To the evaluation of the parameters of the regression equation between the radiometric and geological testing

G A Ivanyukovich1, O A Malafeyev2, I V Zaitseva3,4, A M Kovshov2, V V Zakharov3 and N I Zakharova5

1Institute of Earth Sciences, St. Petersburg State University, University Emb., 7/9, St. Petersburg, 199034, Russia
2St. Petersburg State University, University ave., 35, St. Petersburg, Peterhof, 198504, Russia
3Stavropol State Agrarian University, Zootekhnichesky Lane, 12, Stavropol, 355017, Russia
4Stavropol branch of the Moscow Pedagogical State University, 66G, Dovatortsev Street, Stavropol, 355042, Russia
5North Caucasus Federal University, Pushkin street, 1, Stavropol, 355017, Russia

E-mail: irina.zaitseva.stv@yandex.ru

Abstract. Comparisons of the regression equation between two methods for testing rocks and ores at the Dukat gold-silver deposit are made using two algorithms. In the first loss function defined by the sum of squares of residuals (deviations from the values of the dependent variable of the regression line), in the second - the sum of the absolute values of the residual normalized to sum values of the dependent and independent variables. As a result, it was shown that in the first case, in order to increase the accuracy of estimates of the coefficients, it is necessary to exclude observations in which the absolute value of the residuals normalized to their standard deviation exceeds the critical level of 5-7. In the second case, it is stable to sample formation and the number of observations can be reduced by 2-3 times. Under these conditions, the error in the estimates of the coefficients turns out to be close.

1. Introduction
For the analysis 270 samples of cuttings from drilling and blasting wells of the silver Ducat field were used. The silver content in the samples according to the assay analysis data varied from 20 to 10,000 and more g/t. Discrepancies between the main and repeated mining operations in silver content intervals of 50 - 239, 240 - 999, 1000 g/t and more were 76, 175 and 651 g/t, respectively.

The estimates of the regression equation between two methods of testing using two algorithms are compared. In the first one, the loss function determined the sum of squares of residues (deviations of the dependent variable values from the regression line), in the second one - the sum of absolute residues values normalized by the sum of values of dependent and independent variables.

It complicates the “selection” of regression equation coefficients between RRO and geological sampling data by the wide range of content and extremely uneven distribution of metal in ore zones. For these reasons, the mean square differences between the sampling methods are large and vary widely depending on the metal content. In order to increase the accuracy of regression parameter
estimates, it is necessary to increase the control sample, which makes sampling more expensive and decreases its efficiency with the variability of texture-structural features of ores.

As a result, it is shown that in the first case, in order to improve the accuracy of the coefficient estimates, it is necessary to exclude observations where the absolute value of the residues normalized for their standard deviation exceeds the critical level 5-7. In the second method it is stable to sample formation and the number of observations can be reduced by 2-3 times. Under these conditions the error of coefficients estimations is close.

2. Statement of the problem and the solution to the problem
X-ray radiometric testing (XRT) is widely used at deposits of rare and non-ferrous metals for the purpose [1]:

- delimitation of industrial ore bodies under conditions of their natural occurrence,
- operational testing of stockpiled commercial ore in bulk to clarify plans for its shipment to the concentrator,
- measurements of metal content in transport containers and on a belt of conveyors for rejection of substandard ores.

Analysts for RROs are calibrated by comparing the results with traditional geological testing (ET). This time-consuming operation involves taking a large number of samples and preparing them for analysis in the laboratory. It has to be performed repeatedly in case of changes in the mineralogical composition of ore bodies.

It is necessary to note features of many deposits of rare and precious metals, which complicate the “selection” of regression equation coefficients between the data of RRO and GO [2].

First, a wide range of content of the useful component. Thus, on the gold-silver birthplace of Dukat in the Magadan Region, the results of RRO and GO which are discussed below, of practical interest are ores with a silver content starting from 50 g/t, and the upper limit reaches 10000 g/t and more. As a consequence, the error of metal concentration determination changes in wide limits and this must be taken into account when calculating the regression equation.

Secondly, the distribution of metal in the ore zones is extremely uneven. For this reason, the average square differences between the results of the main and the re-test of the civil defense are large. Thus, at the Dukat deposit in the course of HC the discrepancies in the intervals of silver grade of 50 - 239, 240 - 999, 1000 g/t and more are 76, 175 and 651 g/t, respectively [1]. Another consequence of the unequal silver distribution is the prevalence of samples with low silver content. As a result, the number of samples has to increase, which significantly complicates and increases the cost of grading the analyzers.

Parameters of the regression equation between RRO and GO data

\[ GO^* = B_0 + B_1 \cdot RRO + B_2 \cdot RRO^2 + \ldots + B_k \cdot RRO^k, \]

usually found by the least squares method (LSM), according to which the regression equilibrium coefficients are determined by a condition [2,3]:

\[ \sum_{i=1}^{n} (GO_i - GO_i^*)^2 = \min. \]

The RRO variable is called a predictive variable, GO is a response variable, GO* is a forecast or estimate variable, and the difference between SWO and GO* is a residual variable.

The popularity of ISC is, to a certain extent, explained by the absence of specialized statistical programs in field exploration lots and at mining enterprises.

The method is effective if the deviations of observations from the regression line (residues) are constant throughout the whole content range. Otherwise, the fitting criterion will be determined by the samples with high element co-holding and the error of coefficients estimation increases. Therefore, in case of unequal measurements it is necessary to ascribe weights whose value is inversely proportional to the error of GO values and to calculate the coefficients estimations by the method of weighted least squares [3]. The relationship between weights and metal content can be established if repeated
observations are made at several RRO values [3]. This is a difficult procedure to implement when testing ores.

In a particular case, if the error increases in proportion to the content, the task is simplified. It is enough to execute transformation of a kind \( \log(GO) = GO \) and \( \log(RRO) = RRO \), and to apply MNK, using logarithms [4]. In practice it is necessary to face both an error of measurement of the maintenance which increases in process of increase in the maintenance of elements, and influence of non-uniformity of distribution of elements in ores which grows in process of decrease in the maintenance, and therefore the scope of application of this method is limited.

However, when considering LSM, it should be taken into account that in geology the distribution of elemental content, in most cases, does not comply with normal law. In gold, silver and other element deposits, logarithms of content are more often than not normally distributed [5]. Therefore, in samples that are used for calibration of analyzers, low metal content samples predominate. Thus, their influence on the fitting criterion increases, compensating for the lack of LSM. Naturally, the number of samples needed to reliably estimate the coefficients of equation (1) has to increase.

In the presence of specialized programs for obtaining regression parameters it is rational to use methods of nonlinear regression analysis. In this case it is possible to set the regression equation, loss function [6,7] and the algorithm of searching for solutions. For example, in LSM the loss function is determined by the sum of squares of residues, and therefore individual observations with large residues significantly affect this sum. If the measure of deviation of the observations from the regression line is constant over the whole range of contents, their contribution to the loss function is the same, the error of estimating the equation coefficients decreases, and the sample size can be reduced. For example, the value

\[
(GO_i - GO_i)'/(GO_i + RRO_i)
\]

changes slightly depending on the RRO data. In this case the residues are normalized by the sum \( (GO_i + RRO_i) \), as both methods give the content of the element with some inaccuracy.

Thus, the function of losses at “selection” of coefficients of the equation can be the sum of

\[
\sum \text{abs}(GO_i - GO_i)'/(GO_i + RRO_i) \to \text{min}.
\]

When minimizing loss functions, a quasi-Newton algorithm is used which approximates the second derivative of the loss function and uses it when searching for its minimum [8,9].

Below (see figure) the results of the study of cuttings from drilling and blasting wells in one of the blocks of Dukat field are considered. The correlation field between the RRO and GO data is shown in the left diagram of the figure. The sample includes the results of conjugate well bottomhole cuttings of 270 wells.

![Figure 1](image1.png)

**Figure 1.** Correlation fields: left - between RRP and GT data; right - between RRP and residues \((GT - GT^*)/(GT + RRP)\).

Line of regression (MNC): 1 - throughout the entire sample; 2 - after the exclusion of one abnormal observation.
The figure shows that there is a direct proportional relationship between the methods, and we can limit ourselves to considering the straight line equation (1). Discrepancies between GO and RRO increase with increasing silver content and samples with low concentrations prevail in the sample.

\[
GO^* = B_0 + B_1 \cdot RRO
\]

The regression equations obtained by the MNC have the form

\[
GO^* = (186 \pm 88) + (0.98 \pm 0.02) \cdot RRO
\]

Since the left picture uses logarithmic coordinates and the equation coefficient \(B_0 \neq 0\), the dependence on the graph shows a curve (line 1).

In the area of low silver content, the result does not agree well with the correlation field. It is quite clear that the reason is the outstanding observations, the effect of which on the value (2) is significant, because the squares of residues are summed up. In order to find out the reasons for the discrepancies, we have calculated the residues.

\[
R_i = GO_i - GO_i^*
\]

It is established that the residues are subject to the normal distribution and estimates of its parameters are obtained: arithmetic mean \( R = \sum R_i / n \), where \( n \) - number of observations, and standard deviation of the residues - \( S_R \), as well as their standardized values \( R_i / S_R \). Next, the observations were revealed, where \( \text{abs}(R/S_R) > 4 \). The values of their standardized residues were: -7.43; 6.18; -5.24; 4.36; 4.18. The probability of such extreme deviations was less than 0.0001. Therefore, the coefficients of equation (1) were additionally calculated after successive exclusion of observations with \( \text{abs}(R/S_R) \) values greater than 7; 5 and 4.

Table 1. Estimates of equation coefficients obtained by LSM.

| \( N_0 \) | \( GT^* = B_0 + B_1 \cdot RRP \) | p-level | \( GT^* = B_1 \cdot RRP \) | Removed observations | Critical level \( \text{abs}(R/S_R) \) |
|----------|-------------------------------|---------|-----------------------------|----------------------|---------------------|
| 1        | 186±88                        | 0.03    | 1,000 ± 0.017               | 0                    | >7                  |
| 2        | 118±77                        | 1.04±0.02 | 0.13                       | 1.050 ± 0.016        | 1                   |
| 3        | 66±62                         | 1.07±0.02 | 0.29                       | 1.076 ± 0.014        | 3                   |
| 4        | 67±56                         | 1.05±0.01 | 0.23                       | 1.054 ± 0.012        | 5                   |

Note. To the left of \( B_0 \) and \( B_1 \) are their standard deviations. p-level - probability of hypothesis \( B_0 = 0 \). Additional explanations in the text.

Table 1 presents estimates of \( B_0 \), \( B_1 \) and probability of hypothesis \( B_0 = 0 \). The null-left hypothesis is accepted in 2, 3, and 4 cases, since the p level of a more critical value equal to 0.05, which is set when solving geological problems [4]. Therefore, Table 1 provides additional estimates of the equation coefficient

\[
GO^* = B_1 \cdot RRO
\]

The exclusion of observation with \( R/S_R = -7.43 \) had a significant impact on factor estimates. It turned out that one could limit oneself to the estimation of the angle factor \( B_1 \). It would seem that the conclusion is obvious - in case there is no identifiable element, the analyzer readings should be equal to zero. In practice, the background radiation is measured with an error, and the condition may not be met. The discrepancies of \( B_1 \) values of equations 2, 3 and 4 are within the limits of their estimation error, i.e. they are not statistically significant [14].

The right graph in the figure shows changes in normalized balances (3) depending on the RRP data. They are arranged symmetrically on the abscissa axis. The weak tendency to decrease the abscissa values of residues with increasing silver content does not play a significant role. Thus, the sum (4) can be a loss function. As in the case of the LSM, we can limit ourselves to calculating the angular factor \( B_1 \). The regression equation obtained by the Quasi-Newton method is as follows.
The B₁ estimate was almost identical to the result obtained by the LSM after the exclusion of emissions.

In order to check the stability of the solution, the sample was divided into 5 groups. The groups were formed as follows: observations were ordered in the order of the sum GO+RRO, the first group included observations with the order number: 1, 6, 11,...1+5(k - 1), the second group included observations with the order number 2, 7, 12,...2+5(k -1), etc., where k is the number of observations in the group. The quantiles of silver distribution in groups and the initial sample are given in Table 2. The Kruskal-Wallis test showed that the samples match (p-level = 0.9).

### Table 2. Quantiles of Ag distribution according to RRO and GO data, y/t.

| Group  | n  | RRO  | GO  |
|--------|----|------|-----|
|        |    | 0,25 | 0,50 | 0,75 | 0,25 | 0,50 | 0,75 |
| 1      | 54 | 79   | 350  | 1580 | 87   | 482  | 1910 |
| 2      | 54 | 79   | 354  | 1580 | 80   | 373  | 2080 |
| 3      | 54 | 81   | 356  | 1620 | 80   | 301  | 1800 |
| 4      | 54 | 81   | 362  | 1640 | 111  | 423  | 1560 |
| 5      | 54 | 83   | 369  | 1640 | 124  | 384  | 2500 |
| Selection | 270 | 81 | 357  | 1620 | 89   | 391  | 1850 |

The results of estimates of B₁ values and their standard deviation in groups are given in Table 3.

### Table 3. B₁ value and its middlesquare deviation SB for groups and all observations.

| Group | 1  | 2  | 3  | 4  | 5  | All observations |
|-------|----|----|----|----|----|------------------|
| B₁    | 1,11 | 1,01 | 0,97 | 1,06 | 1,10 | 1,04             |
| S_B   | 0,12 | 0,13 | 0,12 | 0,26 | 0,15 | 0,16             |

Statistical differences of B₁ values in groups are within the limits of estimation errors. Traditionally, the algorithm of equation coefficients calculation, based on the loss function (4), is stable to the sample formation [10, 11].

3. Conclusion

A comparison of the results obtained by the two methods shows the following:

- LSM estimates are sensitive to the excreted samples, and therefore, in order to improve the reliability of the solution, it is necessary to analyze standardized residues and exclude observations where the absolute value of residues exceeds critical level 4-5;
- estimates based on the loss function \( \sum \text{abs}(GO_i - GO_i')/(GO_i + RRO_i) \rightarrow \min \) are resistant to the way the sample is formed, so its number can be reduced by a factor of 2-3 without loss of accuracy [12, 13].

Acknowledgments

The work is partially supported by the RFBR grant # 18-01-00796.

References

[1] Meyyer V A and Ivanyukovich G A 1989 X-ray radiometric methods for ore quality management (Leningrad: Izd. Leningrad State University) p 292
[2] David M 1980 Geostatistical methods for estimating ore reserves (Leningrad: Nedra) p 360
[3] Dreyper N, Smit G Applied regression analysis (Moscow: Finance and Statistics) p 351
[4] Gujarati, Damodar N; Porter, Dawn C 2008 Basic Econometrics (McGraw-Hill Education) p
944

[5] Ivanyukovich G A 2010 Statistical analysis of eco-geological data: Workshop for solving problems using a software package (Statistics) ed I M Khaykovich, V V Kurilenko (St. Petersburg: St. Petersburg university) p 204

[6] Krige D G 1968 The role of mathematical statistics in refining methods for assessing industrial mineralization in mines in South Africa (Leningrad: Nedra) pp 252-271

[7] Wald A 1939 Contributions to the theory of statistical estimation and testing hypotheses. Annals of Mathematical Statistics p 299-326

[8] Zaitseva I, Bogdanova S, Malafeyev O, Kolesov D and Marenchuk Y 2019 Competitive Mechanism For The Distribution Of Labor Resources In The Transport Objective Journal of Physics: Conference Series 1172 012089

[9] Malafeyev O, Lakhina J, Zaitseva I, Redinskikh N, Smirnova T and Smirnov N 2019 A mathematical model of production facilities location Journal of Physics Conference Series 1172 012090

[10] Zaitseva I, Malafeyev O, Poddubnaya N, Vanina A and Novikova E 2019 Solving a dynamic assignment problem in the socio-economic system Journal of Physics Conference Series 1172 012092

[11] Zaitseva I, Malafeev O, Strekopytov S, Bondarenko G and Lovyannikov D 2018 Mathematical model of regional economy development by the final result of labor resources AIP Conference Proceedings 1952 020011

[12] Zaitseva I, Ermakova A, Shlaev D, Shevchenko E and Lugovskoy S 2017 Workforce planning distribution of the region's results Research journal of pharmaceutical biological and chemical sciences 8 1862-66

[13] Zaitseva I, Kolesin I, Malafeyev O, Ermakova A and Shlaev D 2017 Modeling of the Labour Force Redistribution in Investment Projects with Account of their Delay IEEE Int. conf. on power, control, signals and instrumentation engineering (ICPCSI) pp 68-70