Estimation of the Regression Equation Parameters X-Ray Radiometric and Geological Testing on Deposits of Rare and Precious Metals

O A Malafeyev, G A Ivanyukovich, N D Redinskikh, I V Zaitseva, Y A Pichugin, A A Shulga

1Saint-Petersburg State University, Institute of Earth Sciences, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia
2Saint-Petersburg State University, Faculty of Applied Mathematics and Control Processes, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia
3Stavropol State Agrarian University, Zootekhnichesky Lane, 12, 355017, Stavropol, Russia
4Saint-Petersburg State University of Aerospace Instrumentation, Bolshaya Morskaia str. 67, 190000, Saint-Petersburg, Russia

E-mail: g.ivanukovich@spbu.ru

Abstract. Analysers for x-ray radiometric testing (X-RRT) are graded by comparing the results with traditional geological testing (GT). This time-consuming operation involves the collection of several hundred tests and their subsequent analysis in the laboratory. The "selection" of the regression equation coefficients between the X-RRT and GT data is complicated by a large range of metal contents and extremely uneven its distribution in the ore zones. For these reasons, the standard deviations between the testing methods are large and vary widely depending on the metal content. To increase the accuracy of regression parameter estimates, it is necessary to enlarge the control sampling. Such method raises the cost of testing and reduces its efficiency at variability of textural and structural features of ores. 270 slime tests of drilling and blasting wells of silver deposit Dukat are used for the analysis. The silver content in the tests according to the assay analysis was from 20 to 10 000 and more g/t. The differences between the main and repeated GT in the intervals of silver content 50–239, 240–999, 1000 g/t and more were 76, 175 and 651 g/t, respectively. Estimates of the regression equation between the two methods of testing with the help of two algorithms are compared. In the first, the loss function determined by the sum squares residuals (values deviations of the dependent variable from the regression line), in the second – the sum of the absolute values of the residuals normalized by the sum of the values of the dependent and independent variables. As a result, it is shown that in the first case, in order to improve the accuracy of coefficient estimates, it is necessary to exclude observations in which the absolute value of the residuals normalized by their standard deviation exceeds the critical level of 5-7. The second method is resistant to sampling and the number of observations can be reduced by 2 -3 times. Under these conditions, the error of coefficient estimates is close.
1. Introduction
X-ray radiometric testing (X-RRT) is widely used at the deposits of rare and nonferrous metals for the next purposes:
- determination of boundaries of industrial ore bodies in terms of their natural occurrence,
- operational testing of beaten commercial ore in the bulk to clarify the plans of its shipment to the concentrating,
- measurement of metal contents in the shipping containers and on the conveyor belt for the rejection of substandard ores.

Analysers for X-RRT are graded by comparing the results with traditional geological testing (GT). This time-consuming operation involves the collection of a large number of tests and their preparation 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 the features of many deposits of rare and precious metals, which complicate the choice of the regression equation coefficients between the data of X-RRT and GT. First, a large range of useful component content. Thus, at the gold silver Dukat deposit of the Magadan region, the results of X-RRT and GT are discussed below, practical interest is in ores with a silver content starting from 50 g/t and the upper limit reaches 10,000 g/t or more. As a consequence, the error of determination of concentration of the metal varies widely and it is necessary to take into account in the calculation of the regression equation. Secondly, the distribution of metal in the ore zones is extremely irregularly. For this reason, the standard discrepancies between the results of the main and repeated GT are large. Thus at the Dukat deposit the difference at GT of silver content in the intervals 50 – 239, 240 – 999, 1000 g/t and more is 76, 175 and 651 g/t, respectively. Another effect of the uneven distribution of silver is the predominance of test with low silver content. As a result, it is necessary to increase the number of tests, which significantly complicates and increases the cost of the calibration procedure of analysers.

2. Methods and results
The parameters of the regression equation between the data of X-RRT and GT

$$GO^* = B_0 + B_1 \cdot RRO + B_2RRO^2 + \ldots + B_nRRO^n$$

(1)

It is usually found by the least squares method (LSM), according to which the coefficients of the regression equation is determined by the condition:

$$\sum_{i=1}^{n} (GO_i - GO^*_i)^2 = \min.$$  

(2)

The variable \(RRO\) is called predictive, \(GO\) – feedback, \(GO^*\) – forecast or estimation, and the difference \(GO – GO^*\) – the remains.

The popularity of LSM to a certain extent is due to the lack of specialized statistical programs in field geological exploration parties and mining companies. The method is effective if the deviations of observations from the regression line (residuals) are constant over the entire content range. Otherwise, the fitting criterion will be determined by tests with a high element content, and the error of coefficient estimation increases. Therefore, for non-uniform measurements, observations should be given weights that are inversely proportional to the error of the GO values. The coefficients are calculated by weighted least squares method. The dependence between weights and metal content can be established if repeated observations are available for several X-RRT values. When testing ores, this procedure is difficult to implement. In the particular case, if the error increases in proportion to the content, the problem is simplified. It is enough to perform the transformation of the form \(log(GO) = GO\) and \(go(RO) = RO\), and apply the LSM using logarithms. In practice, it has to deal with the error of content measurement, which increases with the growth in the content of elements and with the influence of uneven distribution of elements in ores. Such distribution of elements increases with decreasing content, and therefore the scope of this method is limited. Along with it, considering the LSM, it should be noted that in geology the distribution of the content of elements, in the most cases, does not correspond to the normal distribution. On the deposits of gold, silver and other elements, logarithms of contents are more often normally distributed. Therefore, in samples that are used for calibration of
analysers, samples with low metal content prevail. Thus, their influence on the fitting criterion increases, compensating for the disadvantage of LSM. Naturally, it is necessary to increase the number of samples necessary for a reliable estimate of the equation coefficients (1). In the presence of specialized programs for obtaining regression parameters, it is rational to use methods of nonlinear regression analysis. In this case, you can set the regression equation, the loss function and the algorithm for finding solutions. Thus, in the LSM, the loss function is determined by the sum of the squares of the residuals, and therefore individual observations with large residuals significantly influence on this amount. If the observations deviation measure from the regression line is constant over the entire range of contents, their contribution to the loss function is the same. In such case the estimation error of the equation coefficients is reduced, and the sample size can be decreased. For example, the value of

$$(GO_i - GO^*)(GO_i + RRO_i)$$

varies slightly depending on the X-RRT data. In this case, the balances are normalized by the amount of $(GO_i + RRO_i)$, since both methods give the content of the element with some error. Thus, the function of losses in the choice of the coefficients of the equation can be the sum

$$\sum \text{abs}(GO_i - GO^*)(GO_i + RRO_i) \rightarrow \text{min}.$$ (4)

The Quasi-Newton algorithm is used by minimizing the loss functions. This algorithm approximates the second derivative of the loss function, which is used for finding the minimum of the loss function. The research results of the drilling and blasting slime wells one of the Dukat deposit block are considered below (see figure). The correlation field between the data of X-RRT and GT is on the left chart of the figure. The sampling includes the results of conjugated X-RRT and GT of 270 slime well. The existence of the direct proportional relationship between the methods is followed from the figure and so it is possible to consider only of the line equation (5). The differences between GT and X-RRT are increased with the silver content and in the sampling the tests with low silver content are dominated [1-24].

$$GO^* = B_0 + B_1 RRO.$$ (5)

Since the logarithmic coordinates are used on the left figure and the coefficient of the equation $B_0 \neq 0$, the curve (line 1) represents the relation on the graph. In the field of low silver content, the result is poorly consistent with the correlation field. Very likely, the reason is the stand-out observations essentially influencing on the value (2), as the squares of residuals are summed. To find out the reasons for the discrepancies, the residuals are calculated, $R_i = GO_i - GO^*$. It is established that the residuals follow the normal distribution and its parameters estimates are obtained: the arithmetic mean - \( R = \sum R_i / n \), where \( n \) is the number of observations, and the standard deviation of the residuals $-S_R$, and also their standardized values $- R_i / S_R$. Further, the observations with $\text{abs}(R_i / S_R) > 4$ were identified. The values of their standardized residuals equal: $-7.43$; $6.18$; $5.24$; $4.36$; $4.18$. The probability of such extremal deviations is less than 0.0001. Therefore, the coefficients of equation (5) are additionally calculated after the sequential exclusion of observations in which the values of $\text{abs}(R_i / S_R)$ exceed 7; 5 and 4. In table1 estimates of $B_0$, $B_1$ and the probability of the hypothesis $B_0 = 0$ are presented. The null hypothesis is accepted in 2, 3 and 4 cases, since more critical p-level equals to 0.05 is set by solving geopolitical problems. Therefore, table 1 provides additional estimates of the coefficient of the equation $GO^* = B_1 RRO$. A significant impact on the estimates of the coefficients was made by the exclusion of the observation with $R_i / S_R = -7.43$. It turned out that it is possible to limit the estimation of the angular coefficient $B_1$. It would seem that the conclusion is obvious – in the absence of a certain element, the analyser Indicators should be zero. In practice, background radiation is measured with an error, and the condition may not be met. The differences in the values of $B_1$ of equations 2, 3 and 4 are within the error of their estimation, i.e. are not statistically significant. The changes in the normalized residuals (3) depending on the X-RRT data are shown on the right graph of the figure. They are located symmetrically relative to the x-axis. A weak tendency to decrease the
absolute values of the residuals with an increase in the silver content does not play a significant role. Thus, the loss function can be sum (4). As in the case of LSM, it is possible to calculate only the angular coefficient $B_1$. The regression equation obtained by the Quasi-Newton method has the form $GO^* = (1.04\pm0.16) RRO$. The $B_1$ estimation was almost the same as the result obtained by the LSM after the emission exceptions. In order to check the stability of the solution, the sampling is divided into 5 groups. The groups are formed as follows: observations were ordered in descending order of the sum $GO+RRO$. The first group is included observations with the sequence number: 1, 6, 11,...,1+5(k - 1), the second - 2, 7, 12,...,2+5(k -1), etc., where k is the observation number in the group. The quantiles of the silver distribution in the groups and the initial sample are given in table 2. The Kruskal-Wallis test showed that the sampling are the same ($p$-level = 0.9). The results of the estimates of $B_1$ values and their standard deviation in the groups are given in table 3. Statistical differences $B_1$ in the groups are within the estimation errors. Consequently, the algorithm for calculating the coefficients of the equation, based on the loss function (4), is resistant to the formation of the sampling.

A comparison of the results obtained by the two methods shows the following:
- LSM estimations are sensitive to distinctive tests, and therefore, in order to improve the reliability of the solution, it is necessary to analyse standardized residues and exclude observations in which the absolute value of residuals exceeds the critical level of 4-5;
- estimations based on the loss function $\sum \text{abs}(GO_i - GO^*)/(GO_i + RRO_i) \rightarrow \min$ are stable to the way of sampling forming, therefore its number of observations can be reduced in 2-3 times without loss of estimation accuracy.

Statistical program Statistica 10 is used for statistical data analysis.

![Figure 1](image.png)

**Figure 1.** The correlation fields: left chart – between the data $RRO$ and $GO$; right chart – between the data $RRO$ and residuals $(GO - GO^*)/(GO + RRO)$. Regression line (LSM): 1 – across the sampling; 2 – after excluding one abnormal observation.

| № | $GO^* = B_0 + B_1 \cdot RRO$ | $p$-level | $GO^* = B_1 \cdot RRO$ | Dropped observations | Critical level $\text{abs}(R/S_0)$ |
|---|---|---|---|---|---|
| 1 | $186\pm88$ | 0.98±0.02 | 0.03 | $1,000 \pm 0.017$ | 0 |
| 2 | $118\pm77$ | 1.04±0.02 | 0.13 | $1,050 \pm 0.016$ | 1 | $>7$ |
| 3 | $66\pm62$ | 1.07±0.02 | 0.29 | $1,076 \pm 0.014$ | 3 | $>5$ |
| 4 | $67\pm56$ | 1.05±0.01 | 0.23 | $1,054 \pm 0.012$ | 5 | $>4$ |

**Remark.** To the left of $B_0$ and $B_1$ their standard deviations are specified. $p$-level – the probability of the hypothesis $B_0=0$. Additional explanations in the text.
Table 2. Quantiles of the distribution of Ag according X-RRT and GT, g/t.

| Group | n  | X-RRT | GT |
|-------|----|-------|----|
|       |    | 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 |
| Sampling | 270 | 81 | 357  | 1620 | 89   | 391  | 1850 |

Table 3. Value B1 and its standard 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 |
| SB    | 0.12 | 0.13 | 0.12 | 0.26 | 0.15 | 0.16 |

3. Conclusion
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 is defined by the sum of residuals squares (deviations from the values of the dependent variable of the regression line), in the second, the sum of the absolute values of the residual is normalized to the sum values of the dependent and independent variables. As a result, it is shown that in the first case, in order to increase the estimates accuracy of the coefficients, it is necessary to exclude observations in which the absolute value of the residuals is 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 is minimal.

4. References
[1] Kolokoltsov V N, Malafeyev O A 2017 Mean-Field-Game Model of Corruption Dynamic Games and Applications 7(1) 34-47
[2] Malafeyev O, Alferov G, Andreyeva M 2015 Group strategy of robots in game-theoretic model of interception with incomplete information International Conference on Mechanics - Seventh Polyakhov’s Reading Proc. http://dx.doi.org/10.1109/POLYAKHOV.2015.7106751
[3] Malafeyev O A, Neverova E G, Nemnyugin S A 2014 Multi-criteria Model of Laser Radiation Control 2nd International Conference on Emission Electronics Proceeding http://dx.doi.org/10.1109/Emission.2014.6893966
[4] Malafeyev O A, Nemnyugin S A, Alferov G V 2014 Charged Particles Beam Focusing with Uncontrollable Changing Parameters 2nd International Conference on Emission Electronics Proceeding http://dx.doi.org/10.1109/Emission.2014.6893964
[5] Grigorieva X, Malafeev O A 2014 A competitive many-period postman problem with varying parameters Applied Mathematical Sciences 8(145-148) 7249-7258
[6] Malafeev O A, Nemnyugin S A 1996 Generalized dynamic model of a system moving in an external field with stochastic components Theoretical and Mathematical Physics 107(3) 770-774
[7] Pichugin Y A, Malafeyev O A 2016 Statistical estimation of corruption indicators in the firm Applied Mathematical Sciences 10(41-44) 2065-2073
[8] Drozdov G D, Malafeyev O A, Nemnyugin S A 2015 Multicomponent dynamics of competitive single-sector economy development SCP 2015 Proceedings
[9] Malafeev O A 1977 Stationary strategies in differential games USSR Computational Mathematics and Mathematical Physics 17(1) 37-46

[10] Malafeyev O A, Nemnyugin S A, Ivaniukovich G A 2015 Stochastic models of social-economic dynamics SCP 2015 Proceedings 483-485

[11] Malafeev O A 1974 Equilibrium situations in dynamic games Cybernetics and System Analisys 10(3) 504-513

[12] Neverova E G, Malafeyef O A 2015 A model of interaction between anticorruption authority and corruption groups AIP Conference Proceedings http://dx.doi.org/10.1063/1.4912671

[13] Malafeyev O A, Redinskikh N D 2016 Stochastic analysis of the dynamics of corrupt hybrid networks Proceedings of 2016 International Conference "Stability and Oscillations of Nonlinear Control Systems" (Pyatnitskiy's Conference) http://dx.doi.org/10.1109/STAB.2016.7541208

[14] Malafeev O A 1974 The existence of situations of ε-equilibrium in dynamic games with dependent movements USSR Computational Mathematics and Mathematical Physics 14(1) 88-99

[15] Zaitseva I, Malafeyev O, Strekopytov S, Bondarenko G, Lovyannikov D 2018 Mathematic model of regional economy development by the final result of labor resources AIP Conference Proceedings https://doi.org/10.1063/1.503197

[16] Malafeyev O, Rylow D, Zaitseva I, Zelenkovskii P, Popova M, Novozhilova L 2017 Game-theoretic model of dispersed material drying process AIP Conference Proceedings http://dx.doi.org/10.1063/1.4990216

[17] Zaitseva I, Popova M 2013 Technique to study the employment potential of the region: economic-mathematical aspect World Applied Sciences Journal 22(1) 22-25

[18] Zaitseva I, Kriulina E, Ermakova A, Shevchenko E, Vorokhobina Y 2016 Application of Factor Analysis to Study the Labour Capacity of Stavropol Krai Research Journal of Pharmaceutical Biological and Chemical Sciences 7(4) 2183-2186

[19] Zaitseva I, Popova M, Ermakova A, Bogdanova S, Rezenkov D 2016 Determination Prospects Of Development Labor Potential In Agriculture Stavropol Territory Based On Assessment His Condition Research Journal of Pharmaceutical Biological and Chemical Sciences 7(3) 2592-2595

[20] Zaitseva I, Malafeyev O, Strekopytov S, Ermakova A, Shlaev D 2018 Game-theoretical model of labour force training Journal of Theoretical and Applied Information Technology

[21] Zaitseva I, Malafeyev O, Strekopytov S, Bondarenko G, Lovyannikov D 2018 Mathematic model of regional economy development by the final result of labor resources AIP Conference Proceedings American Institute of Physics

[22] Zaitseva I V, Ermakova A N, Shlaev D V, Shevchenko E A, Lugovskoy S I 2017 Workforce planning redistribution of the region's results Research Journal of Pharmaceutical Biological and Chemical Sciences 8(1) 1862-1866

[23] Kostyukov K, Zaitseva I, Bondarenko G, Svechinskaya T, Nemyayeva S 2016 Workforce Planning as An Element of Control System Research Journal of Pharmaceutical Biological and Chemical Sciences 7(6) 2315–2319

[24] Pesterov A P, Yakovlev V A, Kirillina A A, Solovev D B 2019 Environmental Problems Mining Industry in the Arctic IOP Conference Series: Earth and Environmental Science 272 paper № 022055. [Online]. Available: https://doi.org/10.1088/1755-1315/272/2/022055

Acknowledgments
The work is partly supported by grant RFBR No. 18-01-00796.