

Robust Variance Components Estimation in the PERG Mixed Distributions of Empirical Variances—PEROBVC Method

Gligorije Perović

Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia
Email: perg@grf.bg.ac.rs

How to cite this paper: Perović, G. (2020) Robust Variance Components Estimation in the PERG Mixed Distributions of Empirical Variances—PEROBVC Method. *Open Journal of Statistics*, 10, 640-650.
<https://doi.org/10.4236/ojs.2020.104038>

Received: February 8, 2020

Accepted: August 1, 2020

Published: August 4, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A mixed distribution of empirical variances, composed of two distributions the basic and contaminating ones, and referred to as PERG mixed distribution of empirical variances, is considered. In the paper a robust inverse problem solution is given, namely a (new) robust method for estimation of variances of both distributions—PEROBVC Method, as well as the estimates for the numbers of observations for both distributions and, in this way also the estimate of contamination degree.

Keywords

Non-Homogeneous Sets of Empirical Variances, PERG Mixed Distribution of Empirical Variances, Robust Variance Components Estimation—PEROBVC Method

1. Introduction

A detailed presentation of Variance and Covariance Components Models can be found in Rao and Kleffe [1], as well as in Krishnaiah [2].

However, in the applications deviations appear from the assumptions under which these solutions are obtained, which also leads to negative estimates of variance components. The main source of these deviations is the contaminating distribution of observations (also presence of outliers), which, as a rule, always appears. Therefore, Variance Components (VC) solutions are always studied, which are less sensitive or insensitive to these deviations. Such properties are characteristic for robust solutions, out of which the following ones are here presented:

Maximum Likelihood Approaches to Variance Component Estimation and to

Related Problems [3]; Robust Estimation of Variance Components [4], Estimation and Use of Variance Components [5], Robust Variance Estimation in Linear Regression [6], Robust Variance Estimation for Random Effects Meta-analysis [7], Robust Variance Estimation in Meta-regression with Dependent Effect Size Estimates [8], Robust Estimation of Linear Mixed Models [9].

Robust estimates of variances and of their components have been continually applied in various fields [10]-[21] (a more detailed presentation can be seen in [22]).

Regarding a GPS measurement error as a random process and modeling the process realization structurally as a 2-way hierarchical classification with random effects, Perović [23] studied a robust solution for variance components in GPS measurements—PERG2FH method. Perović [22] also robustly estimated the parameters for both distributions of normally distributed observations that form Tukey's mixed distribution.

From the said above it is seen that robust methods for estimation of Variance Components are studied. However, studies of Variance Components estimates on the basis of empirical variances distributions, except the standard procedure, have not been undertaken. For this reason the present author has devoted himself to VC research on the basis of these distributions, so that the here presented method was invented.

In the present paper a mixed distribution composed of two distributions of empirical variances is considered, where one of them is basic and the other one is contaminating, with an essential assumption that the variances of both distributions are obtained on the basis of primary normal distributions, as well as that they have the same degrees of freedom. The present author names this distribution PERG mixed-distribution of empirical variances (PERG—from initial letters of the author's names: Perović Gligorije), which is presented in **Figure 1** and described by means of the following model

$$F(s^2) = (1 - \varepsilon)F_1(s^2) + \varepsilon F_2(s^2), \quad (0 < \varepsilon < 0.5) \quad (1)$$

where F_1 is the basic distribution, F_2 the contaminating distribution, both of empirical variances s^2 with f d.f. and ε ($0 < \varepsilon < 0.5$) the contaminating degree.

On the basis of such a composition of the mixed distribution of empirical variances the author obtained a robust solution named Perović's Robust Variance Components Estimation Model—PEROBVC¹ method for variances and for observation numbers for both distributions. In this way the method eliminates the influences of gross errors and outliers on the estimates of distribution parameters.

The key difference between the PEROBVC method and the existing ones is that distribution censoring is not used, but instead on the basis of structural decomposition in two distributions, except obtaining parameter estimates of the two distributions, the question of outliers is simultaneously solved.

Consequently, the PEROBVC method has the following properties:

- 1) Unbiased (exact) variance estimates for the basic distribution, as the most

¹PEROBVC is an abbreviation of the initial letters: Perović's Robust Variance Components Estimation Model.

important property,

- 2) Unbiased (exact) variance estimates for the contaminating distribution,
- 3) Estimates of observation numbers for both distributions and in this way percentage estimates concerning fractions of basic and contaminating distributions in the mixed one.

The correctness of the PEROBVC method has been verified on exact (expected) values of some quantities from the mixed PERG distribution, as well as on examples of measurements in 2D control geodetic networks.

The variance estimate for the basic distribution obtained by applying the PEROBVC method is compared to its ML estimate.

The structure of the further presentation is the following: 1) establishing the PEROBVC method, 2) way of solving the formulated problem, 3) presentation of robust ML estimation method, 4) correctness verification for PEROBVC method by comparing its solution to the ML one and 5) verification on examples of geodetic measurements in to 2D control networks.

2. Basis of PEROBVC Method

The idea of PEROBVC method has been presented in [24] (Section 32.14).

Consider a PERG mixed distribution (Figure 1) of empirical variances $s_i'^2, i = 1, 2, \dots, n'$ and $s_j''^2, j = 1, 2, \dots, n''$

$$F(s^2) = (1 - \varepsilon)F(s'^2) + \varepsilon F(s''^2), \quad (0 < \varepsilon < 0.5) \tag{2}$$

with $n' + n'' = n$ and with the same degrees of freedom, $f_i' = f_j'' \equiv f$. The variances are obtained from primary normally distributed observations with expectations

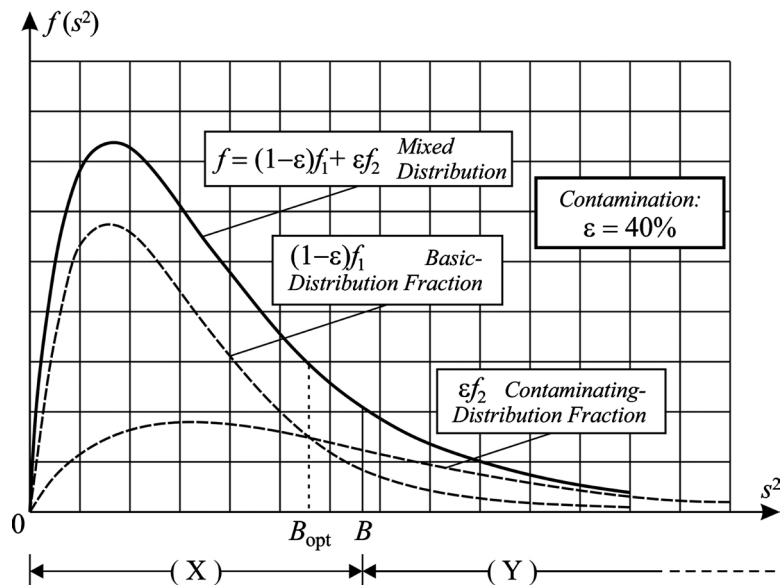


Figure 1. Density plot for PERG mixed distribution $f(s^2)$ of empirical variances with 4 d. f. ($f = 4$); with $f_1(s'^2)$ is the basic with σ_1^2 and $f_2(s''^2)$ is the contaminating distribution with σ_2^2 (here $\sigma_2^2 = 2\sigma_1^2$) and contaminating degree $\varepsilon = 0.40$.

$$\mathbf{E}[s_i'^2] = \sigma_1^2, i = 1, 2, \dots, n', \quad \mathbf{E}[s_j''^2] = \sigma_2^2, j = 1, 2, \dots, n''; \quad \sigma_2^2 > \sigma_1^2,$$

where $F(s'^2)$ and $F(s''^2)$ are the distribution functions of the empirical variances s'^2 and s''^2 . The task is to estimate the variances σ_1^2 and σ_2^2 of both distributions on the basis of the mixed distribution with f known. In addition to the solving of this task, the method also yields estimates of the numbers of observations n' and n'' for both component distributions $F(s'^2)$ and $F(s''^2)$, *i.e.* the contaminating degree is estimated.

The point B divides the domain of s^2 into two sub-domains X and Y (**Figure 1**) and let be:

n' and n'' — numbers of all observations, (s'^2 and s''^2), within X and Y ;
 n'_X and n''_X — numbers of observations for s'^2 and s''^2 within X ;
 n'_Y and n''_Y — numbers of observations for s'^2 and s''^2 within Y ,

$$\left. \begin{aligned} n_X + n_Y &= n & n'_X + n''_X &= n_X \\ n' + n'' &= n & n'_Y + n''_Y &= n_Y \end{aligned} \right\}, \quad (3)$$

The point B will be chosen in such a way that within the interval $(0, B)$ the observations s'^2 dominate, whereas within $[B, \infty)$ the observations s''^2 dominate from where it follow that it must be $B = B_{\text{opt}}$ (also see **Figure 1**).

Here the point B is defined according as

$$B = s_{(n_X)}^2 + d/2, \quad (4)$$

where d is the width of the rounding interval for observations s^2 .

For the sake of simplifying the following designations are introduced:

$$\left. \begin{aligned} \text{(a)} \quad w = s^2 \text{ with } f \text{ d.f.} &\Rightarrow w' = s'^2, \quad w'' = s''^2 \\ \text{(b)} \quad y = f \frac{s^2}{\sigma^2} = \chi_f^2 &\Rightarrow y' = f \frac{s'^2}{\sigma_1^2}, \quad y'' = f \frac{s''^2}{\sigma_2^2} \end{aligned} \right\} \quad (5)$$

Then the density of the random variable s^2 -distribution can be written as

$$f(s^2) = f(w) = k_f^{-1} \sigma^{-f} f^{f/2} w^{f/2-1} e^{-wf/2\sigma^2} = \frac{f}{\sigma^2} f(y), \quad k_f = 2^{f/2} \Gamma\left(\frac{f}{2}\right) \quad (6)$$

where $\Gamma(f/2)$ is the gamma function of $f/2$. Then

$$F(s^2) = F(w) = \int_0^w f(w) dw = \int_0^y f(y) dy = F(y) \quad (7)$$

$$wf(w)dw = \frac{\sigma^2}{f} yf(y)dy \Rightarrow \int_0^w wf(w)dw = \frac{\sigma^2}{f} \int_0^y yf(y)dy. \quad (8)$$

From the equality condition for the ordinates of the two distribution densities, *i.e.* from the condition $(1 - \varepsilon) f(w') = \varepsilon f(w'')$, one obtains optimal point $B(B_{\text{opt}})$

$$B_{\text{opt}} = \frac{\frac{2}{f} \ln \left(\frac{n' \left(\frac{\sigma_2}{\sigma_1} \right)^f}{n'' \left(\frac{\sigma_1}{\sigma_2} \right)^f} \right)}{\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}}. \quad (9)$$

3. PEROBVC Solution

The estimates of the variance components for both distributions are obtained from the event-probability maximum

$$\left[D'_{(1)} \wedge \dots \wedge D'_{(n'_X)} \right] \wedge (s^{n^2} < B) \wedge (s'^2 > B) \wedge \left[D''_{(n''_X+1)} \wedge \dots \wedge D''_{(n'')} \right]$$

where $D'_{(1)} = s'^2 < s'^2 < s'^2 + ds'^2 = w'_{(1)} < w' < w'_{(1)} + dw'$, ..., $D''_{(n'')} = s''^{n^2} < s''^{n^2} < s''^{n^2} + ds''^{n^2} = w''_{(n'')} < w'' < w''_{(n'')} + dw''$, whose likelihood function, to the proportionality constant k , is

$$L = k \cdot \prod_X f(w'_{(i)}) \cdot p(w'' < B) \cdot p(w' > B) \cdot \prod_Y f(w''_{(i)}),$$

where: $\prod_X f(w'_{(i)}) = \prod_{i=1}^{n'_X} f(w'_{(i)})$, $\prod_Y f(w''_{(i)}) = \prod_{i=n'_X+1}^{n''} f(w''_{(i)})$.

In view of (5)-(8) one finds:

$$\left. \begin{aligned} f'_B &= k_f^{-1} y_B'^{f/2-1} e^{-y_B'/2} & f''_B &= k_f^{-1} y_B''^{f/2-1} e^{-y_B''/2} \\ y'_B &= \frac{f \cdot B}{\sigma_1^2} & y''_B &= \frac{f \cdot B}{\sigma_2^2} \\ F'_B &= \int_0^{y'_B} f(y') dy' & F''_B &= \int_0^{y''_B} f(y'') dy'' \\ A'_B &= \int_0^{y'_B} \frac{1}{f} y' f(y') dy' & A''_B &= \int_0^{y''_B} \frac{1}{f} y'' f(y'') dy'' \end{aligned} \right\}. \tag{10}$$

Furthermore $\ln L$ will be used there is:

$$\ln L = c + \sum_X \ln f(w'_{(i)}) + \ln (F''_B)^{n'_X} + \ln (1 - F'_B)^{n''} + \sum_Y \ln f(w''_{(i)}), \tag{11}$$

where c is a constant.

According to the asymptotic theory it is:

$$\sum_X w''_i \xrightarrow[n'' \rightarrow \infty]{p} n''_X \int_0^B w'' f(w'') dw'', \tag{12a}$$

$$\sum_Y w'_i \xrightarrow[n' \rightarrow \infty]{p} n'_Y \int_0^B w' f(w') dw' = n'_Y \sigma_1^2 (1 - A'_B). \tag{12b}$$

Bearing in mind (10)-(12) and introducing the substitutions

$$\left. \begin{aligned} \text{(a)} \quad \sum_X s_i'^2 &= \sum_X s_i^2 - \sum_X s_i^{n^2}, \quad \sum_Y s_i''^2 = \sum_Y s_i^2 - \sum_Y s_i'^2 \\ \text{(b)} \quad \sum_X s_i^{n^2} &= n'' \int_X s^{n^2} f(s^{n^2}) ds^{n^2} = n'' \sigma_2^2 A''_B \\ \text{(c)} \quad \sum_Y s_i'^2 &= n' \int_Y s'^2 f(s'^2) ds'^2 = n' \sigma_1^2 (1 - A'_B) \end{aligned} \right\}$$

from the necessary maximum condition: (a) $\partial \ln L / \partial \sigma_1^2 = 0$ and (b) $\partial \ln L / \partial \sigma_2^2 = 0$ we get the equations

$$\left. \begin{aligned} \text{(a)} \quad \frac{-n'_X}{2\sigma_1^2} + \frac{1}{2\sigma_1^4} \sum_X s_i^2 - \frac{n''_X \sigma_2^2}{2\sigma_1^4} A''_B + \frac{B n'_Y f'_B}{\sigma_1^4 (1 - F'_B)} &= 0 \\ \text{(b)} \quad \frac{-n'_Y}{2\sigma_2^2} + \frac{1}{2\sigma_2^4} \sum_Y s_i^2 - \frac{B n''_X f''_B}{\sigma_2^4 F''_B} &= 0 \end{aligned} \right\}$$

Their solution is given as simple iterations

$$\sigma_{1,k+1}^2 = \frac{1}{n'_{X,k}} \left[\sum_X s_i^2 - n''_{X,k} \sigma_{2,k}^2 A''_{B,k} + 2B \frac{n'_{Y,k} f'_{B,k}}{1 - F'_{B,k}} \right], \tag{13a}$$

$$\sigma_{2,k+1}^2 = \frac{1}{n_{Y,k}''} \left[\sum_Y s_i^2 - n'_k \sigma_{1,k}^2 (1 - A'_{B,k}) - 2B \frac{n_{X,k}'' f_{B,k}''}{F_{B,k}''} \right]. \quad (13b)$$

where the numbers n'_X , n'_Y , n''_X and n''_Y are determined from the expressions

$$\left. \begin{array}{l} \text{(a) } n'_X = n' F'_B, \quad n'_Y = n' - n'_X \\ \text{(b) } n''_X = n'' F''_B, \quad n''_Y = n'' - n''_X \end{array} \right\}, \quad (14)$$

whereas n' and n'' are determined as

$$\left. \begin{array}{l} F'_B n' + F''_B n'' = n_X \\ (1 - F'_B) n' + (1 - F''_B) n'' = n_Y \end{array} \right\} \quad (15)$$

Note 1. The initial values for the variance components can be assumed, for example, as:

$$\sigma_{1,0}^2 = 0.6 \cdot \left(\frac{1}{n} \sum_1^n s_i^2 \right) \quad \text{and} \quad \sigma_{2,0}^2 = 3 \cdot \left(\frac{1}{n} \sum_1^n s_i^2 \right).$$

If the model assumptions are not satisfied, for example, if the distribution instead of a long tail has a shortened tail (from the right), then for one of the variances we obtain negative estimates. Δ

Note 2. Blunders in the observations (measurements with large gross errors) must be rejected before applying the PEROBVC method. For example, first find

$$s^2 = \frac{1}{n} \sum_1^n s_i^2 \quad \text{and then} \quad s_{0,9999}^2 = \frac{1}{f} \chi_{f;0,9999}^2 \cdot s^2.$$

Those s_i^2 exceeding $s_{0,9999}^2$ are rejected, whereas this procedure is repeated once more with non-rejected results. Δ

The advantages of the PEROBVC method are:

- 1) Unbiased estimators for σ_1^2 and σ_2^2 , if B are close to B_{opt} , and
- 2) Minimal variances for σ_1^2 .

The disadvantages of the PEROBVC method are:

1) A sensitivity to the choice of the point B , (point B must be close to B_{opt}), which can result in negative estimates for either of the variances σ_1^2 or σ_2^2 , or for both, and

2) Sensitivity to the choice of the initial values for the variances, σ_1^2 and σ_2^2 , which, also, can result in negative estimates for one or both variances.

The stopping criterion for the iterative process: Let $\mathbf{x}_k^T = [\sigma_{1,k}^2 \quad \sigma_{2,k}^2]$ be the vector of parameter estimates in the k -th iteration and \mathbf{d} the difference vector of these estimates from the $(k+1)$ -th and k -th iterations,

$$\mathbf{d}^T = \left[(\sigma_{1,k+1}^2 - \sigma_{1,k}^2) \quad (\sigma_{2,k+1}^2 - \sigma_{2,k}^2) \right], \quad \text{then the iterations should be stopped if} \\ \|\mathbf{d}\| < 10^{-q} \|\mathbf{x}_k\|, \quad q \in \{5, 6, 7, 8, 9\}. \quad (16)$$

4. Some Robust Estimation

In this case we shall use the method of Maximum Likelihood (ML).

The Robust Maximum Likelihood (ML) Estimation

The ML method is based on the assumption that in the domain X there exists only variances s'^2 ($s'^2 \equiv s^2$; with $E[s^2] = \sigma^2$) of the basic distribution, whereas in the Y interval in addition to basic variances exist the variances of contaminating distributions. Therefore the point B in **Figure 1** will be the censoring point.

The ML solution is obtained from the probability maximum for the event

$$\left[D_{(1)} \wedge \dots \wedge D_{(n_X)} \right] \wedge (s^2 > B),$$

where $D_{(i)} = s_{(i)}^2 < s^2 < s_{(i)}^2 + ds^2$, $i = 1, \dots, n_X$, one obtains the likelihood function (with proportionality constant k):

$$L = k \prod_{i=1}^{n_X} f(s_i^2) [1 - F(B)]^{n_Y}.$$

Hence it follows

$$\ln L \propto \sum_X \ln f(s_i^2) + n_Y \ln [1 - F(B)].$$

From the necessary condition $\partial \ln L / \partial \sigma_1^2 = 0$ one obtains

$$-n_X \sigma^2 + \sum_X s_i^2 + 2B n_Y \frac{f(y_B)}{1 - F(y_B)} = 0.$$

Their solution is given as simple iterations

$$\sigma_{k+1}^2 = \frac{1}{n_X} \sum_X s_i^2 + 2B \frac{n_Y}{n_X} \cdot \frac{f(y_{B,k})}{1 - F(y_{B,k})}, \quad (17)$$

where

$$y_{B,k} = \frac{f \cdot B}{\sigma_k^2}, \quad F_{B,k} = \int_0^{y_{B,k}} f(y) dy, \quad \left(y = f \frac{s^2}{\sigma^2} = \chi_f^2, f(y) = k_f^{-1} y^{f/2-1} e^{-y/2} \right). \quad (18)$$

The stopping criterion for the iterative process can be based on the parameter difference $r_\sigma = (\sigma_{k+1}^2 - \sigma_k^2) / \sigma_{k+1}^2$ between two iterations:

$$|r_\sigma| < 10^{-c}, \quad c \in \{4, 5, 6, 7, 8\}, \quad (19)$$

5. Results and Discussion

Example 1. By using the functions F-RAN1 and F-RAN2 [25], one simulates 10,000, ($n = 10,000$), samples (series) of size 4, $f = 3$, with F-RAN1 5000 samples are simulated ($n' = 5000$), $\sigma_1^2 = 1$ and the variance estimate (from these 5000 samples) $s_1^2 = 0.9997$, with F-RAN2 also 5000 samples ($n'' = 5000$), of the same size, with $\sigma_2^2 = 2$ and variance estimate (from these 5000 samples) $s_2^2 = 1.9886$. Then a PERG mixed distribution (2), is formed ($\varepsilon = 50\%$).

Since the variances σ_1^2 and σ_2^2 are known then, according to (6), one can also find, B optimal, obtaining $B_{\text{opt}} = 2.772589$.

In **Table 1**, for 4 partitions (X, Y), we present the PEROBVC estimates of the variances σ_1^2 and σ_2^2 , as well as the contamination percentage ε , ($\varepsilon = n''/n$).

From **Table 1** it is easily noticed that the variance components, as well as the

numbers of observations for both distributions, are well estimated, *i.e.* their values are very close to the real ones. It is also noticed that the values of the ML estimates for σ^2 are smaller than those of the PEROBVC ones for σ_1^2 (where it is $\sigma^2 \equiv \sigma_1^2$). ▲

Example 2. In the 2D geodetic control network of object “TUZLA” [26] from the measurements of 328 ($n = 328$), horizontal angles, with a theodolite Wild T3, in 6 quick locks, in stable and unstable conditions, a mixed distribution of empirical variances (in [“2”]), with $f = 5$ d.f., is obtained (Table 2).

In Table 3, for 7 partitions (X, Y), we present the estimates of the variances

Table 1. Results of PEROBVC and ML estimations in simulated PERG mixed distribution of empirical variances s^2 with $n = 10000$, $n' = 5000$, $n'' = 5000$ ($\varepsilon = 50\%$), $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $f = 3$ d.f., and $\sum_1^n s^2 = 14941.50$.

n_Y	%	B	$\sum_x s^2$	σ_1^2	σ_2^2	n'	n''	ε [%]	ML: σ_1^2
3990	40	1.3863	4002.870	1.0330	2.0326	5386.9	4613.1	46.1	0.8260
3000	30	1.7715	5557.883	0.9237	1.8834	4055.9	5944.1	59.4	0.8741
2500	25	1.9950	6494.014	1.0561	2.0624	5646.8	4353.2	43.5	0.9241
2000	20	2.2955	7562.874	1.0250	2.0138	5255.6	4744.4	47.4	0.9849
Mean:				1.0095	1.9981	5086.3	4913.7	49.1	0.9023

Table 2. Ordered grouped increasing sequence of 328 ($n = 328$) empirical variances, with $f = 5$ d.f., of horizontal angle measurements with a theodolite Wild T3 in the geodetic 2D control network of object “TUZLA” in epochs 1979, 1981 and 1984.

s^2	Freq.	s^2	Freq.	s^2	Freq.	s^2	Freq.	s^2	Freq.
0.05	0	1.65	3	3.25	4	4.95	5	7.75	1
0.15	1	1.75	8	3.35	1	5.05	1	7.95	1
0.25	4	1.85	10	3.45	3	5.15	1	8.05	1
0.35	6	1.95	10	3.55	3	5.35	1	8.35	1
0.45	6	2.05	11	3.65	2	5.45	1	8.55	1
0.55	8	2.15	5	3.75	6	5.55	2	9.05	1
0.65	7	2.25	12	3.85	3	5.65	1	9.35	1
0.75	15	2.35	6	3.95	2	5.75	1	9.65	1
0.85	8	2.45	7	4.05	0	5.85	3	10.45	1
0.95	16	2.55	6	4.15	2	6.05	1	11.25	1
1.05	12	2.65	4	4.25	1	6.25	3	11.45	1
1.15	11	2.75	1	4.35	1	6.75	1	12.75	1
1.25	17	2.85	7	4.45	2	7.15	1	13.75	1
1.35	15	2.95	4	4.55	1	7.35	1	16.85	1
1.45	11	3.05	5	4.65	2	7.45	2	Σ	n
1.55	15	3.15	3	4.75	1	7.65	1	818.30	328

Table 3. Results of PEROBVC and ML estimates of variances for PERG mixed distribution composed of 328 ($n = 328$), empirical variances of horizontal angles in the 2D control network of object “TUZLA”, in epochs 1979, 1981 and 1984; $f = 5$ d.f., and $\sum_{i=1}^n s_i^2 = 818.30$.

n_Y	%	B	$\sum_x s^2$	σ_1^2	σ_2^2	n'	n''	ε [%]	ML: σ^2
165	50.3	1.8	173.05	1.8788	6.8348	287.2	40.8	12	1.9915
134	41	2.1	233.60	1.9781	12.5460	312.0	16.0	5	1.4628
98	30	2.6	317.90	1.9157	3.1517	297.5	30.5	9	1.4930
82	25	3.0	363.00	1.7716	5.5333	264.9	63.1	19	1.5439
66	20	3.5	414.40	1.6918	5.0078	248.6	79.4	24	1.6216
50	15	4.0	474.30	1.7902	5.6048	267.4	60.6	18	1.7304
33	10	5.3	553.65	1.6501	4.8954	242.6	85.4	26	1.8842
Mean:				1.8109	6.9391	274.3	53.7	16	1.6753

σ_1^2 and σ_2^2 , as well as the contamination percentage ε ($\varepsilon = n''/n$).

According to an a priori analysis for the horizontal-angle measurements in stable conditions, following Činklović [27], one obtains $\sigma_1^2 = 1.848$ [“2], resulting in a very good agreement with 1.8109 [“2] obtained by using PEROBVC method based on the measurements.

Besides, on the basis of the results presented in **Table 3** one concludes about a very high stability of the estimator for the variance component σ_1^2 .

In Example 2 the average value from the ML estimates for σ^2 , ($\sigma^2 \equiv \sigma_1^2$), is also smaller than the average value of the PEROBVC estimates for σ_1^2 ($1.6753 < 1.8109$).

It should be noted that for the partition $n_Y = 165 > n_X = 163$ the numbers of observations n' and n'' are well estimated, although $B = 1.8$ is significantly smaller than $B_{opt} = 8.18$. ▲

Note 3. In Examples 1-2 the ML estimates for σ^2 are smaller than the corresponding PEROBVC estimates for σ_1^2 , ($\sigma^2 \equiv \sigma_1^2$), because higher values of empirical variances are in the end of the distribution so that they do not enter the ML estimate for σ^2 . ▲

On the basis of applications of the PEROBVC method in the treatment of real geodetic measurements, Example 2, the present author concludes that the variance estimator of the basic distribution σ_1^2 , has a good stability.

Generally, the PEROBVC method offers good estimates, especially for σ_1^2 , which is illustrated by Examples 1-2.

6. Conclusions

On the basis of the obtained results in Examples 1-2 we can conclude the following:

1) On the basis of exact (expected) values from Example 1 the validity of the PEROBVC method in the variance estimation, as well as the estimates of num-

bers of observations, for both distributions in the PERG mixed distribution of empirical variances is confirmed. Here the variance estimates for both distributions, basic and contaminating ones, are correct; *i.e.* their values are exact.

2) On the basis of realistic measurements for the horizontal angles in the geodetic 2D control network from Examples 2 good (satisfactory) parameter estimates for both distributions are also confirmed.

3) In Example 2 the variance estimate for the basic distribution is confirmed through the result of the a priori analysis.

Acknowledgements

Dr ZORICA Cvetković from the Astronomical Observatory in Belgrade is the author of the FORTRAN programmes who performed the calculations in the examples. The OBJEKTIV GEO D.O.O. Company from Belgrade (Serbia) has financially supported the publishing of the article.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Rao, C.R. and Kleffe, J. (1988) Estimation of Variance Components and Applications. North-Holland, Amsterdam.
- [2] Krishnaiah, P.R., (1984) Analysis of Variance. North-Holland, Amsterdam.
- [3] Harville, D.A. (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, **72**, 320-338. <https://doi.org/10.1080/01621459.1977.10480998>
- [4] Fellner, W.H. (1986) Robust Estimation of Variance Components. *Technometrics*, **28**, 51-60. <https://doi.org/10.1080/00401706.1986.10488097>
- [5] Robinson, D.L. (1987) Estimation and Use of Variance Components. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **36**, 3-14. <https://doi.org/10.2307/2988267>
- [6] Herbert, J.H. and Kott, Ph.S. (1988) Robust Variance Estimation in Linear Regression. *Journal of Applied Statistics*, **15**, 341-345. <https://doi.org/10.1080/02664768800000044>
- [7] Sidik, K. and Jonkman, J.N. (2006) Robust Variance Estimation for Random Effects Meta-Analysis. *Computational Statistics & Data Analysis*, **50**, 3681-3701. <https://doi.org/10.1016/j.csda.2005.07.019>
- [8] Hedges, L.V., Tipton, E. and Johnson, M.C. (2010) Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates. *Research Synthesis Methods*, **1**, 39-65. <https://doi.org/10.1002/jrsm.5>
- [9] Koller, M. (2013) Robust Estimation of Linear Mixed Models. PhD Thesis, ETH, Zürich.
- [10] Welsh, A.H. and Richardson, A.M. (1997) 13 Approaches to Robust Estimations of Mixed Models. In: *Handbook of Statistics*, Vol. 15, Elsevier Science, Amsterdam, 343-384. [https://doi.org/10.1016/S0169-7161\(97\)15015-5](https://doi.org/10.1016/S0169-7161(97)15015-5)
- [11] Amiri-Simkooei, A.R. (2007) Least-Squares Variance Component Estimation: Theory and GPS Applications. Ph.D. Thesis, Delft University of Technology, Delft.

- [12] Bartlett, J. (2014) The Robust Sandwich Variance Estimator for Linear Regression (Using R). The Stats Geek, February 14, 2014.
- [13] Arango-Castillo, L. and Takahara, G. (2018) Robust Estimation of the Sample Mean Variance for Gaussian Processes with Long-Range Dependence. 2017 *IEEE Global Conference on Signal and Information Processing*, Montreal, 14-16 November 2017, 201-205. <https://doi.org/10.1109/GlobalSIP.2017.8308632>
- [14] Chen, Y. and Jackson, D. (1995) Robust Estimation of Mean and Variance in Fisheries. *Transactions of the American Fisheries Society*, **124**, 401-412. [https://doi.org/10.1577/1548-8659\(1995\)124<0401:REOMAV>2.3.CO;2](https://doi.org/10.1577/1548-8659(1995)124<0401:REOMAV>2.3.CO;2)
- [15] Grossi, L. and Laurini, F. (2011) Robust Estimation of Efficient Mean-Variance Frontiers. *Advances in Data Analysis and Classification*, **5**, 3-22. <https://doi.org/10.1007/s11634-010-0082-3>
- [16] Li, Y., Li, J., Qi, J. and Chen, L. (2019) Robust Cubature Kalman Filter for Dynamic State Estimation of Synchronous Machines under Unknown Measurement Noise Statistics. *IEEE Access*, **7**, 29139-29148. <https://doi.org/10.1109/ACCESS.2019.2900228>
- [17] Singh, A. and Nocerino, J. (2002) Robust Estimation of Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations. *Chemometrics and Intelligent Laboratory Systems*, **60**, 69-86. [https://doi.org/10.1016/S0169-7439\(01\)00186-1](https://doi.org/10.1016/S0169-7439(01)00186-1)
- [18] Vilà-Valls, J., Wei, Q., Closas, P. and Fernández-Prades, C. (2014) Robust Gaussian Sum Filtering with Unknown Noise statistics: Application to Target Tracking. 2014 *IEEE Workshop on Statistical Signal Processing*, Gold Coast, 29 June-2 July 2014, 416-419. <https://doi.org/10.1109/SSP.2014.6884664>
- [19] Wang, Ch., Wang, Y., Lin, Z., Yuille, A.L. and Gao, W. (2014) Robust Estimation of 3D Human Poses from a Single Image. CBMM Memo No. 013. <https://cbmm.mit.edu/sites/default/files/publications/CBMM-Memo-013.pdf>
- [20] Zhou, W.-Z., Bose, K., Fan, J. and Liu, H. (2018) A New Perspective on Robust M-Estimation: Finite Sample Theory and Applications to Ddependence-Adjusted Multiple Testing. *The Annals of Statistics*, **46**, 1904-1931. <https://doi.org/10.1214/17-AOS1606>
- [21] Zhao, J. and Mili, L. (2018) A Framework for Robust Hybrid State Estimation with Unknown Measurement Noise Statistics. *IEEE Transactions on Industrial Informatics*, **14**, 1866-1875. <https://doi.org/10.1109/TII.2017.2764800>
- [22] Perović, G. (2019) Robust Estimation of the Normal-Distribution Parameters by Use of Structural Partitioning-Perobls D Method. *American Journal of Computational Mathematics*, **9**, 302-316. <https://doi.org/10.4236/ajcm.2019.94022>
- [23] Perović, G. (2016) Variance Components Analysis in GPS Measurements. *Imperial Journal of Interdisciplinary Research*, **2**, 357-367.
- [24] Perović, G. (2005) Least Squares. Author, Belgrade.
- [25] William, H.P., *et al.* (1986) Numerical Recipes. Cambridge Univ. Press, Cambridge.
- [26] "Tuzla" Object (1985) Elaborates of Geodetic Measurements in Epochs 1979, 1981 and 1984 for the Purposes of Ground—Deformation Determining. Salt Mine, Tuzla.
- [27] Činklović, N. (1978) Analysis and a Priori Accuracy Estimate for Precise Geodetic Measurements, Monograph. Faculty of Civil Engineering, Belgrade.