

Uncertainty of nickel chemical analysis in paired samples of lateritic ore, NiS and NiO by linear regression

Incertidumbre del análisis químico de níquel en muestras pareadas de mineral laterítico, NiS y NiO por regresión lineal

Armando Rojas-Vargas^{1*} <https://orcid.org/0000-0002-8927-2023>

Aymara Ricardo-Riverón² <https://orcid.org/0000-0002-2014-4314>

Yudith González-Díaz³ <https://orcid.org/0000-0003-1240-1146>

Benigno Labrada-Vázquez³ <https://orcid.org/0009-0003-5654-1519>

¹Empresa de Servicios Técnicos de Computación, Comunicaciones y Electrónica del Níquel (SERCONI), Cuba

²Centro de Investigaciones del Níquel (CEDINIQ), Cuba

³Facultad de Ingeniería Química y Agronomía, Universidad de Oriente, Cuba

*Autor para la correspondencia. Correo electrónico: arojas@eros.moa.minem.cu

ABSTRACT

Interlaboratory comparison is an instrument for evaluating the laboratory performance in specific tests. In this work, the uncertainty in the quantitative chemical analysis of the nickel concentration was determined in paired solid samples of the materials: lateritic ore (LO), nickel sulfide (NiS) and sintered nickel oxide (NiO), using the simple linear regression. The analytical determinations were made by Atomic Absorption Spectrometry (AAS) with the participation of three laboratories of the corporative group for the production of nickel in Cuba. The assumptions of the linear regression model were verified using the Durbin-Watson statistic, homoscedasticity and normality test. The uncertainty in the slope (S_m), the uncertainty in the intercept (S_b) and the standard deviation of the measurement (S_y) were determined. As a result, the expected value (x) of the nickel concentration presented a maximum error according to: LO: $\pm 0,06$ %, NiS: $\pm 0,51$ % y NiO: $\pm 0,33$ % at the extreme ends of

the measurement range. This method provides information to evaluate the laboratories continuous performance, and in turn the historical database of interlaboratory comparison is analyzed.

Keywords: interlab comparison; lineal regression; uncertainty.

RESUMEN

La comparación interlaboratorio es un instrumento para evaluar el desempeño del laboratorio en pruebas específicas. En este trabajo, se determinó la incertidumbre en el análisis químico cuantitativo de la concentración de níquel en muestras sólidas pareadas de los materiales: mineral laterítico (ML), sulfuro de níquel (NiS) y óxido de níquel sinterizado (NiO), mediante regresión lineal simple. Las determinaciones analíticas se realizaron por Espectrometría de Absorción Atómica con la participación de tres laboratorios del grupo empresarial para la producción de níquel en Cuba. Los supuestos del modelo de regresión lineal se comprobaron mediante el estadístico de Durbin-Watson, pruebas de homocedasticidad y normalidad. Se determinó la incertidumbre de la pendiente (S_m), la incertidumbre del intercepto (S_b) y la desviación estándar de la medición (S_y). Como resultado, el valor esperado (x) de concentración de níquel presentó un error máximo según: ML: $\pm 0,06$ %, NiS: $\pm 0,51$ % y NiO: $\pm 0,33$ % en los extremos del intervalo de concentración. El método aporta información para evaluar el desempeño continuo de los laboratorios, a la vez que se analiza la base de datos históricos de comparación interlaboratorio.

Palabras clave: comparación interlaboratorio; regresión lineal; incertidumbre.

Received: 08/05/2024

Accepted: 15/08/2024

Introduction

Interlaboratory comparison is an instrument for evaluating the laboratory performance for specific tests, provides information on continuous performance and additional confidence to laboratory customers. ⁽¹⁾

The chemical analysis laboratories in the corporate group for the nickel production in Cuba apply interlaboratory studies as an external quality control;

therefore, there is a historical database of these controls that includes the quantitative analysis of metals such as: nickel, cobalt, iron, silicon, magnesium, in the different stages and products of the process. Chemical analysis measurements provide a basis for important decisions concerning efficiency, commerce, environmental protection, health, among others.

Common schemes of participation in interlaboratory comparisons consist of circulate the sample to realize the quantity mensuration among participants. Each participant measures according to a standard or specific procedure and reports to the coordinator a measurement result, which includes a measured value and its uncertainty. ⁽²⁾

The purpose of all measurement is to provide information about a quantity of interest. When a quantity is measured, the outcome depends on the measuring system, the measurement procedure, sampling and sample preparation, the skill of the operator, the environment; thus, no measurement is exact.

There are two types of measurement error quantity, systematic and random. ⁽³⁾

A systematic error is a component of error that remains constant or depends in a specific manner on some other quantity. A random error is associated with the fact that when a measurement is repeated it will generally provide a measured quantity value that is different of further quantities; therefore, the next measured quantity value cannot be predicted exactly from previous such values. ⁽⁴⁾

Kadis (2008) assumed Chumpler and Yoe's (1940) classification of the errors as corrigible and incorrigible, both of the random and "systematic" components. Indeed, an estimate of the corrigible error is used to make a correction; the incorrigible error contributes to the uncertainty of the result (figure 1). ^(5, 6)

Measurement uncertainty is defined as non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. ⁽⁴⁾ Uncertainties associated with analytical measurements represent the doubt or level of reliability associated with the measurement. ⁽⁷⁾

All analyzes were performed on paired samples. Paired samples are samples in which matched couplings occur, this generates a data set in which each data point in one sample is uniquely paired to a data point in the second sample.

The chemical analyzes of nickel were realized by the Atomic Absorption Spectrophotometry method, in 3 participating laboratories (table 1).

Table 1- Laboratory and equip for chemical analysis

No.	Laboratory	Spectrophotometer
1	CG	SavantAA Σ
2	RL	Avanta - P
3	UP	SP-9

All chemicals are of analytical-reagent grade and all solutions prepared with deionised water. The lateritic ore is decomposed by mixing 0,5 g of sample with 1:1 HCL and 10 vol% HNO₃. As for NiS and NiO, 0,9 g of sample is mixed with HCl (15 mL), HNO₃ (10 mL) and HClO₄ (5 mL).

Measurement uncertainty by least-squares linear regression

Simple linear regression plots a line graph between two data variables, x and y. The independent variable x (predictive variable) is plotted on the horizontal axis; while the dependent variable y (predicted variable) is plotted on the vertical axis.

The method of least-squares is used to determine the line (1) that best fits the model; this consists of minimizing the errors (2) through the quadratic sum of the vertical distances between the experimental data (y) and the estimated values (y*). Its variables are the slope m and the intercept b (3-9).^(7, 9, 12)

$$y = m \cdot x + b \quad (1)$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y_i^*)^2 = [y_i - (m \cdot x_i - b)]^2 = \text{minimun} \quad (2)$$

$$m = \frac{S_{xy}}{S_x^2} \quad (3)$$

$$b = \frac{\sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i - n \cdot \sum_{i=1}^n y_i \cdot x_i}{(\sum_{i=1}^n x_i)^2 - n \cdot \sum_{i=1}^n x_i^2} \quad (4)$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (6)$$

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (7)$$

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \quad (8)$$

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} \quad (9)$$

where:

\bar{x} , \bar{y} denote the sample means of x and y respectively,

S_x^2 and S_y^2 are the sample variance of x and y respectively,

S_{xy} is the covariance between x and y.

The linear correlation coefficient (r) between x and y, is a measure that reports the linear dependence between the two variables, varies from 0 to 1. The R-squared (R^2) is called the coefficient of determination, is the proportion of variability of y that is explained by x, represents the quality of fit (10). The slope (m) represents the relationship between the correlation and regression coefficients (11).⁽¹²⁾

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\sqrt{\sum_{i=1}^n (y_i^* - \bar{y})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (10)$$

$$m_{y/x} = r \cdot \frac{S_y}{S_x} \quad (11)$$

If r is significant, the x-axis versus y-axis plot shows a linear trend; and the model can be used to predict the value of y in the domain of observed values of x. It is verified by means of the hypothesis test and the t-distribution, for the distribution curve with two points and degrees of freedom (n-2) (12).

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{(1-r^2)}} \quad (12)$$

Null hypothesis H_0 : $r=0$. The correlation coefficient is not significantly different from zero, there is no significant linear correlation between x and y in the population.

Alternative hypothesis H_1 : $r \neq 0$. The correlation coefficient is significantly different from zero, there is a significant linear correlation between x and y in the population.

The confidence limits for the slope m and the intercept b are determined using the t-student distribution, confidence level and degrees of freedom ($n-2$) (13-18).^(9, 12)

$$m \pm t_{n-2}^{\alpha/2} \cdot S_m \quad (13)$$

$$b \pm t_{n-2}^{\alpha/2} \cdot S_b \quad (14)$$

$$S_m = S_y \cdot \sqrt{\frac{n}{n \cdot \sum_{i=1}^n (x_i)^2 - (\sum_{i=1}^n x_i)^2}} \quad (15)$$

$$S_b = S_m \cdot \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n}} \quad (16)$$

$$n \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = n \cdot \sum_{i=1}^n (x_i)^2 - \left(\sum_{i=1}^n x_i \right)^2 \quad (17)$$

$$U_c = \sqrt{S_m^2 + S_b^2} \quad (18)$$

where:

S_m is uncertainty in the slope; S_b is uncertainty in the intercept; U_c is the combined uncertainty.

Uncertainty in the measurements of y (S_y), or also called standard deviation for the lineal regression, is used in constructing confidence intervals and prediction intervals (error bars) for values of y_i (19).

$$S_y = \sqrt{\frac{\sum_{i=1}^n (Y_i - m \cdot x_i - b)^2}{n - 2}} \quad (19)$$

The results of previous equations were based on the assumptions that the measurements of y_i were all equally uncertain and that any uncertainties in x_i were negligible.

The expected value x_0 (20) for a certain value y_0 is expressed from the uncertainty (S_{x_0}) associated with x_0 , both when a single reading of y_0 (21) or when k readings of y_0 (22) are performed, together at the confidence interval (23).

$$x_0 = (y_0 - b) \cdot \frac{1}{m} \quad (20)$$

$$S_{x_0} = \frac{S_y}{m} \cdot \sqrt{1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{m^2 \cdot \sum_i (x_i - \bar{x})^2}} \quad (21)$$

$$S_{x_0} = \frac{S_y}{m} \cdot \sqrt{\frac{1}{k} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{m^2 \cdot \sum_i (x_i - \bar{x})^2}} \quad (22)$$

$$x_0 \pm t_{n-2}^{\alpha/2} \cdot S_{x_0} \quad (23)$$

The best value of y_0 at any point will be the mean of all possible observed values of y_0 at that point. The variance for y_0 is (24) and confidence interval for the mean value of y_0 at x_0 is (25).

$$S_{x_0} = S_x^2 \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}} \quad (24)$$

$$y_0 \pm t_{n-2}^{\alpha/2} \cdot S_{x_0} \quad (25)$$

The assumptions of the linear regression model are:

The residuals must be independent of each other. It is verified by means of the Durbin-Watson (DW) statistic. DW value should be in the range 1,5 to 2,5 with a P-value greater than 0,05 and a confidence level of 95,0 %.

Equality of variances (Homoscedasticity). For each value of the variable X, the variance of the residuals (e_i) must be the same, so the adjustment is equally precise regardless of the values that x takes. It is obtained by observing the graph of predicted values *versus* residuals, no association pattern should be appreciated in the cloud of points. ⁽¹¹⁾

Normality. For each value of the variable X, the residuals e_i have a normal distribution with mean zero. The data from the analytical determinations that come from a normal distribution have standard skewness and kurtosis statistics within the range -2 to 2.

Statistical analysis was performed using StatGraphic Centurion XV and Microsoft Excel.

Results and discussion

The statistics summary for nickel (Ni) determinations with size (n) that were carried out in the CG, RL and UP laboratories are shown in table 2. Measures of central tendency (average) and variability (standard deviation, coefficient of variation, range) are presented. In the case of shape measurements (standardized skewness and standardized kurtosis), these are presented for the paired samples.

Table 2- Statistical summary of nickel concentration in paired samples

Product Laboratory	NiS		NiO		Feed & HC		Feed & HC		Feed	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(3)	(1)	(3)
n	23	23	37	37	19	19	46	46	26	26
Average	14,33	14,51	88,57	88,56	1,13	1,13	1,15	1,152	1,20	1,21
Standard deviation	2,84	2,81	3,38	3,35	0,07	0,07	0,09	0,09	0,09	0,09
Coefficient of variation (%)	19,85	19,34	3,82	3,78	6,09	6,18	7,62	7,41	7,97	7,82
Minimum	10,40	10,64	75,97	76,12	1,00	0,97	0,89	0,89	1,05	1,07
Maximum	20,03	20,45	93,29	93,24	1,23	1,23	1,3	1,32	1,42	1,43
Range	9,63	9,81	17,32	17,12	0,23	0,26	0,41	0,43	0,37	0,36
Standardized skewness	-1.02		0.56		0.76		-1.06		0.65	
Standardized kurtosis	1.21		-1.25		-1.01		0.00		-1.04	

(1) CG; (2) RL; (3) UP

From table 2, the analytical determinations of nickel in all solid samples could reasonably be assumed to have come from normal distributions, according to the shape measurements between -2,0 and 2,0.

The standardized residual did not follow a pattern of association in the cloud of nickel concentration points in each product; therefore, the equality of variances (homoscedasticity) can also be assumed (figures 2 to 6).

The residuals were independent among themselves, according to the satisfactory values of the Durbin-Watson (DW) statistic with a confidence level of 95,0 % ($DW > 1,5$ and $P > 0,05$). The mean absolute error (MAE) varied between 0,017 and 0,172 (table 3).

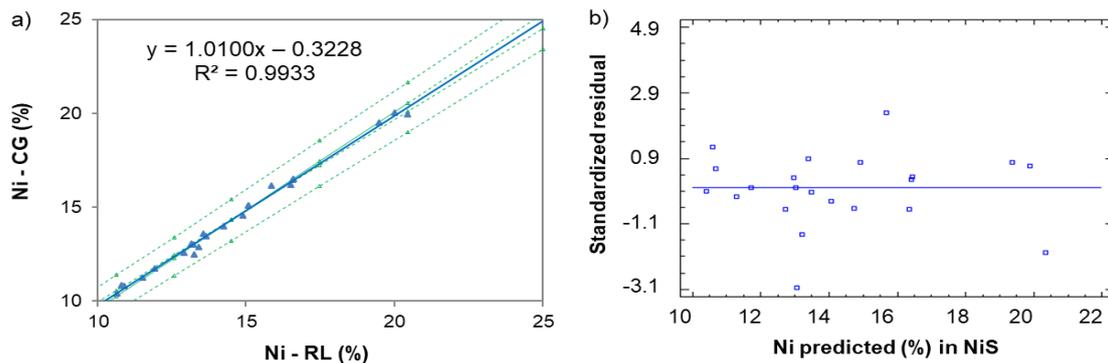


Fig. 2- Nickel determination in Nickel Sulfide between laboratories CG-RL, a) Lineal model for NiS, b) Standardized residual

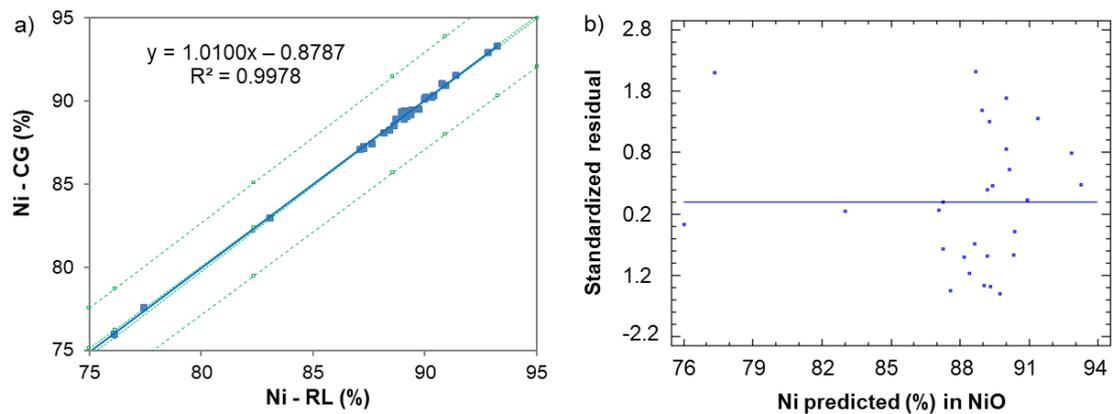


Fig. 3- Nickel determination in Nickel Oxide between laboratories CG-RL a) Lineal model for NiO, b) Standardized residual

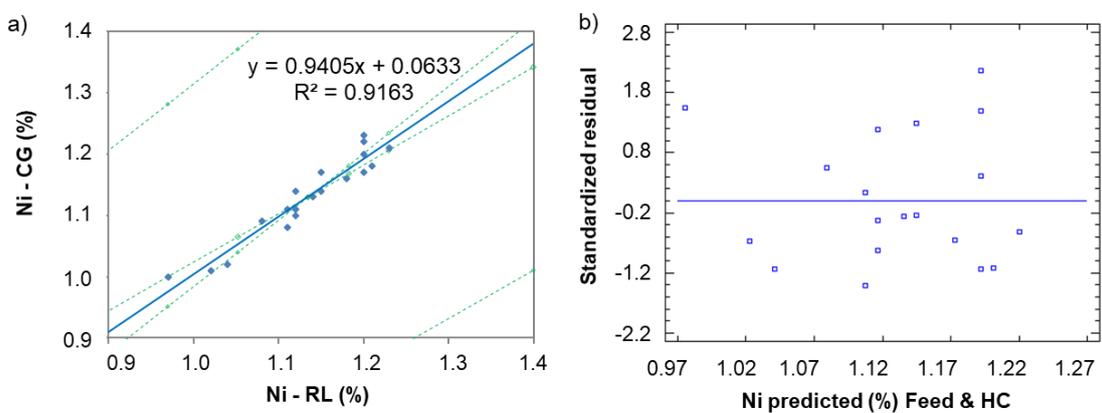


Fig. 4- Nickel determination in ore feed with petroleum additive between laboratories CG-RL, a) Lineal model, b) Standardized residual

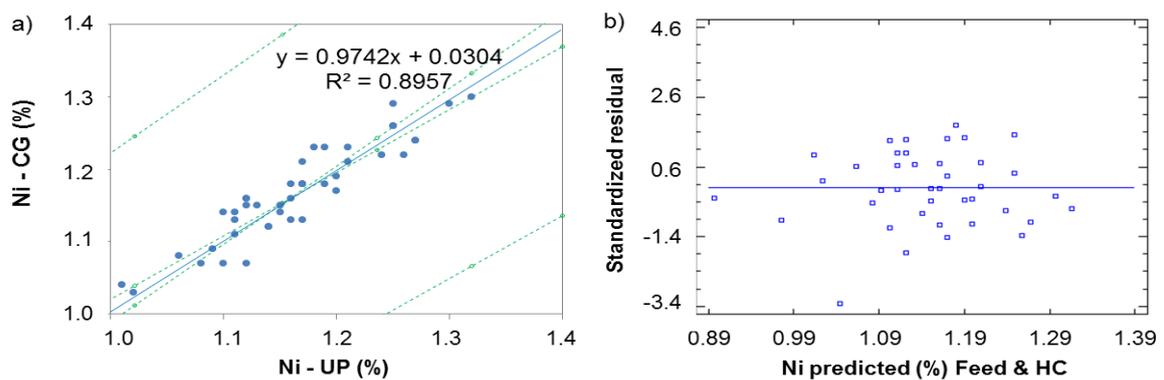


Fig. 5- Nickel determination in ore feed with petroleum additive between laboratories CG-UP, a) Lineal model, b) Standardized residual

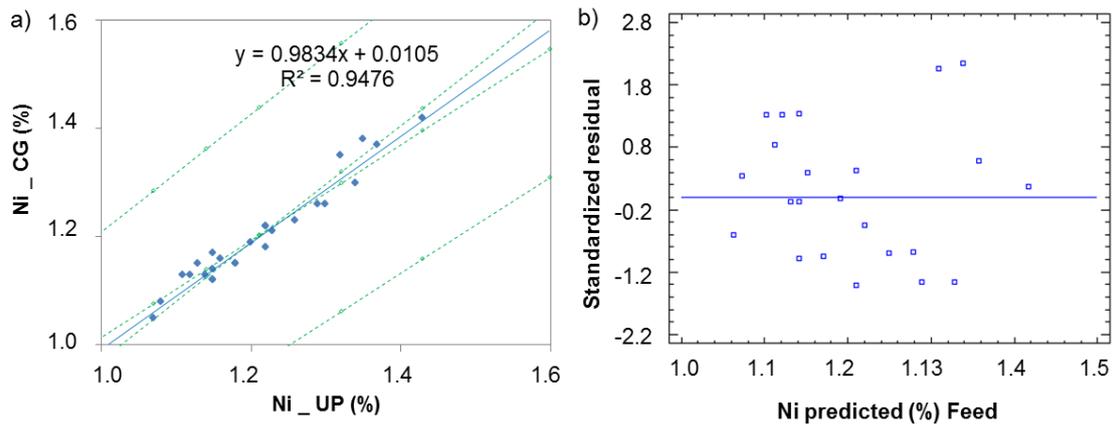


Fig. 6- Nickel determination in ore feed without petroleum additive between laboratories CG-UP, a) Lineal model, b) Standardized residual

Table 3- Parameters of linear models for nickel concentration

Items	Product				
	NiS ^(1,2)	NiO ^(1,2)	Feed & HC ^(1,2)	Feed & HC ^(1,3)	Feed ^(1,3)
DW	2.051	1.500	1.716	1.767	1.877
(P)	(0.542)	(0.061)	(0.243)	(0.174)	(0.299)
MAE	0.172	0.125	0.017	0.023	0.018
R ²	0.993	0.998	0.916	0.896	0.948
S _m	0.018	0.008	0.069	0.050	0.047
S _b	0.267	0.714	0.078	0.050	0.057
U _c	0.268	0.714	0.104	0.077	0.074
S _y	0.238	0.162	0.021	0.029	0.022
t _{α, n-2}	2.080	2.030	2.110	2.015	2.060
m ± t · S _m	1.010 ± 0.038	1.01 ± 0.016	0.941 ± 0.145	0.974 ± 0.101	0.983 ± 0.097
b ± t · S _b	-0.323 ± 0.556	-0.879 ± 1.449	0.063 ± 0.165	0.030 ± 0.117	0.011 ± 0.119

⁽¹⁾ CG; ⁽²⁾ RL; ⁽³⁾ UP

From table 3, uncertainty in the slope (S_m) and uncertainty in the intercept (S_b), the combined uncertainty (U_c) and the standard deviation (S_y) of the measurements were higher for the chemical determinations of nickel concentration in sulfide (NiS) and oxide (NiO). On the other hand, the mineral samples fed to the reduction furnaces with petroleum additive presented greater combined uncertainty (U_c) in the nickel determinations than without the additive addition (figures 2 to 6).

The expected value of x_i (RL or UP) given a value of y_i (CG) was determined by equations (20) to (23) for 3 values included in the nickel concentration ranges that were evaluated in the experiments (table 4).

The uncertainty associated (S_{x0}) with x₀ was for NiS 0,25 % in the concentration range [10,4 – 20,5 %], NiO 0,16 % [76,0 – 93,2 %], lateritic ore a maximum of 0,03 % [0,9 – 1,4 %]. The interlaboratory error varied between 0,05 % and 0,51 % for the samples of the different products. At a nickel concentration in NiS

of 17,18 % in the CG laboratory, an expected value of 17, 33 % in the RL laboratory corresponds with an error of 0,51 %, that is $17,33 \pm 0,51$ %, and the confidence interval [16,82 %; 17.84 %].

The expected limits for the compounds are: NiS $\pm 0,15$ % within [20 - 30 %], NiO $\pm 0,70$ % [70 - 80 %], lateritic ore $\pm 0,02$ % [1 - 2 %]; then, increasing the sample size and adjusting the concentration limits may be of interest for further investigation.

Table 4 - Nickel concentration x_i (RL or UP) given a value of y_i (CG)

Ítems	Products				
	NiS ^(1,2)	NiO ^(1,2)	Feed & HC ^(1,2)	Feed & HC ^(1,3)	Feed ^(1,3)
Y_1 (%)	11.49	85.19	1.06	1.06	1.11
x_1 (%)	11.69	85.21	1.06	1.06	1.11
S_{x_0}	0.25	0.16	0.02	0.03	0.02
$\pm t \cdot S_{x_0}$	0.51	0.33	0.05	0.06	0.05
Y_2 (%)	14.33	88.57	1.13	1.15	1.20
x_2 (%)	14.51	88.56	1.13	1.15	1.21
S_{x_0}	0.24	0.16	0.02	0.03	0.02
$\pm t \cdot S_{x_0}$	0.50	0.33	0.05	0.06	0.05
Y_3 (%)	17.18	91.95	1.20	1.24	1.30
x_3 (%)	17.33	91.91	1.21	1.24	1.31
S_{x_0}	0.25	0.16	0.02	0.03	0.02
$\pm t \cdot S_{x_0}$	0.51	0.33	0.05	0.06	0.05

⁽¹⁾ CG; ⁽²⁾ RL; ⁽³⁾ UP

Conclusions

1. Measurements uncertainty for the interlaboratory chemical analysis of nickel in paired samples was estimated by simple linear regression, resulting for Nickel Sulfide 0,238 %, Nickel Oxide 0,162 %, and the lateritic ore between 0,021 % and 0,029 %, which allowed the construction of confidence and prediction intervals.
2. Measurement errors were determined using t-distribution, preceding a maximum for NiS $\pm 0,51$ %, NiO $\pm 0,33$ % and lateritic ore $\pm 0,06$ %.

Nomenclature

b : Linear function origin	S_{xy} : Covariance between x and y
DW: Statistic Durbin-Watson	S_y : Standard deviation of the estimate
m : Lineal function slope	S_y^2 : Sample variance of y
MAE : Mean absolute error	U_c : Combined uncertainty
S_b : Uncertainty in the origin	\bar{x}, \bar{y} : Sample means of x and y respectively
S_m : Uncertainty in the lope	y^* : Estimated value, dependent variable
S_x^2 sample variance of x	y : Experimental date, dependent variable

References

1. NC ISO / IEC GUÍA 43-1/1997, IDT. Ensayos de aptitud por comparaciones interlaboratorios – parte 1: desarrollo y funcionamiento de programas de ensayos de aptitud.
2. BERNI, R., CAROBBI, C. Alternative statistical analysis of interlaboratory comparison measurement results. XXI IMEKO World Congress “Measurement in Research and Industry” August 30 - September 4, 2015, Prague, Czech Republic. <https://www.imeko.org/publications/wc-2015/IMEKO-WC-2015-TC21-382.pdf>
3. KADIS, R.L. Measurement Uncertainty and Chemical Analysis. *J. Anal. Chem.* [en línea]. 2008, **63**(1). 95–100. ISSN 1061-9348. DOI: 10.1134/S1061934808010188
4. JCGM 200:2008. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM). 3rd Edition (Bilingual E/F). Cited by: Evaluation of measurement data – An introduction to “Guide to the expression of uncertainty in measurement” and related documents OIML G 1-104 Edition 2009 (E).
5. KADIS, R. *Evaluation of measurement uncertainty in analytical chemistry: related concepts and some points of misinterpretation*. Dissertation for the commencement of the Degree of Doctor of Philosophy in Chemistry. University of Tartu, 2008. <http://dspace.ut.ee/handle/10062/6720?show=full>
6. CRUMPLER, T.B., YOE, J.H. *Chemical computations and errors*. New York: John Wiley and Sons; London: Chapman and Hall, 1940. Ch. VII.

7. QUEENIE, S.H.C. Uncertainties Related to Linear Calibration Curves: A Case Study for Flame Atomic Absorption Spectrometry. *J. Braz. Chem. Soc.* [en línea]. 2007, **18**(2). 424-430. ISSN: 1678-4790. DOI:10.1590/S0103-50532007000200027
8. YATKIN, S., TRZEPLA, K., HYSLOP, P., et al. Comparison of a priori and interlaboratory-measurement-consensus approaches for value assignment of multi-element reference materials. *Microchem. J.* [en línea]. 2020. **158**, 105225. ISSN 0026-265X. DOI: 10.1016/j.microc.2020.105225
9. RAWSKI, R.I., SANECKI, P.T., KIJOWSKA, K.M., et al. Regression Analysis in Analytical Chemistry. Determination and Validation of Linear and Quadratic Regression Dependencies. *S. Afr. J. Chem.* [en línea]. 2016, **69**. 166-173. ISSN 0379-4350. DOI:10.17159/0379-4350/2016/v69a20
10. CHEN, H.Y.; CHEN, C. Evaluation of Calibration Equations by Using Regression Analysis: An Example of Chemical Analysis. *Sensors* [en línea]. 2022. **22**(2). 447. ISSN 1424-8220. DOI: 10.3390/s22020447
11. RAPOSO, F. Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review. *Trends. Analyt. Chem.* [en línea]. 2016, **77**. 167–185. ISSN: 0165-9936. DOI: 10.1016/j.trac.2015.12.006
12. DEVORE, J.L. *Probabilidad y Estadística para Ingeniería y Ciencias*. Séptima Edición. California Polytechnic State University, San Luis Obispo. 2008, 742 p. ISBN-13: 978-607-481-338-8

Conflicts of Interest

The authors declare that there are not conflicts of interest.

Contribution of the authors

Armando Rojas-Vargas: conceptualization and design, investigation and acquisition of data, analysis and interpretation, writing and editing.

Aymara Ricardo-Riverón: investigation, acquisition of data, analysis, review.

Yudith González-Díaz: design, analysis, interpretation of data and review.

Benigno Labrada-Vázquez: design, analysis, interpretation of data and review.

The authors approved the final version for publication.