i + e_{ijk}$, so that the model is reduced to 1-way with random effects $\alpha$;
- if $H_{a,\beta}$ also accepted, then all the three variance components, $\sigma_a^2$, $\sigma_\beta^2$, and $\sigma_\epsilon^2$, are estimated (based on model (1)). In this case the first factor influences should be verified to gross errors.

3.4. ANOVA estimators

For the purpose of estimating the variance components it is necessary to equalise the square sums to their expectations (8) to (10), i.e. to use the ANOVA principle – ANOVA estimation method, where in the equations formed in such a way the variance components are replaced by their estimators $m_2^2$, $m_2^2$, and $m_2^2$ ($m_2^2$ = $m_1^2$). Thus the estimates can be obtained directly from expressions (8) to (10) by solving equations:

$$m_2^2 = A_0 m_2^2 + A_0 m_2^2 + m_2^2,$$  (21)

$$m_2^2 = B_0 m_2^2 + m_2^2,$$  (22)

$$m_2^2 = m_2^2,$$  (23)

so **ANOVA estimators** $m_2^2 = \bar{m}_a^2$, $m_2^2 = \bar{m}_\beta^2$ and $m_2^2 = \bar{m}_\epsilon^2$ will be

$$m_2^2 = m_2^2,$$  - with $f_\epsilon = N - B$,  (24)

$$m_2^2 = \frac{1}{B_0} (m_2^2 - m_2^2),$$  - with $f_\beta$,  (25)

$$m_2^2 = \frac{1}{A_0} (m_2^2 - A_0 m_2^2 - m_2^2),$$  - with $f_\alpha$  (26)

and then:

$$A_0 = \frac{1}{a-1} \left( N - \frac{1}{N} \sum_i N_i^2 \right),$$  (27)

$$A_\beta = \pi' \text{ (see (18)),}$$  (28)

- are „average” number of observations in group for $\alpha$,

$$B_\beta = \pi', \text{ (see (17)).}$$  (29)

The procedure of finding the variance components comprises the following steps:

**Step 1.** gross errors test in $\epsilon$;

**Step 2.** determination of $m_1^2$;

**Step 3.** gross errors test in $\beta$;

**Step 4.** determination of $m_2^2$;

**Step 5.** gross errors test in $\alpha$;

**Step 6.** determination of $m_3^2$.

The process is iterative because VC’s have to be used in gross error tests. Therefore the steps are interconnected: Step 1 and Step 2, Step 3 and Step 4, and Step 5 and Step 6.

Besides, [12] finds **degrees of freedom estimates**:

$$\hat{f}_\beta = \frac{(m_2^2 - m_2^2)^2}{(m_2^2)^2 + (m_2^2)^2} = \frac{B-a}{N-B},$$  (30)
and
\[ f_a = \frac{(m_i^2 - E_m^2 + E_i^2)}{(m_i^2)^2} \] for \( f_a \), \( a = 1 \) to \( N \).

where \( \Pi \) is from (17), and:
\[ E_1 = \frac{J_0 - \Pi}{\Pi}, \quad E_2 = \frac{J_0}{\Pi} \] and \( \Pi' \) is from (18).

For a large data number we use variances, \( \sigma^2 \), instead of \( m_i^2 \), \( \sigma_i^2 \) instead of \( m_i^2 \) and \( \sigma^2 \) instead of \( m_i^2 \), so instead of estimates (30) and (31), we obtain degrees of freedom:
(a) \[ f_b = \frac{(\sigma_i^2 - \sigma^2)^2}{(\sigma_i)^2 + (\sigma^2)^2} \]
(b) \[ f_a = \frac{(\sigma_i^2 - E_2 \sigma_j^2 + E_i \sigma_j^2)}{(\sigma_i)^2 + E_2 (\sigma_j)^2 + E_i (\sigma_j)^2} \] for \( f_a \), \( a = 1 \) to \( N \).

4. PERG2FH method of VC analysis in GPS measurements

For the purpose of VC analysis of GPS measurements the PERG2FH method is used. The analysis has been based on GPS measurement error presentation by means of its three structure groups – as explained in Sec. 1 – Introduction. In its application the gross errors test is simultaneously used, where as the test statistics the deviation from the mean value within a group (cell) is used. Due to the hierarchical structure of the measurements errors (caused by influences of corresponding factors) the test are rather complicated, so that a new testing method is formed. Its presentation will not be given here because of saving space, it is left for one of the future articles.

Note. The application of gross error tests is preceded by elimination of the results containing large gross errors (blunders, gross outliers). These are the results outlying by order of 5 centimeter, and more.

4.1. GPS measurements as random processes

Modern electronic technologies enable us to achieve a large number of measurements for one quantity (as well as several quantities) in time. Then (in time) the measurements are affected by some factors leading to measurement errors variable in time. Such are, for instance, the coordinate measurements for points: by using GPS, by airborne laser scanning of asphalt roads, on satellite images, and the like.

A group of factors affecting GPS measurements are residual ionospheric and tropospheric influences. Their influences are variable in time, with oscillation periods of 24 hours, one year, and in the case of the ionospheric influences even 11 years. Since the temperature of the atmospheric layers along the path of the rays has an utmost influence on the change of the ray path and at a single point the temperature differences day – night are of order of 15° C, the annual ones of order of 45° C, the annual variations of the tropospheric influences are significantly larger than the diurnal ones. In an analogous way the variations in ion number density in the ionosphere result in much higher variations of ionospheric influences over the 11-years period than for the cases of the annual and diurnal levels.

In his monograph about refraction – terrestrial, for ground observations towards Earth artificial satellites, and astronomical – Yunoshev (1969) presented long – period and detailed examinations of Russian scientists – first of all geodesists, noting that the quasi-stationary refraction blocks are crucial factor in producing refraction. The quasi-stationary refraction blocks are defined as atmospheric layers along the ray path at which stable gradients of refraction coefficients appear (10). (also see: [9] – Subsec. 6.5, and [12] – Subsec. 9.5). The duration of these blocks is between about 10 minutes and 2-3 hours. This would be another group of factors affecting the GPS measurements. By defining quasi-stationary blocks refraction studies by applying statistical methods becomes possible, and in the case of GPS measurements it becomes also possible to undertake statistical studies of a group of errors variable in time.

Therefore, from a theoretical viewpoint point GPS measurements of the coordinates of a single point during a sufficiently long time interval appear as a continuous random process, whereas the real measurements – „several times“ registrations appear as intersections of this random process, i. e. a discrete random process.

4.2. PERG2FH ANOVA GPS measurements

Modern electronic technologies not only enable us a large number of intersections (measurements) of a stochastic process over time, but simultaneously with intersection registering make it possible to register the summary influence of all factors in the measurement result.

Now, for the purpose of analysing, the errors of GPS measurements can be structurally decomposed into three groups due to three sources:

(10) In Yunoshev’s monograph there is also a description of the ways how to determine the refraction angles in real atmosphere, in particular: following the results of meteorological measurements, by means of measuring angles of differential refraction and by means of geodetic measurements; as well as agreement between the obtained results.
1. random process noise (purely random errors), $\varepsilon$, and two other factors,

2. first factor – tropospheric-ionospheric influence, $\alpha$, and

3. second factor – quasi-stationary refraction blocks influence, with summary influence $\beta$; its action on the GPS measurements is always present in time, but has irregular variations with approximately constant values within intervals from 5-10 minutes to 2-3 hours.

Some methods for stochastic process noise variance determination have been given by [10], [11].

Then each measurement contains a different influence of every individual factor, but the influences of the same factor over a short time interval are mutually slightly different; for instance, the influences of the second factor $\beta$ are within the 10-minutes interval are negligibly different from another, whereas in the case of the first factor such an interval should be examined – but generally an interval of 2-6 hours is acceptable. Besides, it is possible that there also exist seasonal influences, which is also to be studied.

Now the GPS measurements can be represented with $Y_i$, with the following structure:

$$Y_i = \mu + \alpha_i + \beta_i + \varepsilon_i, \quad i = 1, 2, \ldots, N$$

– for observations,

$$\Delta_i = \alpha_i + \varepsilon_i, \quad i = 1, 2, \ldots, N$$

– for true errors.

However, the real state of factor values is as given in Fig. 1, so that the second factor, $\beta$, over the interval $(i, j)$ produces approximately constant effects, $\beta_{ij} \approx const_{\beta}$, whereas the first factor, $\alpha$, produces approximately constant effects, $\alpha_i \approx const_{\alpha}$, over the interval $(i)$, i.e.:

(a) $\alpha_i \approx const_{\alpha}$ – over interval $(i)$

(b) $\beta_{ij} \approx const_{\beta}$ – over interval $(i, j)$

Therefore, the 2-way hierarchical classification model – from (1) with (1a), to (4b) with random effects, can be applied. So the GPS measurements model, which can be equally applied to any coordinate of a point and to any coordinate difference for the base vector, can be represented as [12]:

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}, \quad (with \Delta_{ijk} = \alpha_i + \beta_{ij} + \varepsilon_{ijk}) \quad (34)$$

with $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, $k = 1, 2, \ldots, n_{ij}$,

$$N_i = \sum_{j=1}^{b_i} n_{ij} = n_i, \quad N = \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij}, \quad \text{and with conditions: } n_{ij} \geq 2, \quad b_i \geq 2, \quad a \geq 2.$$

So the first factor, $\alpha$, would be the tropospheric-ionospheric influence producing the scale factor variations, whereas the quasi-stationary refraction blocks influence would be the second factor, $\beta$, producing irregular variations with respect to the trend – given through the first factor influence (Fig. 1). Finally there are also purely random deviations $\varepsilon$ – in the sense of the Gaussian errors.

Thus the GPS measurements can be treated by using the 2-way hierarchical classification model with random effects, knowing that then a small piece of information will be lost.

![Fig. 1. Presentation of errors of GPS coordinates (without constant systematic error) by using 2-way hierarchical classification in PERG2FH method.](image)

5. Results and discussion

Example 1. For the purpose of illustrating this theory, the base vector 83 km long is chosen, between the permanent GPS stations Bor (B) and Kruševac (K) in Serbia. The fixed solutions are obtained for each registration regularly spaced 30seconds intervals for a diurnal (24-hours) measurements in a January 22, 2014. The number of the satellites, $n_S$, during measurements was: $\min n_S = 4$, $\text{mean} n_S = 7.3$, $\max n_S = 10$.

The choice of intervals for the VC analysis is:

$I_\beta = 7.5 \text{ minutes for } \beta$ and $I_\alpha = 6 \text{ hours for } \alpha$, for all the three coordinate differences: in longitude $\Delta \lambda$, in latitude $\Delta \varphi$ and in altitude $\Delta h$.

The initial number of results was $N = 2876$, the number of groups for beta was $B = 192$, and for alpha $a = 4$, while initial $n_y$ were 15 – except in the final group in which $n_y$ was 11.
The gross error tests were applied to deviations of single results from the mean value (local tests) with the significance level $\alpha_0 = 0.001$. The number of results removed is given in Table 1.

### Table 1. Number of results removed at $I_a = 6$ h.

<table>
<thead>
<tr>
<th>$\Delta \lambda$</th>
<th>$\Delta \phi$</th>
<th>$\Delta h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test for $Y_{ij}$</td>
<td>21, [0.7%]</td>
<td>35, [1.2%]</td>
</tr>
<tr>
<td>Test for $Y_{ij, \lambda}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Test for $Y_{ij, \phi}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$N$</td>
<td>2855</td>
<td>2841</td>
</tr>
<tr>
<td>$B$</td>
<td>192</td>
<td>192</td>
</tr>
<tr>
<td>$a$</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

The application of PERG2FH method of VC analysis yields the following results:

For coordinate differences in longitude $\Delta \lambda$, \(^{11)}\):

\[
N = 2855, \quad B = 192, \quad a = 4, \\
SSE = 53373, \quad SSB = 71335, \quad SSA = 11253, \\
N - B = 2663, \quad B - a = 188, \quad a - 1 = 3, \\
m^2_{\lambda} = 20.04, \quad m^2_{\phi} = 379.4, \quad m^2_h = 3751; \text{ - in mm}^2, \\
A_{\lambda} = 713.75, \quad A_{\phi} = 14.89, \quad B_{\lambda} = 14.87, \\
m^2_{e,\lambda} = 20.04, \quad m^2_{e,\phi} = 24.17, \quad m^2_{e, h} = 4.72; \text{ - in mm}^2, \\
F_{\lambda} = 12.05 > 1.37 = F_{0.999;188,2663}.
\]

- accepted $H_{\lambda}$ and estimate $m^2_{\lambda}$,

\[
F_{\phi} = 9.89 > 5.65 = F_{0.999;3,188}
\]

- accepted $H_{\phi}$ and estimate $m^2_{\phi}$.

In the same way for both of other differences – latitude and altitude – the VC’s were obtained:

For $\Delta \phi$:

\[
m^2_{\phi,\lambda} = 20.04, \quad m^2_{\phi,\phi} = 183.6, \quad m^2_{\phi, h} = 126.9, \\
For \Delta h
\]

\[
m^2_{\lambda, h} = 31.49, \quad m^2_{\phi, h} = 39.01, \quad m^2_{h, h} = 9.20.
\]

The final results are presented in Table 2 (the variance estimate $m^2_{a} = \frac{12}{5} = m^2_{\lambda} + m^2_{\phi} + m^2_{h}$ is given).

### Table 2. ANOVA results survey at $I_a = 6$ h.

<table>
<thead>
<tr>
<th>$m_e$ [mm]</th>
<th>$m_f$ [mm]</th>
<th>$m_g$ [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>4.5</td>
<td>4.9</td>
</tr>
<tr>
<td>5.6</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>4.5</td>
<td>4.5</td>
<td>13.5</td>
</tr>
</tbody>
</table>

\(^{11)}\) After the gross error test in $i$; (in $\beta$ and $\alpha$ there were no outliers)

The relative error $r_{a} = m_{a} / b$ of the base vector of length $b$ is the Scale factor variance, i.e. the variance of the Tropo-Iono influence $m_{T-I}$ on the coordinate differences of the base vector. So it is:

\[
r_{a,\lambda} = 0.3 \cdot 10^{-7} = m_{T-I,\lambda}, \quad r_{a,\phi} = 0.4 \cdot 10^{-7} = m_{T-I,\phi}, \quad r_{a, h} = 1.4 \cdot 10^{-7} = m_{T-I, h}.
\]

Since the average number of satellites is 7.3, then the Tropo-Iono „average” variations of the base vector per satellite will be approximately $\sqrt{7.3}$ times higher, i.e. decomposed in the influences on the coordinate differences will be:

\[
m_{T-I,\lambda} = 0.7 \cdot 10^{-7}, \quad m_{T-I,\phi} = 1.0 \cdot 10^{-7} \quad \text{and} \quad m_{T-I, h} = 3.7 \cdot 10^{-7}.
\]

Now the Tropo-Iono „average” variations which concern the pseudorange – from one measured point to a satellite – decomposed in the influences on the coordinate differences will be $\sqrt{2}$ times lower:

\[
m_{T-I,\lambda} = 0.5 \cdot 10^{-7}, \quad m_{T-I,\phi} = 0.7 \cdot 10^{-7} \quad \text{and} \quad m_{T-I, h} = 2.6 \cdot 10^{-7}.
\]

Since these are projections of the Tropo-Iono influence variance $m_{T-I}$ of the pseudorange onto the coordinate axes $\lambda, \phi$ and $h$ and it will be:

\[
m^2_{T-I} = m^2_{T-I,\lambda} + m^2_{T-I,\phi} + m^2_{T-I, h}, \quad \text{so} \quad m_{T-I} = 2.7 \cdot 10^{-7},
\]

which, in fact, is the standard Tropo-Iono variation on the pseudorange (per satellite) for the mean value of the measuring result within time interval $I_a = 6$ hours.

Rigorously, this is valid for the base vector Bor-Kruševac of 83 km length. However, since the local influences on the Tropo-Iono variations are negligible, then this result can concern any vector of the same length, but under the same or very similar Tropo-Iono influences.

In another interval division: $I_{a} = 7.5 \text{ minutes}$ for $\beta$ and $I_{a} = 3 \text{ hours}$ for $\alpha$, the following results are obtained – Table 3.

### Table 3. ANOVA results survey at $I_a = 3$ h.

<table>
<thead>
<tr>
<th>$m_e$ [mm]</th>
<th>2.2</th>
<th>3.0</th>
<th>11.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_f$ [mm]</td>
<td>8.9</td>
<td>7.0</td>
<td>18.2</td>
</tr>
<tr>
<td>$m_g$ [mm]</td>
<td>2663</td>
<td>2649</td>
<td>2642</td>
</tr>
<tr>
<td>$f_{\beta}$</td>
<td>128.63</td>
<td>169.02</td>
<td>184.28</td>
</tr>
<tr>
<td>$f_{\alpha}$</td>
<td>2.42</td>
<td>2.51</td>
<td>2.82</td>
</tr>
</tbody>
</table>
A good agreement between the results for the 3-hour and 6-hour interval divisions is obvious.

The elements of the VC matrix are calculated following Searle et al. (1992) – Section F.2, so that for the VC correlation coefficients $\Delta \lambda$ the following values are obtained:

$ r_{m_{\alpha};m_{\beta}} = -0.04, r_{m_{\alpha};m_{\varepsilon}} = 0, $ and $ r_{m_{\varepsilon};m_{\beta}} = -0.01. $

For $\Delta \phi$ approximately the same values are obtained for the corresponding VC coefficients, whereas the correlation coefficients for $h$ are:

$ r_{m_{\alpha};m_{\beta}} = -0.02, r_{m_{\alpha};m_{\varepsilon}} = 0 $ and $ r_{m_{\varepsilon};m_{\beta}} = 0.00. $

The results of the correlation matrix confirm the assumptions about the mutual independence between the $\alpha, \beta$ factors and between them and the noise $\varepsilon$ – assumptions (12.2.4b). In this way the assumptions about agreement of the 2-way hierarchical classification model with GPS measurements results are confirmed, i.e. the justification of introducing this model is confirmed.

Example 2. The base vector $\text{Bar} - \text{Podgorica} (B-P)$ of 40 km length. For the base vector between the permanent GPS stations $\text{Bar} - \text{Podgorica} (B-P)$, of 40 km length, during three years, 2008, 2009, and 2010, fixed solutions for regularly spaced registrations of 30 seconds are found, yielding a total of about 3 100 000 measurements. In Fig. 2. (taken from: Perović 2015) the coordinate difference $\Delta \phi$ are presented the standards $\sigma = \sigma_{\varepsilon}, \sigma_{\beta}$ and $\sigma_{\alpha}$ from the VC PRG2FH ANOVA analysis, within intervals of 5 minutes for the quasistationary blocks $\beta$ and of 2 hours for the Tropo-Iono influences $\alpha$. In the figure the plot of the total error standard $\sigma_{\Delta}$ is also presented. The plot points follow the months so that the seasonal influences are clearly visible with a maximum in the summer and a minimum in the winter. The annual mean values of the standards are:

$\sigma = 7.5 \text{ mm}, \quad \sigma_{\beta} = 6.1 \text{ mm}, \quad \sigma_{\alpha} = 4.6 \text{ mm}$

and $\sigma_{\Delta} = 10.7 \text{ mm},$
so that the standard errors of the air light speed (of a single measurement) is:

$$v = \frac{0.4 \text{ mm}}{4000 \text{ mm}} \approx 1 \times 10^{-7},$$

which is in a good agreement with the results obtained by other scientists (see Introduction). The standard values determined for months of the year during the three years, from 2008 to 2010, are within the limits:

$$5.8 \text{ mm} < \sigma < 8.5 \text{ mm},$$

$$4.6 \text{ mm} < \sigma_{\beta} < 8.2 \text{ mm},$$

$$2.7 \text{ mm} < \sigma_{\alpha} < 7.1 \text{ mm}$$

and

$$7.9 \text{ mm} < \sigma_{\Delta} < 13.3 \text{ mm}.$$  
In this analysis at first the results containing gross errors are removed. Here for the $\varepsilon$ errors the $3\sigma_{\varepsilon}$ criterion within intervals of $\beta$ is applied.

Example 3. Base Vector $\text{Kragujevac} - \text{Batočina} (K-B); 20 \text{ km}$. For the base vector Kragujevac - Batočina of 20 km, between the AGROS Network points, 8640 three – days GPS measurements - for regularly spaced registrations of 30 seconds are analysed. The variance components are analysed for two interval widths in the factors:

- first, with $I_{\alpha} = 6 \text{ h} \quad$ and $I_{\beta} = 7.5 \text{ min},$ and
- second, with $I_{\alpha} = 2 \text{ h} \quad$ and $I_{\beta} = 7.5 \text{ min}.$

The results of VC analysis for the K-B vector are given in Table 12.6.4.

| $m_{\varepsilon} \quad [\text{mm}]$ | 5.5 | 5.5 | 8.0 | 8.0 | 14.3 | 14.3 |
| $m_{\beta} \quad [\text{mm}]$ | 5.2 | 4.7 | 6.2 | 5.8 | 14.8 | 12.3 |
\begin{align*}
m_{\alpha} \text{ [mm]} & = 3.4, 2.7, 2.3, 3.1, 16.4, 12.1, \\
m_{\beta} \text{ [mm]} & = 8.3, 7.7, 10.4, 10.4, 26.3, 22.4.
\end{align*}

On the basis of Table 4 we infer that the variance estimate of the factor \( \alpha \) for different interval widths is stable.

For the VC correlation coefficients for each of the three coordinate differences \( \Delta \lambda \), \( \Delta \varphi \) and \( \Delta h \), the same values are obtained:

\[
r_{m_{\alpha}^2,m_{\beta}^2} = -0.02, \quad r_{m_{\beta}^2,m_{\beta}^2} = 0, \quad \text{and} \quad r_{m_{\alpha}^2,m_{\alpha}^2} = -0.02.
\]

Thus, a correlation between any pair of VC practically does not exist. This is another confirmation that the VC analysis model is correctly introduced.

**Example 4. Base Vector Beograd (BGD): 2 km.** For the base vector BGD of 20 km, between the AGROS Network points, 7475 three-days GPS measurements – for regularly spaced registrations of 30 seconds are analysed.

Here VC are also analysed for two interval widths in the factors:

- first, with \( I_{\alpha} = 6 \text{ h} \) and \( I_{\beta} = 7.5 \text{ min} \), and
- second, with \( I_{\alpha} = 2 \text{ h} \) and \( I_{\beta} = 7.5 \text{ min} \).

The results of VC analysis for the BGD vector are given in Table 5.

Also for such a short vector – 2 km length – from Table 5 we infer that the variance estimate of the factor \( \beta \) for different interval widths is stable.

For the VC correlation coefficients for each of the three coordinate differences \( \Delta \lambda \), \( \Delta \varphi \) and \( \Delta h \), the same values are obtained

\[
r_{m_{\alpha}^2,m_{\beta}^2} = -0.03, \quad r_{m_{\alpha}^2,m_{\beta}^2} = 0, \quad \text{and} \quad r_{m_{\alpha}^2,m_{\alpha}^2} = -0.02,
\]

so that also for this vector on the basis of ANOVA we infer that a correlation between any VC pair practically does not exist. This is another confirmation that the VC analysis model is correctly introduced.

**Table 5. ANOVA results for \( I_{\beta} = 7.5 \text{ min} \).**

<table>
<thead>
<tr>
<th>( I_{\alpha} )</th>
<th>For ( \Delta \lambda )</th>
<th>For ( \Delta \varphi )</th>
<th>For ( \Delta h )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 h</td>
<td>2.2</td>
<td>2.2</td>
<td>3.3</td>
</tr>
<tr>
<td>2 h</td>
<td>3.3</td>
<td>3.3</td>
<td>5.4</td>
</tr>
<tr>
<td>6 h</td>
<td>5.4</td>
<td>5.4</td>
<td>5.4</td>
</tr>
</tbody>
</table>

6. Conclusions

On the basis of the obtained results, PERG2FH VC ANOVA applied to the GPS measurements of base vectors 2 km, 20 km, 40 km and 80 km long, we can draw the following conclusions and questions.

**A) General conclusions.** The results of GPS measurements agree with the proposed 2-way hierarchical model with random effects. This is confirmed on the basis of the following results where PERG2FH ANOVA is applied to the GPS measurements of base vectors from 2 to 80 km long:

- The variance components are not correlated, which is in agreement with assumptions in defining the GPS results by using 2-way hierarchical model with random effects.
- The variance components are practically insensitive to the interval choice, which is confirmed by close VC values obtained for different interval widths.
- For all vectors from 2 km to 80 km existence of VC of GPS measurements from both vector kinds – from quasistationary atmospheric blocks influences (factor \( \beta \)) and Tropo-Iono influences (faktor \( \alpha \)) – is confirmed.

**B) Special conclusions.** In the special conclusions we can emphasize the following.

- It is necessary to examine optimal interval widths, \( I_{\alpha} \) and \( I_{\beta} \).
- The research of seasonal VC should be continued.
- The study of VC dependence on the base vector length should be continued.
- VC should be studied specially for pseudoranges – in determination of coordinates of a single point.
- It would be of importance to establish standards in this field.

7. Acknowledgements

The Republic Geodetic Authority of Serbia provided the measurement data at permanent stations for the base vector Bor – Kruševac 83 km in length.

Prof. Dr Dragan Blagojević obtained the fixed solutions for the base vector Bor – Kruševac.

MsC Darko Andić calculated VC for the base vector B-P.
8. References


[11] Perović, G. 2009: "Principle of Determining the Influences of Time Variable Errors on the Baseline Vectors". Paper presented at the International Conference on Geodesy, Cartography and Cadastre in the 21th Century which was been held within framework of celebrating the 230th Anniversary of Moscow State University of Geodesy and Cartography on May 25-27, 2009, Moscow,

