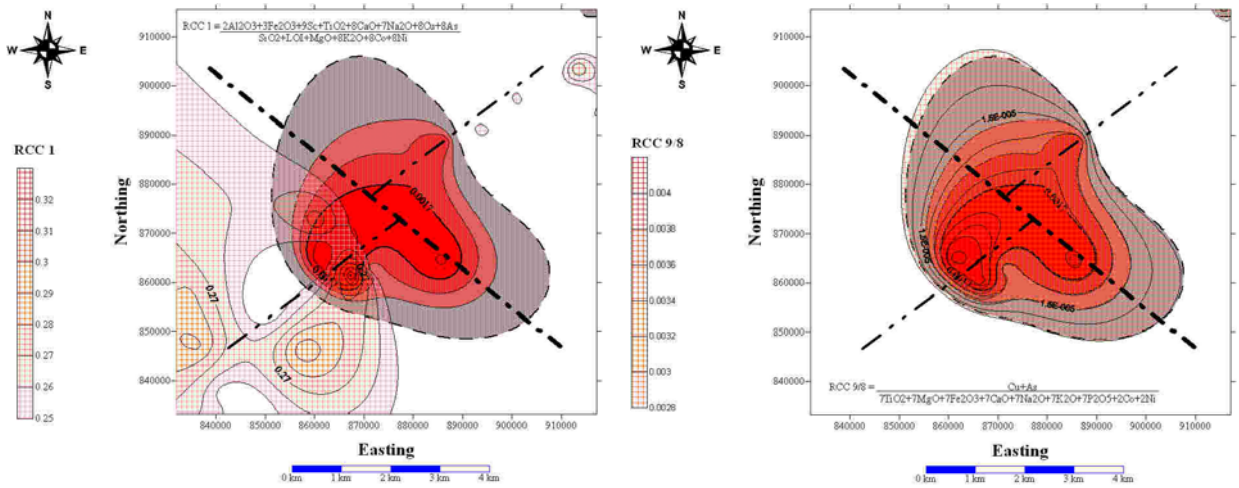
VALLS GEOCONSULTANT

WHY, AND HOW, WE SHOULD USE COMPOSITIONAL DATA ANALYSIS



A Step-by-Step Guide for the Field Geologists

By

Ricardo A. Valls, P. Geo., M. Sc

Special Edition for Wikibooks

January 2nd, 2008

Ricardo A. Valls, P. Geo., M. Sc.
Valls Geoconsultant

WHY, AND HOW, WE SHOULD USE COMPOSITIONAL DATA ANALYSIS.

A Step-by-Step Guide for the Field Geologists

Toronto, Ontario
January 2nd, 2008

Mente et Maleo… and Computers.

ISBN 1-89664-30-X

Production:	Wikibooks
Editing:	Somerset Parker.
Design and layout:	Ricardo A. Valls
IllustRations:	Ricardo A. Valls

Picture in the cover.

A comparison between correlations obtained by processing the original data using standard statistical methods (left) with the same methods applied to data previously transformed using the ARL method of the CoDaPack software. It is clear that the transformed data better maps the targeted anomaly than the initial dataset.

Cette publication est disponible en anglais seulement.
Esta publicación está disponible sólo en Inglés.

Table of Contents

List of Figures

List of Tables

List of Equations

The most recent link to the CoDaPack freeware is http://ima.udg.es/CoDaPack.

Summary

Compositional data arise naturally in several branches of science, including geology. In geochemistry, for example, these constrained data seem to occur typically, when one normalizes raw data or when one obtains the output from a constrained estimation procedure, such as parts per one, percentages, ppm, ppb, molar concentRations, etc.

Compositional data have proved difficult to handle statistically because of the awkward constraint that the components of each vector must sum to unity. The special property of compositional data (the fact that the determinations on each specimen sum to a constant) means that the variables involved in the study occur in constrained space defined by the simplex, a restricted part of real space.

Pearson was the first to point out dangers that may befall the analyst who attempts to interpret correlations between Ratios whose numerators and denominators contain common parts. More recently, Aitchison, Pawlowsky-Glahn, S. Thió, and other statisticians have develop the concept of Compositional Data Analysis, pointing out the dangers of misinterpretation of closed data when treated with “normal” statistical methods

It is important for geochemists and geologists in general to be aware that the usual multivariate statistical techniques are not applicable to constrained data. It is also important for us to have access to appropriate techniques as they become available. This is the principal aim of this book.

From a hypothetical model of a copper mineralization associated to a felsic intrusive, with specific relationships between certain elements, I will show how “normal” correlation methods fail to identify some of such embedded relationships and how we can obtain other spurious correlations. From there, I will test the same model after transforming the data using the ***CRL***, ***ARL***, and ***IRL*** transformations with the aid of the *CoDaPack* software.

Since I addressed this publication to geologists and geoscientists in general, I have kept to a minimum the mathematical formulae and did not include any theoretical demonstRation. The “mathematical curios geologist”, if such category exists, can find all of those in a list of recommended sources in the reference section.

So let us start by introducing the model of mineralization that we will be testing.

The Model

Figure 1 shows a very simplistic version of a copper mineralization associated to a zone of fractures within a granodiorite intrusive.

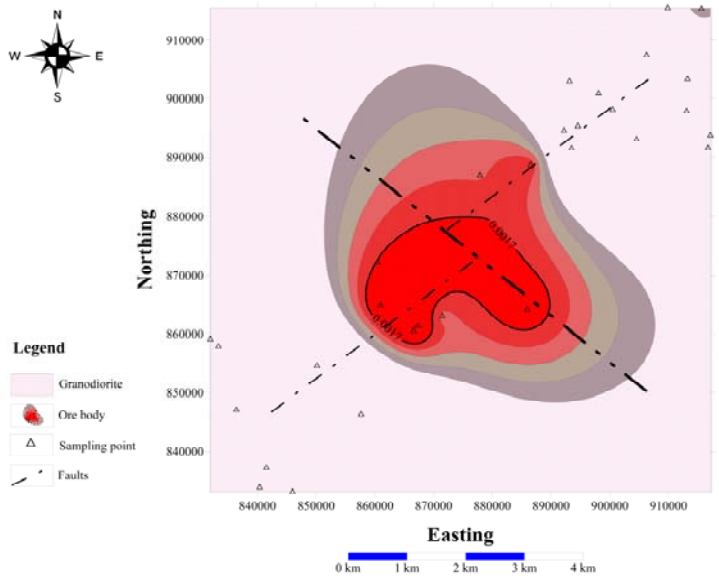


Figure 1. Ore deposit model of a copper mineralization associated to a tectonic zone within a granodiorite intrusive.

Table 1 shows the results of a random sampling using existing access to the top of the intrusive.

Table 1. Chemical composition of the sampling of the granodiorite intrusive.

Sample	Easting	Northing	SiO_2	TiO_2	Al_20_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5
259021	845962.7	833189	56.00	2.20	6.00	3.62	5.11	2.08	5.03	7.08	3.00
259022	831978.5	859092	57.00	1.25	4.00	2.54	5.40	2.19	5.08	7.19	3.74
259023	857644.3	846310	55.00	2.05	7.00	3.86	5.55	2.29	5.19	7.29	3.46
259024	840369.6	833920	62.00	2.19	0.00	2.47	5.63	2.29	5.29	7.29	3.80
259025	833367.4	857854	56.04	1.14	6.00	3.66	5.69	2.43	5.29	7.43	3.14
259026	841526.8	837288	58.36	1.48	2.00	2.75	5.89	2.45	5.43	7.45	3.70
259027	850152.3	854634	61.00	1.60	1.00	2.26	5.95	2.49	5.45	7.49	3.17
259028	836406.1	847121	57.34	1.58	4.02	2.94	6.25	2.54	5.49	7.54	3.26
259029	893494.7	891521	62.00	1.25	0.66	2.21	6.28	2.55	5.54	7.55	3.08
259030	898079.8	900849	61.00	1.60	1.85	3.19	6.44	2.60	5.55	7.60	3.96
259031	917124.2	893796	61.00	2.05	1.49	2.55	6.46	2.63	5.60	7.63	3.66
259032	904564.7	893086	59.02	1.15	1.04	2.70	6.47	2.65	5.63	7.65	3.85
259033	900474.9	898141	59.00	1.36	2.47	3.53	6.49	2.68	5.65	7.68	3.44
259034	894560.4	895311	59.00	1.61	2.34	3.56	6.82	2.71	5.68	7.71	3.36
259035	913232.2	903356	57.80	2.15	3.09	2.63	6.98	2.86	5.71	7.86	3.45
259036	893128.5	902967	56.44	1.15	6.00	3.48	6.98	2.91	5.86	7.91	3.40
259037	892202.8	894592	59.00	1.41	2.30	2.59	7.18	2.99	5.91	7.99	3.13
259038	913071.1	897870	57.30	1.06	4.00	3.62	7.22	3.02	5.99	8.02	3.82

Sample	Easting	Northing	SiO_2	TiO_2	Al_20_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5
259039	906249.3	907340	58.09	1.71	3.50	2.51	7.74	3.10	6.02	8.10	3.66
259040	916724.1	891607	57.00	1.16	4.00	3.27	7.83	3.15	6.10	8.15	3.83
259041	909848.3	915457	60.00	1.47	1.44	2.47	7.89	3.38	6.15	8.38	3.27
259042	915608.1	915306	57.47	1.83	3.00	3.57	7.95	3.56	6.38	8.56	3.28
259043	871460	863020	55.66	1.03	7.00	2.23	7.96	3.58	6.56	8.58	3.08
259044	877894.5	886973	59.00	1.13	2.00	3.33	8.07	3.68	6.58	8.68	3.99
259045	886538.3	888754	59.00	1.80	2.80	2.36	8.10	3.71	6.68	8.71	3.60
259046	860434.9	872223	53.00	1.36	5.00	3.16	8.28	3.82	6.71	8.82	3.90
259047	885897.4	864158	57.00	1.43	4.00	2.09	8.30	3.88	6.82	8.88	3.91
259048	867351	861364	54.30	2.16	4.00	3.98	8.57	3.90	6.88	8.90	3.83
259049	860971.4	864924	54.20	1.77	3.36	3.72	8.59	3.91	6.90	8.91	3.82
259050	866588.3	860513	58.50	1.00	2.00	3.10	8.83	1.00	6.91	2.00	3.65

Sample	Easting	Northing	Cu	As	Pb	Co	Ni	Sc	L.O.I.	Other
259021	845962.7	833189	0.0019	0.0019	0.0000	0.0000	0.0000	0.0193	6.19	9.88
259022	831978.5	859092	0.0019	0.0019	0.0000	0.0000	0.0000	0.0176	8.16	11.61
259023	857644.3	846310	0.0018	0.0019	0.0000	0.0001	0.0001	0.0205	5.09	8.31
259024	840369.6	833920	0.0018	0.0019	0.0000	0.0001	0.0001	0.0001	5.97	9.03
259025	833367.4	857854	0.0017	0.0017	0.0000	0.0001	0.0001	0.0183	6.60	9.18
259026	841526.8	837288	0.0016	0.0017	0.0000	0.0002	0.0002	0.0105	8.01	10.48
259027	850152.3	854634	0.0015	0.0016	0.0000	0.0002	0.0002	0.0034	7.40	9.59
259028	836406.1	847121	0.0015	0.0016	0.0000	0.0002	0.0003	0.0137	6.97	9.03
259029	893494.7	891521	0.0013	0.0014	0.0000	0.0002	0.0003	0.0003	7.00	8.88
259030	898079.8	900849	0.0011	0.0012	0.0000	0.0002	0.0003	0.0008	4.00	6.22
259031	917124.2	893796	0.0011	0.0011	0.0000	0.0003	0.0003	0.0011	4.00	6.92
259032	904564.7	893086	0.0010	0.0011	0.0000	0.0003	0.0003	0.0079	8.00	9.84
259033	900474.9	898141	0.0009	0.0010	0.0000	0.0003	0.0003	0.0072	5.00	7.69
259034	894560.4	895311	0.0009	0.0009	0.0000	0.0003	0.0004	0.0062	4.00	7.20
259035	913232.2	903356	0.0008	0.0009	0.0000	0.0003	0.0004	0.0106	5.00	7.47
259036	893128.5	902967	0.0007	0.0007	0.0000	0.0004	0.0005	0.0136	3.00	5.88
259037	892202.8	894592	0.0006	0.0007	0.0000	0.0004	0.0005	0.0062	5.00	7.51
259038	913071.1	897870	0.0006	0.0007	0.0000	0.0004	0.0005	0.0100	2.00	5.94
259039	906249.3	907340	0.0006	0.0007	0.0000	0.0005	0.0006	0.0081	1.50	5.57
259040	916724.1	891607	0.0006	0.0006	0.0000	0.0005	0.0006	0.0103	1.56	5.50
259041	909848.3	915457	0.0005	0.0006	0.0000	0.0005	0.0006	0.0022	2.00	5.55
259042	915608.1	915306	0.0005	0.0006	0.0000	0.0006	0.0008	0.0070	1.23	4.39
259043	871460	863020	0.0004	0.0005	0.0000	0.0007	0.0009	0.0127	1.12	4.32

Sample	Easting	Northing	Cu	As	Pb	Co	Ni	Sc	L.O.I.	Other
259044	877894.5	886973	0.0004	0.0005	0.0000	0.0007	0.0009	0.0017	0.35	3.53
259045	886538.3	888754	0.0004	0.0005	0.0000	0.0008	0.0010	0.0015	0.08	3.24
259046	860434.9	872223	0.0002	0.0003	0.0000	0.0008	0.0010	0.0085	1.25	5.96
259047	885897.4	864158	0.0002	0.0003	0.0000	0.0009	0.0010	0.0064	0.08	3.68
259048	867351	861364	0.0001	0.0002	0.0000	0.0009	0.0011	0.0046	0.05	3.46
259049	860971.4	864924	0.0000	0.0001	0.0000	0.0009	0.0011	0.0014	0.89	4.81
259050	866588.3	860513	0.0000	0.0001	0.0000	0.0009	0.0011	0.0009	6.94	13.01

The initial data respond to the following pre-established conditions:

1. Is a close system, meaning that the sum of all the values is equal or very close to 100%
2. There are no zero values present[1].
3. There are no statistical outliers (hurricane values) present.
4. There is a big difference between the concentRations of the major oxides with respect to the trace elements.
5. As shown in Figs. 2-4, there are embedded positive and significant correlation between Cu and As, Ni and Co, and MgO and K_2O.
6. Finally, I introduced significant negative correlations between K_2O and CaO (Fig. 5), SiO_2 and Al_2O_3 (Fig 6), K_2O and Cu (Fig. 7), Na_2O and K_2O (Fig. 8), and Co with Cu (Fig 9).

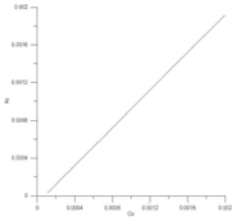


Figure 2. Positive correlation between Cu and As.

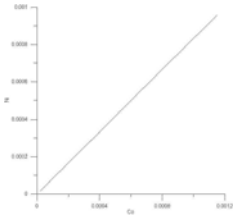


Figure 3. Positive correlation between Ni and Co.

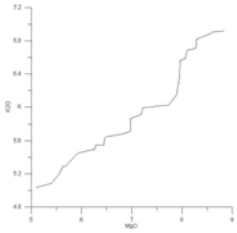


Figure 4. Positive correlation between MgO and K_2O.

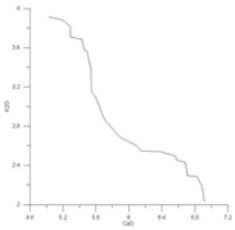


Figure 5. Negative correlation between K_2O and CaO.

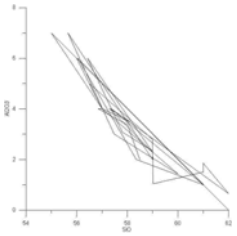


Figure 6. Negative correlation between SiO_2 and Al_2O_3.

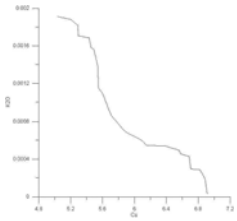


Figure 7. Negative correlation between $K_2O$and Cu.

[1] I will deal with zero values in a separate example to avoid complicating the initial model.

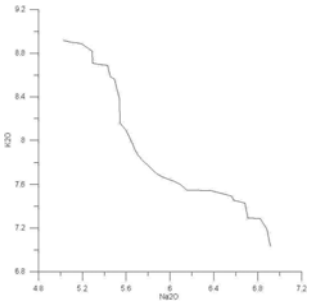


Figure 8. Negative correlation between Na_2O and K_2O.

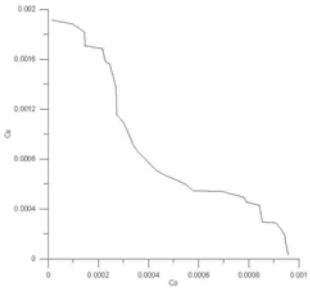


Figure 9. Negative correlation between Co and Cu.

According to these embedded conditions, any correlation analysis must give us two coefficients (Range Correlation Coefficient or RCC):

Equation 1. Range Correlation Coefficient type A for the initial data, according to the embedded correlations.

$$RCC_A = \frac{SiO_2}{Al_2O_3}$$

Equation 2. Range Correlation Coefficient type B for the initial data, according to the embedded correlations.

$$RCC_B = \frac{CaO + Na_2O + Cu + As}{MgO + K_2O + Ni + Co}$$

Normal Processing of the data

Case One

For this case, I will process the initial dataset as a whole, without differentiating between major and trace elements.

Since we established that there are no statistical outliers and no zero values within the data, the first step was to determine their distribution, using the kurtosis and skewness test as described by Kashdan *et al*, (1979). As one can see from Table 2, all elements responded to a Normal Distribution Law, except for CaO and Na_2O, so the next step was to transform those values into Logarithms before testing their correlations.

Table 2. Results of the analysis of kurtosis and skewness.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5
Kurtosis	0.20	0.83	0.94	0.13	0.08	0.78	0.72	7.42	0.52
Skewness	2.64	2.04	2.56	1.68	1.83	3.53	1.93	18.30	1.77

Kurtosis	Cu	As	Pb	Co	Ni	Sc	L.O.I.
Skewness	0.72	0.72	0.33	0.72	0.72	1.18	0.07
	1.93	1.93	2.06	1.93	1.93	2.45	1.64

Using Excel data analysis capabilities, I then determined the correlation analysis of the data (Table 3).

Table 3. Correlation analysis of the initial dataset.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O
SiO_2	1.00							
TiO_2	0.11	1.00						
Al_20_3	-0.92	-0.07	1.00					
Fe_2O_3	-0.46	0.10	0.41	1.00				
MgO	-0.13	-0.21	0.01	0.03	1.00			
CaO	0.10	0.24	0.01	-0.06	-0.98	1.00		
K_2O	-0.16	-0.19	0.04	0.03	0.98	-0.94	1.00	
Na_2O	0.10	0.24	0.01	-0.06	-0.98	1.00	-0.94	1.00
P_2O_5	0.07	-0.04	-0.21	0.10	0.35	-0.34	0.36	-0.34
Cu	0.10	0.24	0.01	-0.06	-0.98	1.00	-0.94	1.00
As	0.10	0.24	0.01	-0.06	-0.98	1.00	-0.94	1.00
Pb	0.30	0.04	-0.30	-0.18	-0.08	0.17	-0.09	0.17
Co	-0.16	-0.19	0.04	0.03	0.98	-0.94	1.00	-0.94
Ni	-0.16	-0.19	0.04	0.03	0.98	-0.94	1.00	-0.94
Sc	-0.80	-0.05	0.80	0.31	-0.47	0.48	-0.46	0.48
L.O.I.	0.17	-0.44	-0.30	-0.52	-0.31	0.27	-0.37	0.27

(continue)

	P_2O_5	Cu	As	Pb	Co	Ni	Sc	L.O.I.
P_2O_5	1.00							
Cu	-0.34	1.00						
As	-0.34	1.00	1.00					
Pb	0.09	0.17	0.17	1.00				
Co	0.36	-0.94	-0.94	-0.09	1.00			
Ni	0.36	-0.94	-0.94	-0.09	1.00	1.00		
Sc	-0.32	0.48	0.48	-0.21	-0.46	-0.46	1.00	
L.O.I.	-0.30	0.27	0.27	0.08	-0.37	-0.37	0.15	1.00

To determine the significance of the obtained correlations, I then proceeded to calculate the critical value of Student using equation 3 (Table 4).

Equation 3. Critical value of Student to determine the significance of the obtained correlations.

$$t_c = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

Where,

tc- critical value of Student

r- correlation

n- amount of data

Table 4. Critical values of Student for the correlation analysis of the initial dataset.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O
SiO_2								
TiO_2	0.59							
Al_20_3	12.84	0.36						
Fe_2O_3	2.76	0.55	2.41					
MgO	0.68	1.12	0.04	0.17				
CaO	0.53	1.30	0.05	0.34	24.01			
K_2O	0.83	1.02	0.24	0.18	23.31	14.85		
Na_2O	0.53	1.30	0.05	0.34	24.01	100.00	14.85	
P_2O_5	0.40	0.21	1.14	0.53	1.95	1.91	2.04	1.91
Cu	0.53	1.30	0.05	0.34	24.01	100.00	14.85	100.00
As	0.53	1.30	0.05	0.34	24.01	100.00	14.85	100.00
Pb	1.66	0.19	1.69	0.98	0.44	0.90	0.50	0.90
Co	0.83	1.02	0.24	0.18	23.31	14.85	100.00	14.85
Ni	0.83	1.02	0.24	0.18	23.31	14.85	100.00	14.85
Sc	7.01	0.27	6.95	1.73	2.80	2.86	2.73	2.86
L.O.I.	0.91	2.61	1.69	3.20	1.75	1.48	2.12	1.48

	P_2O_5	Cu	As	Pb	Co	Ni	Sc
Cu	1.91						
As	1.91	100.00					
Pb	0.46	0.90	0.90				
Co	2.04	14.85	14.85	0.50			
Ni	2.04	14.85	14.85	0.50	100.00		
Sc	1.81	2.86	2.86	1.12	2.73	2.73	
L.O.I.	1.66	1.48	1.48	0.40	2.12	2.12	0.83

It is a common practice in geology that for n > 30 and a probability of 0.05 (95%), if $t_c > 3$, then the correlation is significant. Table 6 shows which correlations are significant from the initial dataset.

Table 5. Results of the significant correlation analysis of the initial dataset.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O
SiO_2								
TiO_2								
Al_20_3	-0.92							
Fe_2O_3	-0.46		0.41					
MgO								
CaO					-0.98			
K_2O					0.98	-0.94		
Na_2O					-0.98	1.00	-0.94	
P_2O_5								
Cu					-0.98	1.00	-0.94	1.00
As					-0.98	1.00	-0.94	1.00
Pb								
Co					0.98	-0.94	1.00	-0.94
Ni					0.98	-0.94	1.00	-0.94
Sc	-0.80		0.80		-0.47	0.48	-0.46	0.48
L.O.I.		-0.44		-0.52			-0.37	

	P2O5	Cu	As	Pb	Co	Ni	Sc	L.O.I.
As		1.00						
Pb								
Co		-0.94	-0.94					
Ni		-0.94	-0.94		1.00			
Sc		0.48	0.48		-0.46	-0.46		
L.O.I.					-0.37	-0.37		

Selecting the proper RCCs

Since the use of Range Correlation Coefficients (RCC) is not a common practice, I will explain in more detail the methodoLogy for their selection from the significant correlations.

a. We start by ranging all correlations from the highest positive to the lowest negative correlation.

b. Start at the bottom of the list and select all the negative significant correlations first to create the initial coefficient.

c. After the first time you use a correlation pair, every time you get the same element, put a dot on top as shown in equation 4.

Equation 4. Example of the calculation of the multiplicative factor on a RCC.

$$\frac{Cu^{*} + As^{*}}{Co^{**}}$$

This example means that you had a significant positive correlation between Cu and As and two significant negative correlations between Co and Cu and Co and As.

d. Once you finish with the negative significant correlations, go back to the top of the list and repeat the same process for the positive correlations.

e. If it is possible, do combine the obtained coefficients.

f. If you get a contradictory result, e.g. an element that has "conflicting correlations" with previous elements, take those elements out from the coefficient and start a new one.

You can reduce the size of the obtained RCC by eliminating the less frequent elements and subtracting their influence from the overall coefficient. For example, let us assume that we obtained the RCC represented in equation 5 as follow:

Equation 5. Hypothetical RCC to demonstrate the reduction process.

$$RCC = (...) + 9CaO + 8Cu + 3Sc + 2L.O.I.$$

If we would like to eliminate the Sc and the L.O.I., we first eliminate the L.O.I. and subtract 2 from every element from equation 5.

Equation 6. Hypothetical RCC without the L.O.I. component.

$$RCC = (...) + 7CaO + 6Cu + Sc$$

Then we would subtract the remaining Sc as shown in equation 7.

Equation 7. Hypothetical RCC without the remaining Sc.

$$RCC = (...) + 6CaO + 5Cu$$

RCC from the initial dataset

According to Table 5, and using the methodoLogy just described, I obtained the RCC shown in equation 8.

Equation 8. RCC1 for the initial dataset.

$$RCC_1 = \frac{2Al_2O_3 + 3Fe_2O_3 + 9Sc + TiO_2 + 8CaO + 7Na_2O + 8Cu + 8As}{SiO_2 + L.O.I + MgO + 8K_2O + 8Co + 8Ni.}$$

In addition, because L.O.I. has a conflicting correlation with some elements from RCC1, I created a separated RCC for this case as shown in equation 9.

Equation 9. RCC 2 for the initial dataset.

$$RCC_2 = \frac{L.O.I.}{K_2O + Co + Ni}$$

Before we use SURFER v. 8.0 to graphically plot these coefficients over our model of mineralization, let us graphically analyze these coefficients using Grapher v. 7.0, also from the Golden Software suite of programs (www.goldensoftware.com).

Analysis of the RCCs from the initial dataset.

The objective of the processing of the data should be to obtain a RCC that will be as similar as possible to our theoretical ones represented by equations 1 and specially equation 2. We can see that we obtained all the embedded correlations, but "masked" by the presence of spurious (inexistent) ones. For example:

1. There is no correlation whatsoever between Al_2O_3 and the other elements (Fig. 10)
2. Same situation with the Fe_2O_3 (Fig. 11).
3. Same situation with the Sc (Fig 12).
4. Same situation with the TiO_2 (Fig. 13).
5. Same situation with the L.O.I. (Fig. 14).

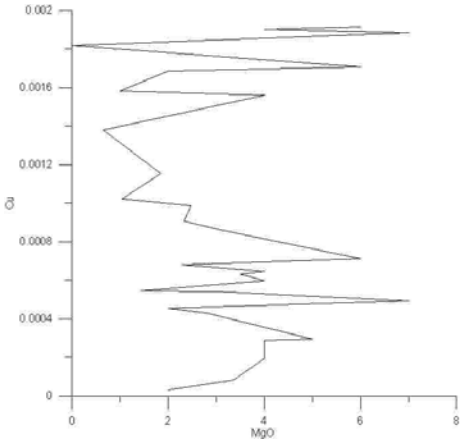


Figure 10. Absence of correlation between Al_2O_3 and Cu.

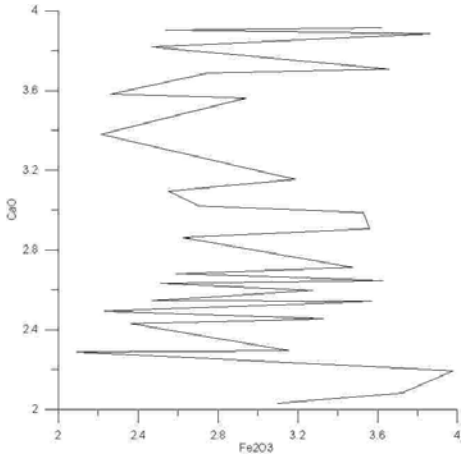


Figure 11. Absence of correlation between Fe2O3 and CaO.

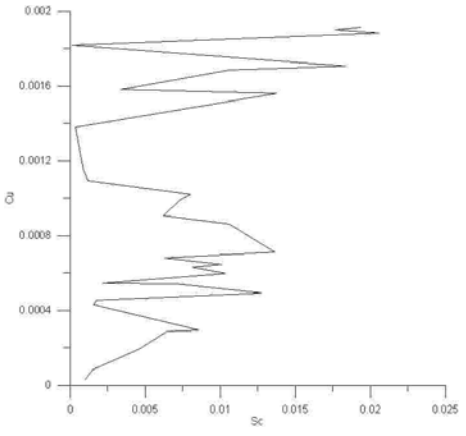


Figure 12. Absence of correlation between Sc and Cu.

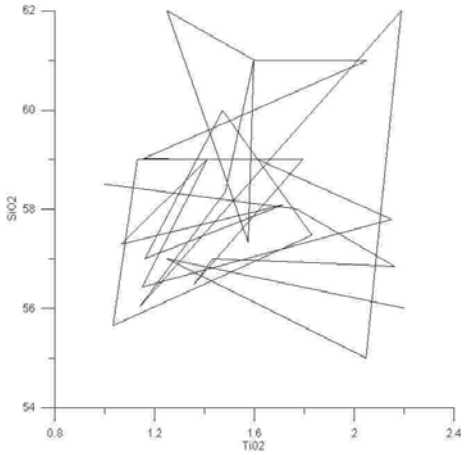


Figure 13. Absence of correlation between SiO_2 and TiO_2.

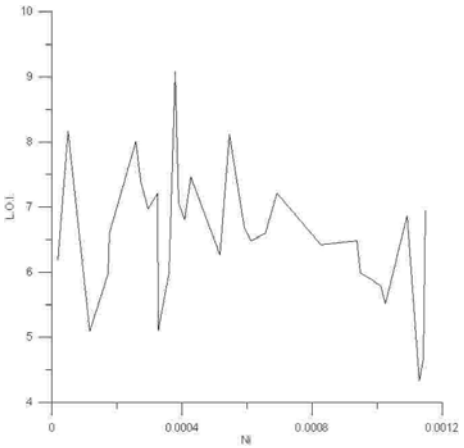


Figure 14. Absence of correlation between L.O.I. and Ni.

Note however, that if we only choose the most strong correlations (r>0.92), then we get a RCC with just the embedded correlations.

Case Two

A more common way to study this kind of data will be to separate the major oxides from the trace elements and treat them separately. It is often intuitively clear for geologists that when mixing percentages with ppm or ppb, the existing correlation between the trace elements is masked or eliminated by the relationships between the major oxides.

So I proceeded to separate the initial dataset into major oxides and trace elements (Tables 7 and 8 in the file ***Initial data.xls*** [worksheet "Processing"], located in the attached CD).

Major oxides

Table 6. Correlation for the major oxides from the initial dataset.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	L.O.I.
SiO_2	1.00									
TiO_2	0.11	1.00								
Al_20_3	-0.92	-0.07	1.00							
Fe_2O_3	-0.46	0.10	0.41	1.00						
MgO	-0.13	-0.21	0.01	0.03	1.00					
CaO	0.10	0.24	0.01	-0.06	-0.98	1.00				
K_2O	-0.16	-0.19	0.04	0.03	0.98	-0.94	1.00			
Na_2O	0.10	0.24	0.01	-0.06	-0.98	1.00	-0.94	1.00		
P_2O_5	0.07	-0.04	-0.21	0.10	0.35	-0.34	0.36	-0.34	1.00	
SiO_2	0.17	-0.44	-0.30	-0.52	-0.31	0.27	-0.37	0.27	-0.30	1.00

Table 7. Critical value of Student for the correlations of the major oxides of the original dataset.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	L.O.I.
SiO_2										
TiO_2	0.59									
Al_20_3	12.84	0.36								
Fe_2O_3	2.76	0.55	2.41							
MgO	0.68	1.12	0.04	0.17						
CaO	0.53	1.30	0.05	0.34	24.01					
K_2O	0.83	1.02	0.24	0.18	23.31	14.85				
Na_2O	0.53	1.30	0.05	0.34	24.01	100.00	14.85			
P_2O_5	0.40	0.21	1.14	0.53	1.95	1.91	2.04	1.91		
L.O.I.	0.91	2.61	1.69	3.20	1.75	1.48	2.12	1.48	1.66	

Table 8. Significant correlations for the major oxides from the initial dataset.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	L.O.I.
SiO_2										
TiO_2										
Al_20_3	-0.92									
Fe_2O_3	-0.46		0.41							
MgO										
CaO					-0.98					
K_2O					0.98	-0.94				
Na_2O					-0.98	1.00	-0.94			
P_2O_5										
L.O.I.		-0.44		-0.52			-0.37			

Analysis of the RCCs from the major oxides of the initial dataset.

As it was the case when we processed the whole dataset, here we obtain a RCC that contains the hypothetical one that we are looking to obtain (equation 1), but it is masked by the presence of other elements, many of them without real correlations between them (equation 10).

Equation 10. RCC3 for the major oxides of the initial dataset.

$$RCC_3 = \frac{SiO_2 + 2L.O.I. + 3CaO + 3Na_2O}{2Al_2O_3 + 3Fe_2O_3 + TiO_2 + 4K_2O + MgO}$$

Trace elements

Table 9. Correlation analysis of the trace elements of the initial dataset.

	Cu	As	Pb	Co	Ni	Sc
Cu	1.00					
As	1.00	1.00				
Pb	0.17	0.17	1.00			
Co	-0.94	-0.94	-0.09	1.00		
Ni	-0.94	-0.94	-0.09	1.00	1.00	
Sc	0.48	0.48	-0.21	-0.46	-0.46	1.00

Table 10. Critical value of Student for the correlations of the trace elements of the original dataset.

	Cu	As	Pb	Co	Ni	Sc
Cu						
As	100.00					
Pb	0.90	0.90				
Co	14.85	14.85	0.50			
Ni	14.85	14.85	0.50	100.00		
Sc	2.86	2.86	1.12	2.73	2.73	

Table 11. Significant correlations for the trace elements from the initial dataset.

	Cu	As	Pb	Co	Ni	Sc
Cu						
As	1.00					
Pb						
Co	-0.94	-0.94				
Ni	-0.94	-0.94		1.00		
Sc	0.48	0.48		-0.46	-0.46	

Analysis of the RCCs from the Trace Elements of the initial dataset.

As it was the case when we processed the whole dataset, here we obtain a RCC that contains the hypothetical one that we are looking to obtain (equation 2), but it is masked by the presence of other elements, many of them without real correlations between them (equation 11).

Equation 11. RCC 4 for the trace elements of the initial dataset.

$$RCC_4 = \frac{4As + 4Sc + Cu}{4Ni + 4Co}$$

Graphical representation of the RCCs

Using SURFER v.8 from Golden Software Inc., (you can download the demos from www.goldensoftware.com), I obtained Figures 15 – 18.

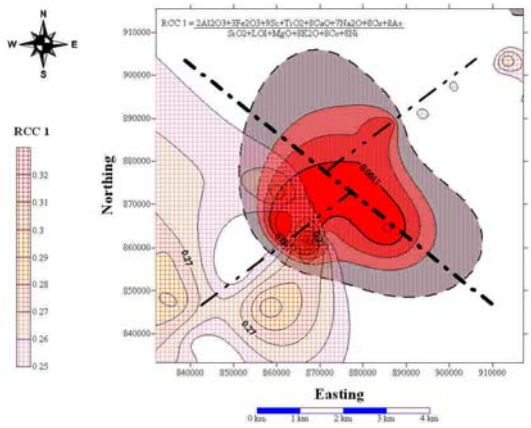


Figure 15. Isopachs of the RCC1 over the granodiorite intrusive.

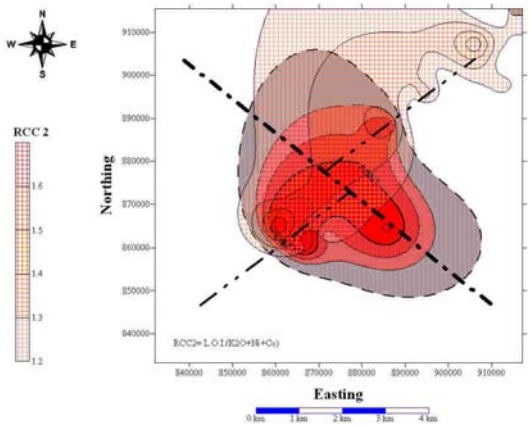


Figure 16. Isopachs of the RCC2 over the granodiorite intrusive.

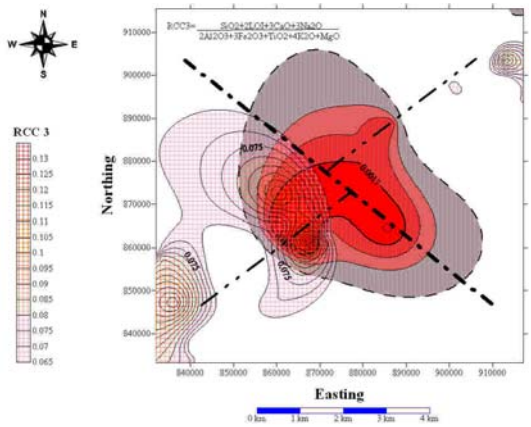


Figure 17. Isopachs of the RCC3 over the granodiorite intrusive.

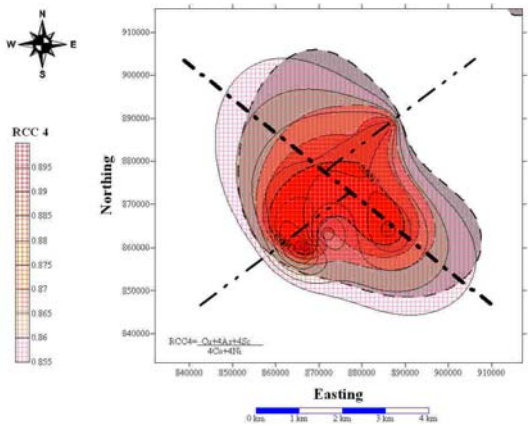


Figure 18. Isopachs of the RCC4 over the granodiorite intrusive.

RCC1 only maps the southwestern border of the mineralized target in a much-dispersed fan that makes it impossible to use as a targeting tool. This was our best RCC from the analysis of the whole dataset.

The only reason why RCC2 partially covers the ore body is the strong and real correlations between Ni, Co, and K_2O. None of these elements have however, a correlation with L.O.I:, therefore, this is a classic example of the formation of a spurious correlation because we applied correlation analysis to a "closed" dataset.

RCC3 contains one of the embedded correlations (SiO_2 vs. Al_2O_3), but it also contains several spurious correlations and, since it is a petrographic association, has little to do with the location of the ore body.

Finally, RRC4 is mostly our main embedded correlation, and although it also contains some spurious components (e.g. correlation with Sc), it is not surprising that it maps perfectly the ore body.

Conclusions and recommendations from the processing of the initial dataset

Closed systems do provoke spurious correlations that mask the effectiveness of the established RCC. This is especially true when processing datasets that contain a combination of major oxides and trace elements. In those cases, I recommend to use only the extremely intense correlations.

A more useful solution is to separate major oxides from trace elements, and concentrate again only on the intense correlations. The disadvantage here is that we do not use the combine information of both groups of elements.

Will the transformation of the data be more efficient in the creation of RCCs that will help us target the mineralized zone within the granodiorite intrusive?

Compositional Data Analysis

The CoDaPack software (which is included in the attached CD and the user guide is presented in Appendix 1) offers three type of transformation, the Centered Log-Ratio transformation (CRL), the Additive Log-Ratio transformation (ARL), and the Isometric Log-Ratio transformation (IRL). The last two require a column with the residual (100 minus the sum of all the other components).

Centered Log-Ratio Transformation (CLR)

Appendix 1 contains the instructions on how to use the CoDaPack software. Since there are no zero values in our dataset, we can proceed directly to the CLR transformation (see table).

Table 12. Results of the CLR transformation of the initial dataset.

SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	Cu	As	Pb	Co	Ni	Sc	L.O.I.	Sum
5.750	1.681	2.933	2.812	2.597	3.045	3.614	3.868	2.976	-4.577	-4.537	-8.323	-5.271	-5.088	-5.241	3.760	0
5.684	2.002	2.335	2.379	1.641	2.982	3.561	3.818	3.004	-4.667	-4.625	-8.384	-5.360	-5.178	-3.406	4.214	0
6.403	3.060	4.222	3.180	4.067	3.104	3.941	4.262	3.611	-5.874	-5.633	-7.486	-6.567	-6.385	-6.691	2.788	0
6.215	2.311	4.034	2.883	3.880	3.020	3.800	4.108	3.213	-5.435	-5.299	-7.920	-6.128	-5.946	-5.866	3.128	0
6.041	2.649	3.722	2.867	3.540	2.885	3.653	3.958	3.229	-5.490	-5.367	-7.996	-6.183	-6.001	-4.804	3.299	0
6.149	2.510	3.830	3.197	3.648	2.973	3.751	4.059	3.415	-5.476	-5.341	-8.855	-6.169	-5.987	-5.028	3.322	0
6.055	2.412	3.736	2.760	3.554	2.842	3.641	3.953	3.097	-5.752	-5.592	-7.930	-6.445	-6.263	-3.736	3.667	0
5.755	1.799	3.064	2.879	2.776	2.953	3.561	3.827	3.062	-4.773	-4.724	-8.338	-5.466	-5.284	-4.685	3.594	0
5.917	2.182	3.225	2.790	2.938	2.907	3.615	3.907	2.981	-5.165	-5.082	-9.606	-5.858	-5.676	-3.230	4.156	0
5.848	2.140	3.363	2.658	3.140	2.903	3.571	3.852	2.938	-5.010	-4.944	-8.466	-5.703	-5.521	-4.367	3.598	0
5.715	2.226	2.753	2.969	2.348	3.016	3.587	3.841	2.994	-4.610	-4.569	-8.724	-5.303	-5.120	-4.855	3.732	0
6.434	2.613	3.084	3.322	2.391	3.124	4.017	4.348	3.710	-7.006	-6.338	-8.128	-7.699	-7.517	-1.647	5.293	0
5.890	2.288	3.198	3.082	2.911	2.798	3.549	3.851	3.025	-5.480	-5.370	-8.367	-6.173	-5.991	-3.267	4.055	0
5.944	2.170	3.252	3.127	2.965	2.840	3.597	3.901	3.103	-5.476	-5.361	-8.985	-6.169	-5.987	-3.061	4.140	0
5.745	2.016	2.112	2.860	1.005	3.022	3.614	3.875	3.072	-4.661	-4.616	-8.815	-5.354	-5.172	-3.051	4.348	0
5.986	2.313	3.131	2.930	2.778	2.807	3.611	3.925	3.228	-5.832	-5.664	-8.176	-6.525	-6.343	-2.629	4.460	0
5.585	2.092	2.893	2.367	2.606	2.811	3.407	3.669	2.789	-4.879	-4.833	-8.711	-5.572	-5.390	-4.947	6.112	0
5.589	2.319	2.162	2.929	1.382	2.906	3.478	3.733	2.892	-4.725	-4.684	-9.647	-5.419	-5.236	-3.831	6.154	0
6.931	3.640	-3.984	4.276	-3.984	3.709	4.571	4.897	4.165	-5.628	-5.286	-7.419	-6.321	-6.139	-0.961	7.529	0
5.739	2.447	2.711	2.648	2.268	2.679	3.424	3.725	2.921	-5.565	-5.460	-8.624	-6.259	-6.076	-2.865	6.287	0
6.149	2.161	1.709	2.933	-4.778	3.400	4.011	4.277	3.255	-4.332	-4.283	-9.136	-5.025	-4.843	-2.233	6.735	0
5.669	1.776	2.319	2.812	1.626	2.756	3.433	3.717	2.970	-5.191	-5.121	-8.467	-5.884	-5.702	-2.944	6.231	0
5.859	2.264	2.661	2.889	2.104	2.723	3.513	3.824	2.993	-5.806	-5.658	-8.487	-6.499	-6.317	-2.479	6.415	0
5.857	1.963	2.185	3.070	0.992	2.875	3.592	3.886	3.047	-5.234	-5.146	-9.372	-5.927	-5.745	-2.473	6.429	0
5.627	2.105	2.693	2.487	2.302	2.672	3.361	3.648	2.864	-5.320	-5.246	-8.270	-6.013	-5.831	-3.248	6.171	0
5.827	1.894	3.141	2.744	2.855	2.718	3.478	3.782	3.097	-5.616	-5.498	-9.262	-6.309	-6.126	-3.080	6.355	0
5.572	2.126	2.426	2.793	1.908	2.739	3.374	3.647	2.711	-5.065	-5.009	-8.585	-5.758	-5.576	-3.431	6.126	0
6.959	3.722	2.934	4.220	-3.974	3.642	4.550	4.884	4.033	-7.440	-6.180	-7.796	-8.133	-7.950	-1.010	7.539	0
6.189	2.295	2.206	3.459	-0.946	2.993	3.829	4.150	3.307	-5.971	-5.733	-7.564	-6.664	-6.482	-1.836	6.768	0
5.668	1.683	2.451	2.907	1.879	2.714	3.409	3.698	2.961	-5.300	-5.224	-8.289	-5.994	-5.811	-2.978	6.225	0

As Table 12 shows, the dataset is now “open”, since the sum of all the components is equal to zero, not 100%.

Once I achieved this transformation, I processed the data following the same steps as with the initial dataset. Tables 13 to 15 show the results of this process.

Table 13. Correlation analysis of the CLR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2	1.00															
TiO_2	0.87	1.00														
Al_20_3	-0.32	-0.33	1.00													
Fe_2O_3	0.87	0.78	-0.49	1.00												
MgO	-0.51	-0.43	0.68	-0.61	1.00											
CaO	0.82	0.71	-0.52	0.77	-0.74	1.00										
K_2O	0.96	0.83	-0.45	0.88	-0.67	0.94	1.00									
Na_2O	0.98	0.85	-0.43	0.89	-0.64	0.92	1.00	1.00								
P_2O_5	0.97	0.85	-0.39	0.88	-0.51	0.83	0.95	0.96	1.00							
Cu	-0.69	-0.65	-0.09	-0.61	0.12	-0.26	-0.53	-0.58	-0.65	1.00						
As	-0.57	-0.54	-0.16	-0.48	-0.05	-0.05	-0.36	-0.42	-0.53	0.96	1.00					
Pb	0.56	0.51	-0.15	0.43	-0.11	0.35	0.48	0.50	0.55	-0.49	-0.47	1.00				
Co	-0.69	-0.65	-0.09	-0.61	0.12	-0.26	-0.53	-0.58	-0.65	1.00	0.96	-0.49	1.00			
Ni	-0.69	-0.65	-0.09	-0.61	0.12	-0.26	-0.53	-0.58	-0.65	1.00	0.96	-0.49	1.00	1.00		
Sc	0.34	0.26	-0.56	0.49	-0.69	0.30	0.37	0.39	0.35	-0.42	-0.35	0.00	-0.42	-0.42	1.00	
L.O.I	0.14	0.18	-0.58	0.33	-0.71	0.20	0.21	0.22	0.16	-0.23	-0.16	-0.07	-0.23	-0.23	0.76	1.00

Table 14. Critical value of Student of the CLR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2																
TiO_2	9.32															
Al_20_3	1.78	1.86														
Fe_2O_3	9.27	6.60	2.99													
MgO	3.12	2.52	4.93	4.12												
CaO	7.61	5.30	3.20	6.47	5.81											
K_2O	18.35	7.97	2.67	9.62	4.72	15.22										
Na_2O	23.88	8.43	2.53	10.06	4.44	12.48	79.36									
P_2O_5	20.04	8.39	2.26	9.85	3.17	7.94	16.26	18.63								
Cu	5.03	4.51	0.47	4.02	0.66	1.42	3.31	3.74	4.58							
As	3.67	3.35	0.83	2.86	0.28	0.28	2.06	2.45	3.31	19.40						
Pb	3.55	3.13	0.79	2.49	0.61	1.99	2.92	3.08	3.45	2.97	2.82					
Co	5.03	4.51	0.47	4.02	0.66	1.42	3.31	3.74	4.58	100.00	19.40	2.97				
Ni	5.03	4.51	0.47	4.02	0.66	1.42	3.31	3.74	4.58	100.00	19.40	2.97	100.00			
Sc	1.90	1.43	3.54	2.94	5.06	1.64	2.14	2.21	1.99	2.44	1.98	0.02	2.44	2.44		
L.O.I	0.77	0.95	3.74	1.87	5.36	1.06	1.16	1.17	0.88	1.26	0.87	0.39	1.26	1.26	6.27	

Table 15. Significant correlations of the CLR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2																
TiO_2	0.87															
Al_2O_3																
Fe_2O_3	0.87	0.78	-0.49													
MgO	-0.51	-0.43	0.68	-0.61												
CaO	0.82	0.71	-0.52	0.77	-0.74											
K_2O	0.96	0.83	-0.45	0.88	-0.67	0.94										
Na_2O	0.98	0.85	-0.43	0.89	-0.64	0.92	1.00									
P_2O_5	0.97	0.85	-0.39	0.88	-0.51	0.83	0.95	0.96								
Cu	-0.69	-0.65		-0.61			-0.53	-0.58	-0.65							
As	-0.57	-0.54		-0.48			-0.36	-0.42	-0.53	0.96						
Pb	0.56	0.51		0.43			0.48	0.50	0.55	-0.49	-0.47					
Co	-0.69	-0.65		-0.61			-0.53	-0.58	-0.65	1.00	0.96	-0.49				
Ni	-0.69	-0.65		-0.61			-0.53	-0.58	-0.65	1.00	0.96	-0.49	1.00			
Sc			-0.56	0.49	-0.69		0.37	0.39		-0.42			-0.42	-0.42		
L.O.I			-0.58		-0.71										0.76	

Using SYSTAT SPSS 10.0 for Windows I constructed a matrix of scatter plots (Fig. 19) to confirm the results from table 15, as well as some individual graphics using Grapher 7.0 which clearly show that all the correlations now are real (Figs. 20 – 23).

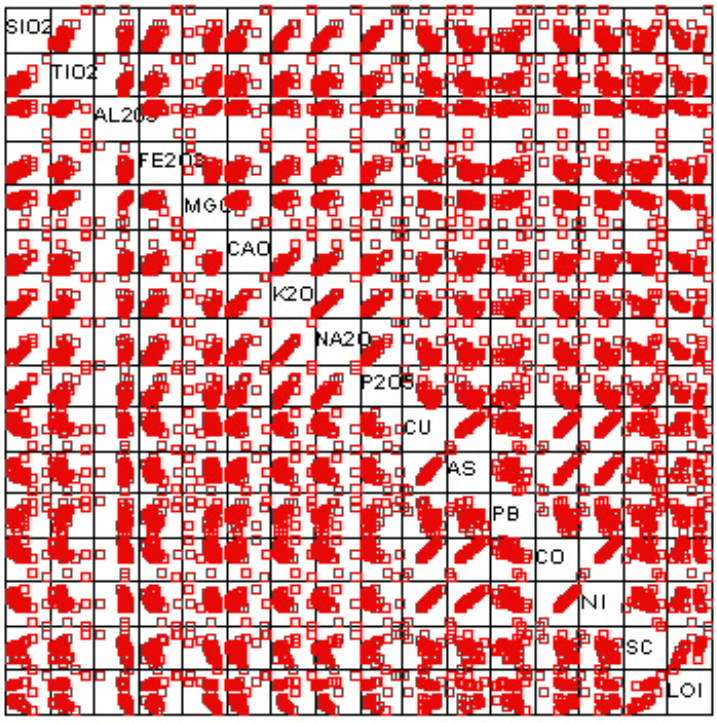


Figure 19. Matrix of scatter plots for the CLR transformed data.

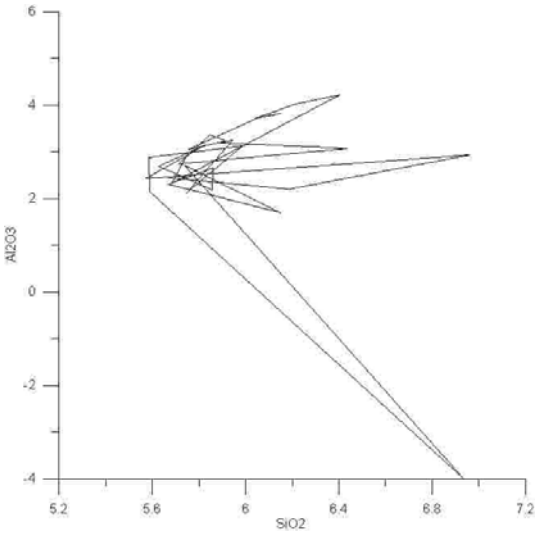


Figure 20. Absence of significant correlation between SiO_2 and Al_2O_3 for the CLR transformed data.

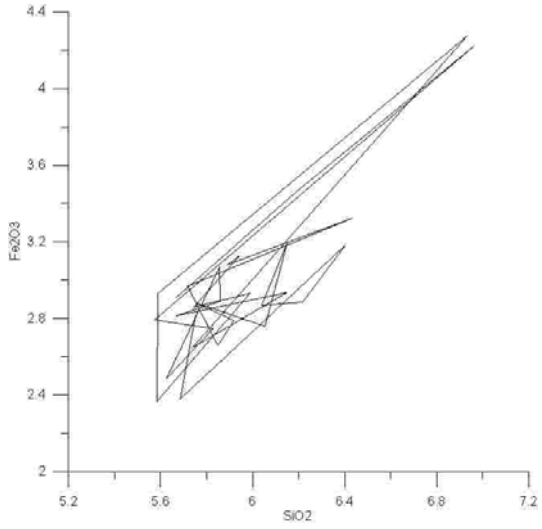


Figure 21. Positive significant correlation between SiO_2 and Fe_2O_3 for the CLR transformed data.

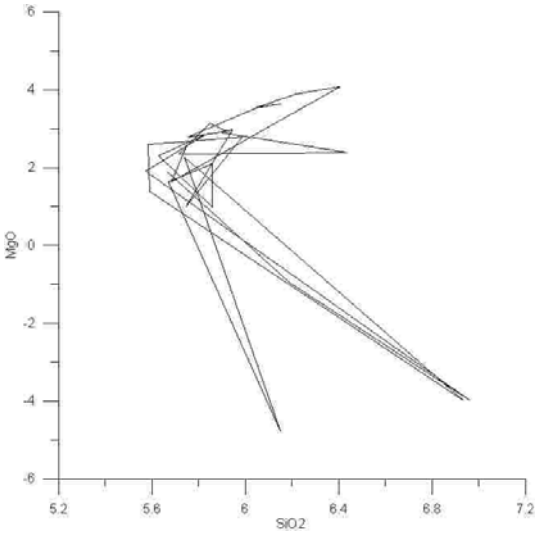


Figure 22. Negative and weak correlation between SiO_2 and MgO for the CLR transformed data.

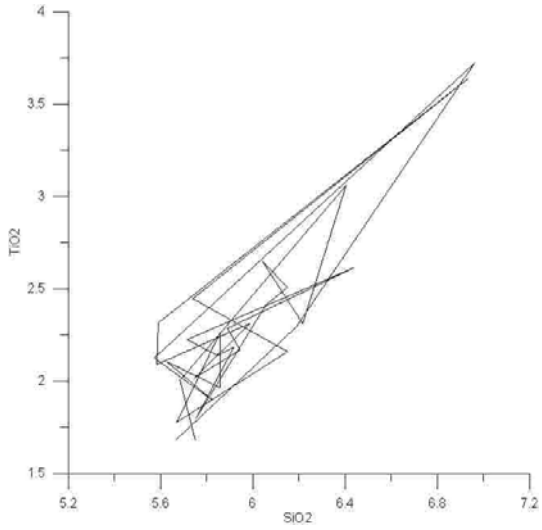


Figure 23. Positive significant correlation between SiO_2 and TiO_2 for the CLR transformed data.

Equation 12 shows the RCC determined for the CLR transformed data and equation 13 shows the same RCC, but reduced by eliminating the SiO_2 and the L.O.I.

Equation 12. RCC5 for the CLR transformed data.

$$RCC_5 = \frac{11Cu + 11Co + 11Ni + 10As + 10MgO + 6Al_2O_3}{14Fe_2O_3 + 14P_2O_5 + 13Na_2O + 12K_2O + 10Pb + 10TiO_2 + 9Sc + 8CaO + 3L.O.i. + SiO_2}$$

Equation 13. RCC5a for the CLR transformed data after reducing the SiO_2 and the L.O.I.

$$RCC_{5a} = \frac{7Cu + 7Co + 7Ni + 6As + 6MgO + 2Al_2O_3}{10Fe_2O_3 + 10P_2O_5 + 9Na_2O + 8K_2O + 6Pb + 6TiO_2 + 5Sc + 4CaO}$$

Fig. 24 shows that RCC5a can effectively target the copper mineralization within the granodiorite intrusive.

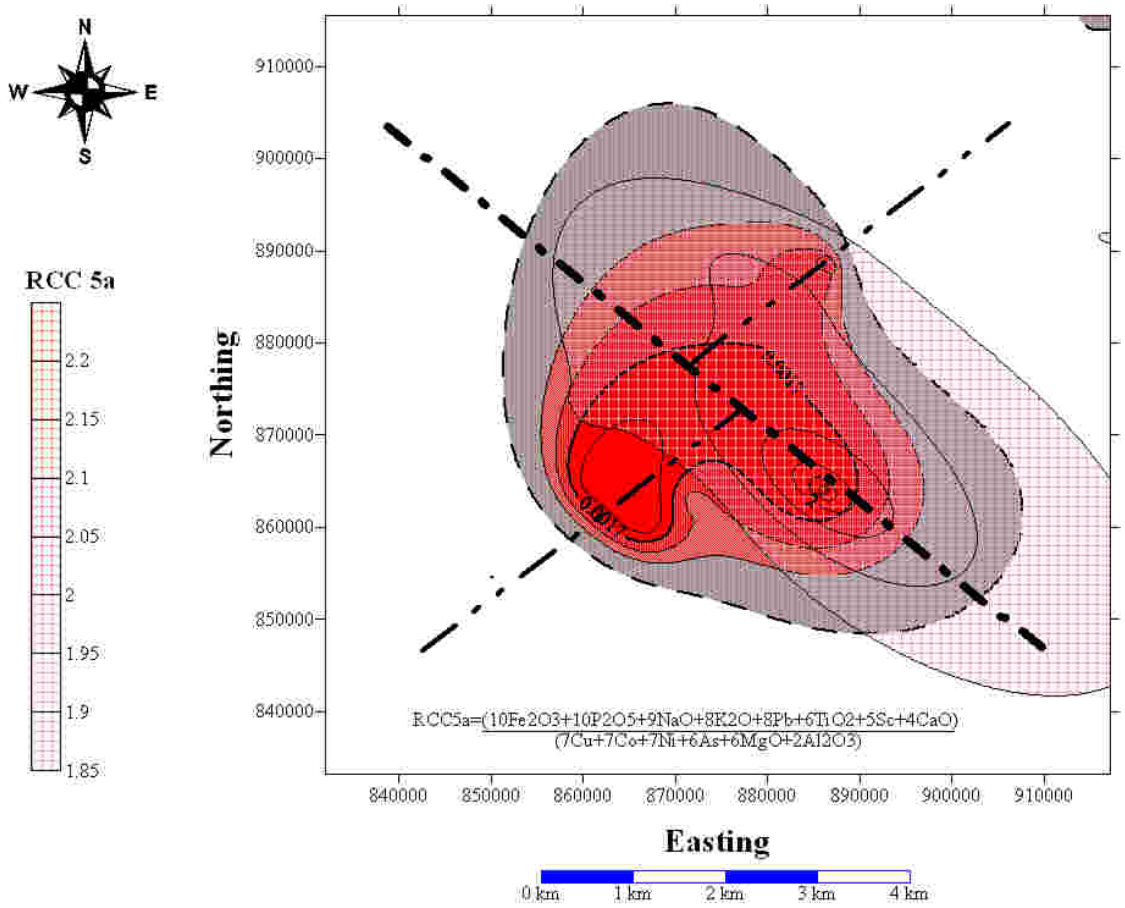


Figure 24. The RCC5a can effectively target the copper mineralization within the granodiorite intrusive.

Now, if we will use only the strongest correlations (r>0.95) then we will obtain RCC6 and RCC7 (equations 14 and 15).

Equation 14. RCC 6 for correlations stronger than ±0.95 for the CLR transformed data.

$$RCC_6 = 3As + 3Ni + 3Co + 2Cu$$

Equation 15. RCC 7 for correlations stronger than ±0.95 for the CLR transformed data.

$$RCC_7 = 3K_2O + 3Na_2O + 3P_2O_5 + 2SiO_2$$

As one can see from Fig. 25, RCC 6 is an almost perfect match with the location of the ore body. The RCC 7 (Fig 26) is similar to RCC 3 and represents a petroLogic association.

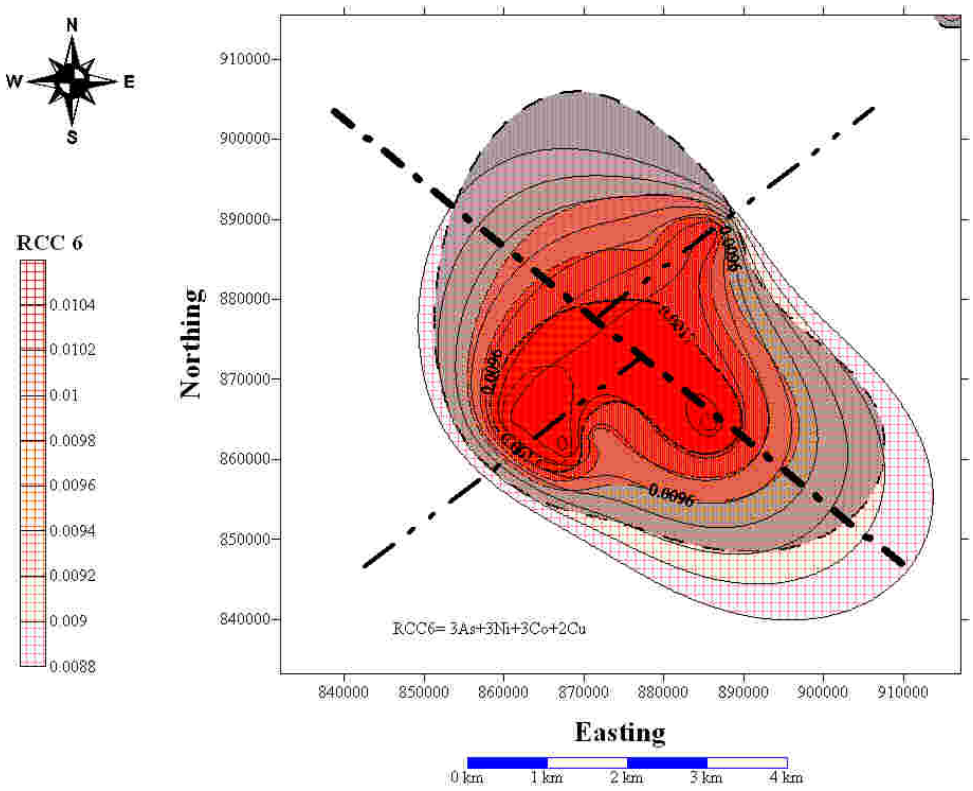


Figure 25. Almost perfect correspondence between the RCC 6 and the location of the ore body.

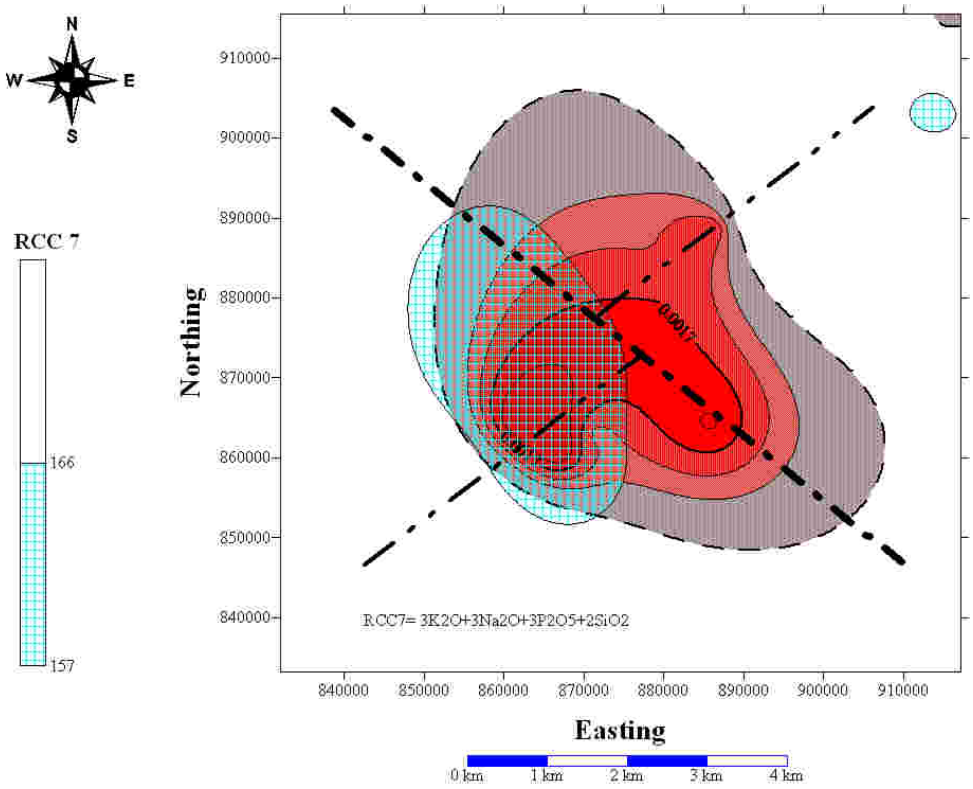


Figure 26. RCC 7 represents a petroLogic association of major oxides.

Additive Log-Ratio Transformation (ARL)

Table 16 shows the results of the transformation of the original dataset[2]. Tables 17 through 19 show the results of the correlation analysis.

Table 16. ALR transformed data.

SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	Cu	As	Pb	Co	Ni	Sc	L.O.I.
2.916	-0.322	0.682	0.176	0.521	-0.235	0.506	0.892	-0.011	-7.368	-7.327	-11.839	-12.176	-11.994	-5.054	0.713
3.048	-0.772	0.392	-0.063	0.691	-0.058	0.631	1.027	0.324	-7.259	-7.218	-11.513	-11.084	-10.902	-5.033	1.105
3.087	-0.204	1.026	0.432	0.794	0.047	0.727	1.112	0.321	-7.194	-7.153	-11.263	-10.165	-9.983	-4.805	0.708
3.307	-0.036	-7.728	0.084	0.908	0.168	0.846	1.219	0.515	-7.129	-7.086	-10.582	-9.663	-9.481	-9.787	0.967
3.533	-0.360	1.299	0.803	1.245	0.568	1.173	1.572	0.651	-6.865	-6.820	-10.220	-9.320	-9.137	-4.492	1.394
3.679	0.006	0.306	0.623	1.386	0.694	1.304	1.685	0.921	-6.774	-6.728	-10.484	-8.832	-8.650	-4.936	1.693
4.529	0.886	0.418	1.234	2.201	1.597	2.115	2.528	1.571	-6.032	-5.983	-9.456	-7.971	-7.789	-5.262	2.420
4.637	1.042	1.979	1.667	2.420	1.777	2.291	2.702	1.771	-5.874	-5.825	-9.709	-7.722	-7.539	-3.701	2.529
6.582	2.678	2.034	3.250	4.293	3.691	4.167	4.588	3.580	-4.131	-4.076	-7.553	-5.761	-5.579	-5.499	4.401
4.813	1.174	1.316	1.861	2.565	1.944	2.416	2.838	2.079	-6.061	-5.995	-10.191	-7.505	-7.322	-6.364	2.089
3.890	0.498	0.181	0.716	1.645	1.022	1.502	1.916	1.078	-7.037	-6.968	-10.146	-8.334	-8.152	-6.955	1.166
6.110	2.176	2.076	3.027	3.899	3.284	3.761	4.172	3.379	-4.853	-4.779	-8.979	-6.026	-5.844	-2.797	4.112
4.957	1.183	1.785	2.141	2.750	2.218	2.611	3.056	2.116	-6.041	-5.964	-9.972	-7.156	-6.974	-4.048	2.489
4.479	0.878	1.253	1.672	2.321	1.782	2.139	2.596	1.614	-6.602	-6.519	-9.778	-7.583	-7.401	-4.677	1.788
5.576	2.285	2.647	2.485	3.462	2.901	3.262	3.714	2.759	-5.538	-5.451	-8.786	-6.421	-6.239	-3.027	3.129
5.188	1.294	2.946	2.400	3.098	2.586	2.923	3.372	2.377	-6.092	-5.987	-10.041	-6.596	-6.414	-3.142	2.253
7.545	3.811	4.299	4.418	5.440	4.908	5.244	5.689	4.610	-3.824	-3.715	-7.978	-4.229	-4.047	-1.601	5.077
4.747	0.762	2.085	1.987	2.676	2.255	2.489	2.975	2.040	-6.643	-6.528	-9.209	-6.914	-6.732	-3.898	1.392
4.589	1.067	1.780	1.449	2.573	2.102	2.323	2.813	1.826	-6.838	-6.720	-9.308	-7.051	-6.869	-4.287	0.933
4.856	0.963	2.199	1.999	2.871	2.401	2.620	3.105	2.157	-6.607	-6.484	-9.280	-6.697	-6.515	-3.757	1.258
5.031	1.323	1.298	1.841	3.003	2.537	2.754	3.235	2.121	-6.578	-6.443	-9.283	-6.521	-6.338	-5.184	1.630
5.085	1.639	2.132	2.306	3.107	2.634	2.887	3.332	2.223	-6.489	-6.353	-9.072	-6.246	-6.063	-3.918	1.241
5.567	1.580	3.494	2.351	3.622	3.171	3.429	3.857	2.673	-6.068	-5.920	-9.718	-5.607	-5.425	-2.815	1.661
5.155	1.199	1.771	2.279	3.165	2.708	2.962	3.391	2.463	-6.618	-6.458	-8.938	-6.066	-5.883	-5.285	0.028
8.201	4.709	5.153	4.983	6.216	5.787	6.023	6.453	5.405	-3.628	-3.460	-6.095	-2.956	-2.773	-2.330	1.598
3.412	-0.253	1.051	0.591	1.555	1.109	1.345	1.774	0.803	-8.691	-8.453	-11.084	-7.623	-7.441	-5.319	-0.335
5.583	1.900	2.926	2.278	3.656	3.258	3.460	3.899	2.903	-6.610	-6.368	-8.485	-5.462	-5.279	-3.508	-0.986
7.715	4.491	5.107	5.102	5.870	5.443	5.650	6.081	5.064	-4.831	-4.489	-7.474	-3.246	-3.064	-1.658	0.725
6.454	3.032	3.673	3.775	4.612	4.224	4.393	4.843	3.800	-6.935	-6.267	-7.917	-4.496	-4.314	-4.049	2.345
2.267	-1.802	-1.109	-0.671	0.376	-1.802	0.132	-1.109	-0.507	-12.176	-10.916	-11.806	-8.753	-8.571	-8.724	0.135

[2] Remember to add a column of the residuals to the original dataset.

Table 17. Correlation analysis for the ALR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2	1.00															
TiO_2	0.99	1.00														
Al_2O_3	0.71	0.69	1.00													
Fe_2O_3	0.99	0.98	0.75	1.00												
MgO	0.99	0.98	0.74	0.99	1.00											
CaO	0.99	0.98	0.74	0.98	0.99	1.00										
K_2O	1.00	0.98	0.73	0.99	1.00	0.99	1.00									
Na_2O	0.99	0.98	0.72	0.98	0.98	1.00	0.98	1.00								
P_2O_5	1.00	0.98	0.72	0.99	1.00	0.99	1.00	0.99	1.00							
Cu	0.81	0.81	0.51	0.78	0.75	0.80	0.77	0.84	0.79	1.00						
As	0.87	0.87	0.56	0.85	0.82	0.85	0.84	0.89	0.85	0.99	1.00					
Pb	0.94	0.93	0.62	0.92	0.94	0.94	0.94	0.93	0.94	0.75	0.81	1.00				
Co	0.92	0.89	0.69	0.92	0.95	0.93	0.94	0.90	0.93	0.57	0.65	0.90	1.00			
Ni	0.92	0.89	0.69	0.92	0.95	0.93	0.94	0.90	0.93	0.57	0.65	0.90	1.00	1.00		
Sc	0.70	0.69	0.88	0.74	0.71	0.74	0.71	0.74	0.71	0.66	0.68	0.59	0.61	0.61	1.00	
L.O.I	0.48	0.45	0.24	0.45	0.42	0.41	0.44	0.44	0.45	0.63	0.64	0.38	0.29	0.29	0.34	1.00

I used SYSTAT SPSS 10.0 for Windows to construct a matrix of scatter plots (Fig. 27) to confirm the results from table 17.

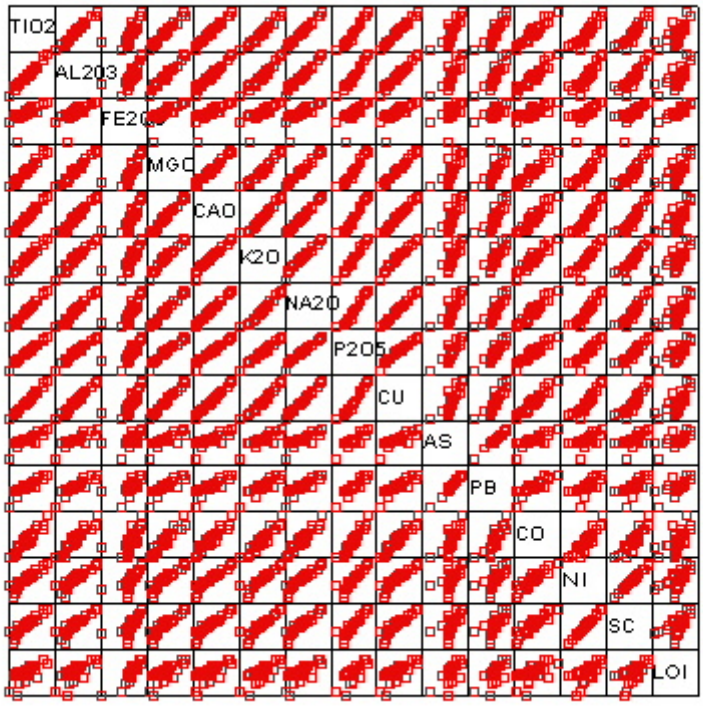


Figure 27. Matrix of scatter plots for the ARL transformed data.

Table 18. Critical values of Student for the ARL transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2																
TiO_2	31.32															
Al_20_3	5.41	5.00														
Fe_2O_3	35.09	25.48	6.03													
MgO	51.30	24.81	5.76	34.14												
CaO	31.02	23.51	5.75	26.76	32.83											
K_2O	72.03	27.59	5.70	37.43	140.28	33.75										
Na_2O	32.38	26.33	5.56	26.42	26.85	70.59	29.60									
P_2O_5	77.26	29.35	5.47	37.49	60.64	33.95	81.61	32.53								
Cu	7.19	7.36	3.14	6.55	5.97	7.01	6.32	8.10	6.72							
As	9.36	9.48	3.59	8.36	7.58	8.65	8.08	10.19	8.63	38.77						
Pb	15.21	13.93	4.20	12.65	14.87	14.24	14.93	13.82	14.98	5.93	7.28					
Co	12.63	10.27	4.99	12.12	16.35	13.16	14.79	11.07	13.57	3.65	4.53	11.04				
Ni	12.63	10.27	4.99	12.12	16.35	13.16	14.79	11.07	13.57	3.65	4.53	11.04	100.00			
Sc	5.26	5.04	9.92	5.78	5.30	5.78	5.34	5.86	5.27	4.60	4.87	3.83	4.05	4.05		
L.O.I	2.91	2.66	1.33	2.69	2.47	2.35	2.56	2.61	2.63	4.28	4.45	2.17	1.60	1.60	1.90	

Table 19. Significant correlations of the ALR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2																
TiO_2	0.99															
Al_20_3																
Fe_2O_3	0.99	0.98														
MgO	0.99	0.98		0.99												
CaO	0.99	0.98		0.98	0.99											
K_2O	1.00	0.98		0.99	1.00	0.99										
Na_2O	0.99	0.98		0.98	0.98	1.00	0.98									
P_2O_5	1.00	0.98		0.99	1.00	0.99	1.00	0.99								
Cu																
As										0.99						
Pb																
Co					0.95											
Ni					0.95								1.00			
Sc																
L.O.I																

Equations 16 and 17 shows the RCCs determined for the ALR transformed data. It is interesting to note that all the correlations here are positive.

Equation 16. RCC 8 of the ARL transformed data.

$$RCC_8 = 7TiO_2 + 7MgO + 7Fe_2O_3 + 7CaO + 7Na_2O + 7K_2O + 7P_2O_5 + 2Co + 2Ni$$

Equation 17. RCC 9 of the ARL transformed data.

$$RCC_9 = Cu + As$$

Figures 28 and 29 show the spatial behavior of these RCCs with respect to the location of the ore body. If we combine both RCCs, we obtain equation 18 (Fig. 30).

Equation 18. Combination of RCC 9 and RCC 8 for the ARL transformed data.

$$RCC_{9/8} = \frac{Cu + As}{7TiO_2 + 7MgO + 7Fe_2O_3 + 7CaO + 7Na_2O + 7K_2O + 7P_2O_5 2Co + 2Ni}$$

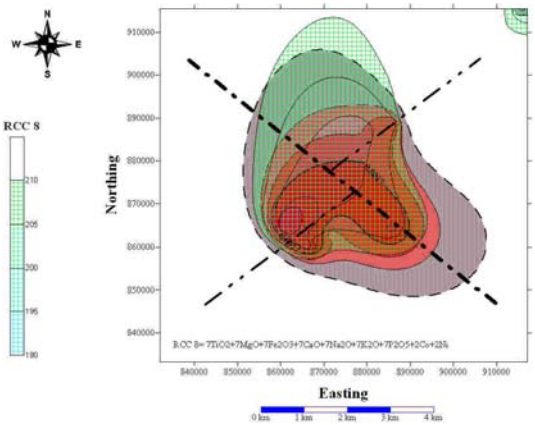


Figure 28. A perfect correspondence of the RCC 8 with the ore body.

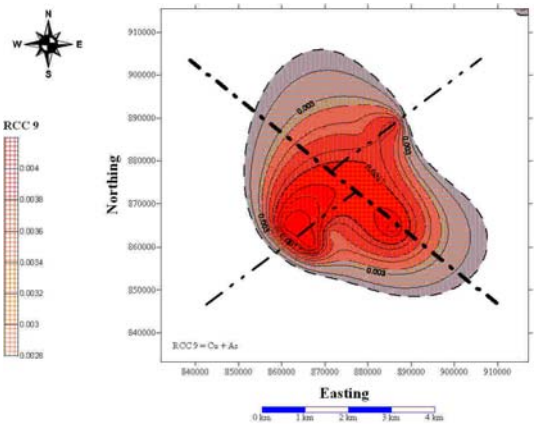


Figure 29. Another perfect spatial correspondence between RCC 9 and the location of the ore body.

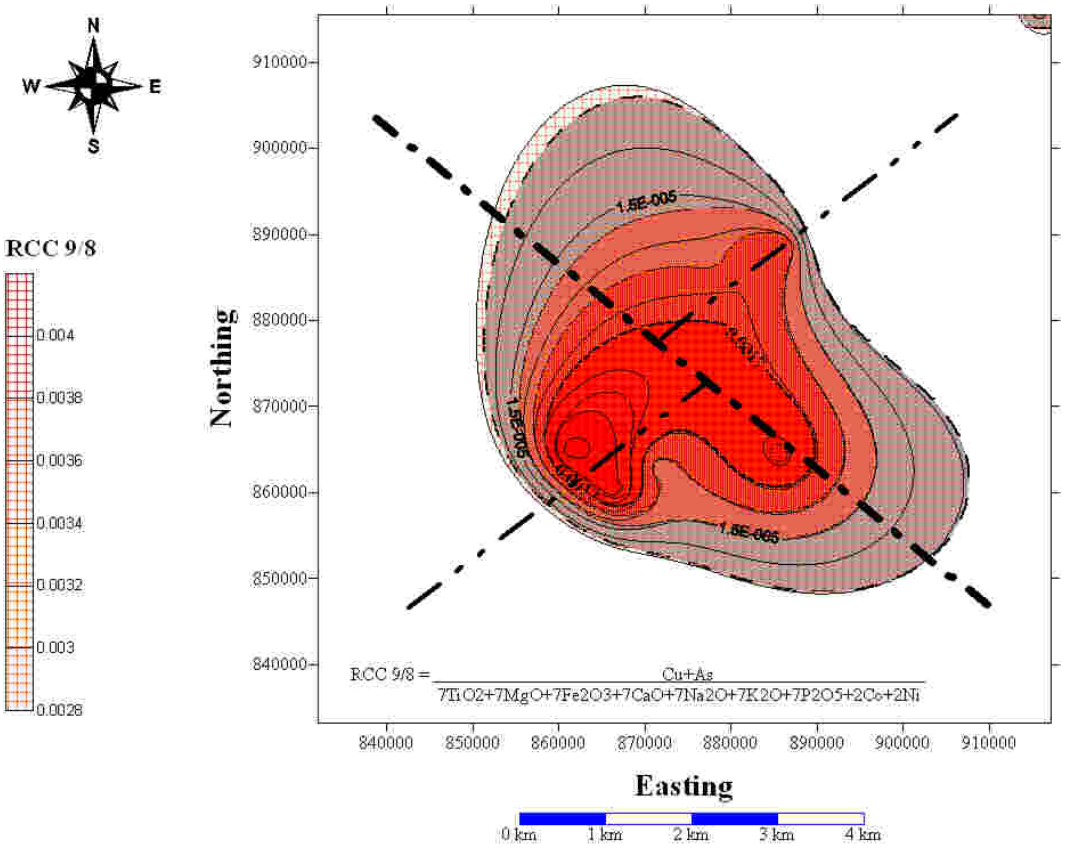


Figure 30. Best possible spatial correspondence between RCC 9/8 and the location of the copper mineralization within the granodiorite intrusive.

Isometric Log-Ratio Transformation (IRL)

Table 20 shows the results of the transformation of the original dataset[3]. Tables 21 through 23 show the results of the correlation analysis.

Table 20. IRL transformed data.

SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
2.289	0.502	0.793	0.306	0.940	0.108	-0.267	0.615	7.530	6.773	10.502	9.984	9.068	1.737	-3.959	-3.027
2.701	0.610	0.825	-0.036	0.655	-0.085	-0.444	0.271	7.437	6.688	10.218	8.987	8.144	1.912	-4.155	-2.831
2.327	0.339	0.755	0.260	0.894	0.126	-0.251	0.525	7.599	6.834	10.174	8.304	7.512	1.991	-3.475	-2.578
2.364	7.645	-1.359	-1.790	-0.786	-1.292	-1.468	-0.631	6.687	6.009	8.832	7.241	6.528	6.374	-4.451	-3.242
2.753	0.235	0.595	0.066	0.672	0.007	-0.367	0.545	7.618	6.848	9.506	7.879	7.119	2.140	-3.697	-2.121
2.598	1.255	0.613	-0.208	0.462	-0.175	-0.508	0.273	7.544	6.780	9.785	7.414	6.689	2.639	-3.950	-2.068
2.576	1.869	0.615	-0.389	0.234	-0.281	-0.631	0.346	7.523	6.758	9.494	7.307	6.589	3.693	-3.983	-1.394
2.542	0.702	0.767	-0.080	0.522	-0.035	-0.414	0.513	7.711	6.928	10.043	7.329	6.609	2.445	-3.746	-1.065
2.760	2.120	0.446	-0.588	0.070	-0.382	-0.725	0.311	7.594	6.816	9.552	7.065	6.365	5.849	-4.115	0.404
2.573	1.370	0.496	-0.245	0.368	-0.126	-0.504	0.271	7.965	7.141	10.536	7.111	6.408	5.040	-3.471	-1.234
2.399	1.644	0.699	-0.289	0.332	-0.164	-0.528	0.324	7.989	7.160	9.579	7.070	6.370	4.774	-3.397	-2.060
2.782	1.688	0.370	-0.493	0.159	-0.308	-0.651	0.173	7.965	7.134	10.534	6.852	6.168	2.799	-4.071	0.165
2.668	1.050	0.434	-0.209	0.315	-0.097	-0.501	0.445	8.136	7.286	10.489	6.943	6.252	2.994	-3.529	-0.900
2.546	1.164	0.460	-0.224	0.309	-0.069	-0.488	0.496	8.238	7.372	9.850	6.952	6.261	3.197	-3.269	-1.336
2.327	1.048	0.881	-0.191	0.356	-0.033	-0.452	0.502	8.321	7.443	9.988	6.915	6.226	2.694	-3.441	-0.197
2.753	0.240	0.643	-0.126	0.364	-0.004	-0.424	0.564	8.540	7.625	10.841	6.663	5.993	2.418	-2.962	-0.596
2.641	1.126	0.693	-0.377	0.178	-0.161	-0.556	0.528	8.473	7.560	10.982	6.501	5.843	3.077	-3.588	1.555
2.818	0.546	0.472	-0.251	0.180	-0.065	-0.511	0.431	8.623	7.690	9.587	6.614	5.947	2.799	-2.504	-1.001
2.491	0.856	0.892	-0.315	0.173	-0.058	-0.509	0.482	8.650	7.712	9.518	6.587	5.923	3.019	-2.230	-1.190
2.753	0.580	0.584	-0.328	0.161	-0.067	-0.511	0.443	8.711	7.761	9.762	6.498	5.841	2.773	-2.261	-0.904
2.622	1.534	0.614	-0.563	-0.034	-0.230	-0.649	0.478	8.679	7.722	9.769	6.332	5.687	4.179	-2.689	-0.944
2.437	1.004	0.559	-0.283	0.200	-0.064	-0.472	0.629	8.828	7.855	9.774	6.275	5.634	3.173	-2.027	-0.701
2.820	0.065	1.036	-0.335	0.139	-0.122	-0.506	0.670	8.892	7.901	10.849	6.031	5.408	2.513	-1.984	-0.251
2.797	1.148	0.372	-0.505	0.006	-0.230	-0.601	0.345	8.924	7.919	9.603	6.074	5.448	4.494	-0.941	-0.857
2.470	1.063	0.898	-0.407	0.060	-0.169	-0.548	0.505	9.021	7.999	9.825	6.022	5.400	4.599	0.498	2.018
2.592	0.431	0.704	-0.317	0.148	-0.093	-0.482	0.490	9.446	8.317	10.112	5.976	5.357	2.938	-2.079	-2.277
2.604	0.666	1.033	-0.433	0.010	-0.179	-0.565	0.441	9.418	8.289	9.593	5.920	5.305	3.227	0.577	-0.414
2.280	0.813	0.580	-0.238	0.196	-0.027	-0.426	0.583	9.909	8.636	10.743	5.819	5.212	3.494	0.961	1.606
2.420	0.874	0.529	-0.339	0.078	-0.091	-0.500	0.543	10.670	9.014	9.809	5.736	5.135	4.524	-1.959	0.435
2.877	1.095	0.395	-0.630	1.474	-0.545	0.689	0.040	11.105	8.844	8.926	5.278	4.710	4.533	-4.338	-3.943

[3] Remember to add a column of the residuals to the original dataset.

Table 21. Critical value of Student of the ILR transformed data

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2	1.00															
TiO_2	-0.22	1.00														
Al_20_3	0.06	-0.90	1.00													
Fe_2O_3	-0.09	-0.86	0.78	1.00												
MgO	0.11	-0.57	0.45	0.66	1.00											
CaO	-0.08	-0.92	0.85	0.95	0.46	1.00										
K_2O	0.22	-0.64	0.51	0.53	0.90	0.43	1.00									
Na_2O	-0.06	-0.88	0.85	0.80	0.29	0.92	0.36	1.00								
P_2O_5	0.14	-0.38	0.27	-0.01	0.15	0.14	0.54	0.28	1.00							
Cu	0.11	-0.44	0.35	0.04	0.06	0.23	0.46	0.39	0.98	1.00						
As	-0.03	-0.47	0.43	0.50	0.11	0.54	0.05	0.54	-0.03	0.06	1.00					
Pb	-0.26	0.03	0.00	0.42	0.33	0.19	-0.09	-0.03	-0.77	-0.78	0.14	1.00				
Co	-0.26	0.03	0.00	0.42	0.33	0.19	-0.09	-0.03	-0.77	-0.78	0.14	1.00	1.00			
Ni	-0.05	0.68	-0.58	-0.76	-0.48	-0.71	-0.34	-0.59	0.11	0.06	-0.49	-0.40	-0.40	1.00		
Sc	-0.16	-0.29	0.32	0.03	-0.32	0.25	-0.04	0.40	0.56	0.67	0.11	-0.55	-0.55	0.05	1.00	
L.O.I	-0.03	-0.20	0.29	0.01	-0.43	0.23	-0.24	0.43	0.27	0.38	0.43	-0.46	-0.46	0.11	0.58	1.00

Figure 31 shows the result of a matrix of scatter plots constructed with SYSTAT SPSS 10.0 for Windows to test the results from table 21.

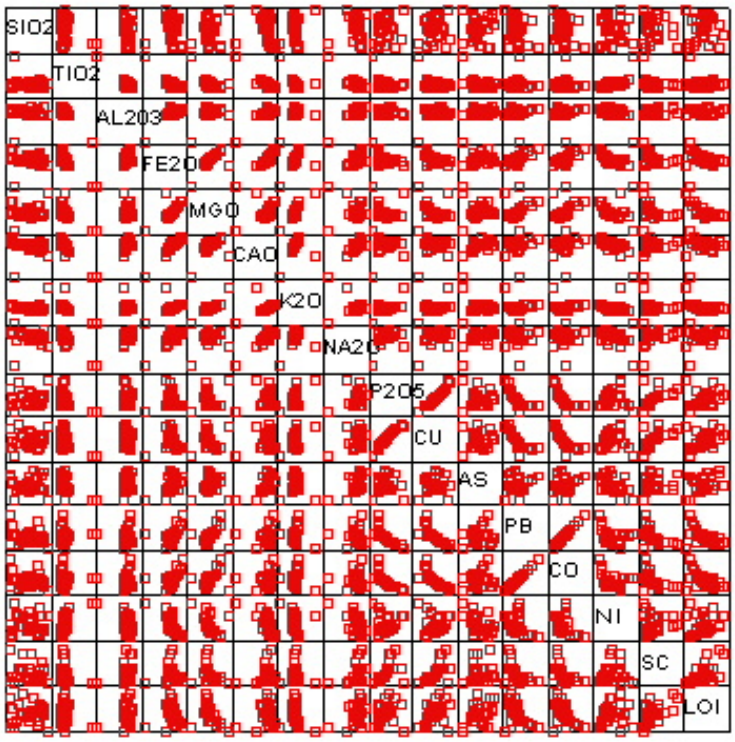


Figure 31. Matrix of scatter plots for the IRL transformed data.

Table 22. Critical value of Student of the ILR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2																
TiO_2	1.21															
Al_20_3	0.33	11.08														
Fe_2O_3	0.45	9.08	6.69													
MgO	0.56	3.67	2.68	4.66												
CaO	0.42	12.02	8.40	15.47	2.74											
K_2O	1.18	4.44	3.10	3.27	11.02	2.51										
Na_2O	0.30	9.57	8.40	7.02	1.59	12.18	2.02									
P_2O_5	0.73	2.15	1.49	0.07	0.79	0.72	3.37	1.55								
Cu	0.60	2.60	1.96	0.22	0.32	1.26	2.72	2.24	28.19							
As	0.18	2.78	2.48	3.06	0.57	3.43	0.27	3.39	0.15	0.31						
Pb	1.45	0.14	0.01	2.44	1.83	1.03	0.50	0.17	6.35	6.69	0.77					
Co	1.45	0.14	0.01	2.44	1.83	1.03	0.50	0.17	6.35	6.69	0.77	100.00				
Ni	0.27	4.85	3.75	6.19	2.87	5.40	1.93	3.91	0.60	0.31	2.99	2.32	2.32			
Sc	0.84	1.57	1.78	0.14	1.81	1.34	0.23	2.29	3.61	4.72	0.56	3.51	3.51	0.27		
L.O.I	0.18	1.09	1.62	0.06	2.49	1.26	1.29	2.53	1.50	2.17	2.52	2.77	2.77	0.56	3.74	

Table 23. Significant correlations of the ILR transformed data.

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	K_2O	Na_2O	P_2O_5	*Cu*	*As*	*Pb*	*Co*	*Ni*	*Sc*	*L.O.I.*
SiO_2																
TiO_2																
Al_20_3		-0.90														
Fe_2O_3		-0.86														
MgO																
CaO		-0.92	0.85	0.95												
K_2O					0.90											
Na_2O		-0.88	0.85	0.80		0.92										
P_2O_5																
Cu									0.98							
As																
Pb																
Co												1.00				
Ni																
Sc																
L.O.I																

Equations 18 through 20 shows the RCCs determined for the ALR transformed data.

Equation 19. RCC 10 for the IRL transformed data.

$$RCC_{10} = \frac{TiO_2}{3caO + 3Na_2O + 2Al_2O_3 + 2Fe_2O_3}$$

Equation 20. RCC 11 for the IRL transformed data.

$$RCC_{11} = MgO + K_2O$$

Equation 21. RCC 12 for the IRL transformed data.

$$RCC_{12} = P_2O_5 + Cu$$

Equation 22. RCC 13 for the IRL transformed data.

$$RCC_{13} = Pb + Co$$

Figures 32 -35 show the result of the use of these RCCs as targeting tools.

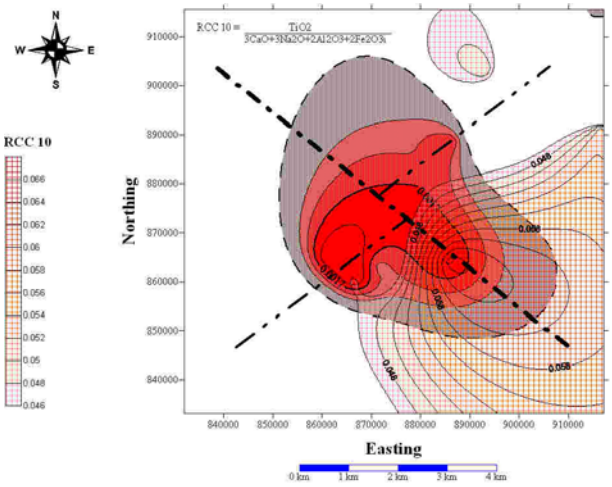


Figure 32. RCC 10 is not useful as an exploRation tool in this case.

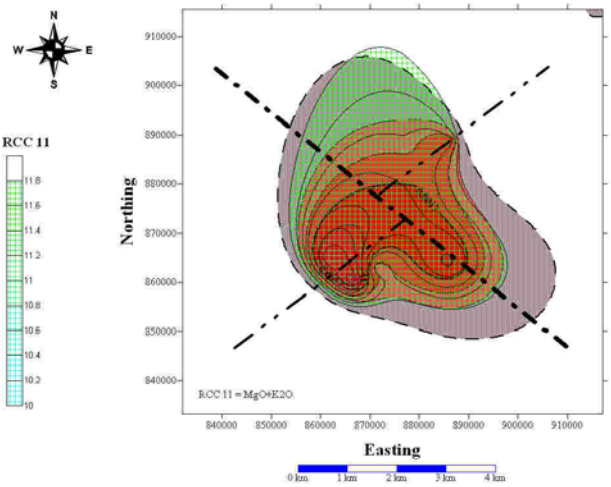


Figure 33. The lower concentRations of RCC 11 perfectly correspond to the limits of the ore body.

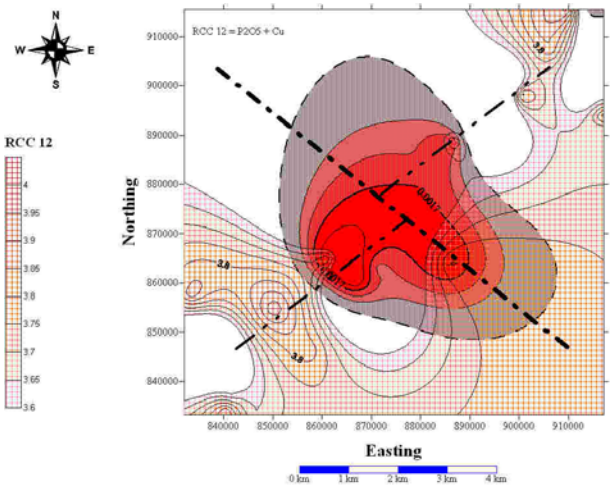


Figure 34. RCC 12 also fails to map the location of the ore body.

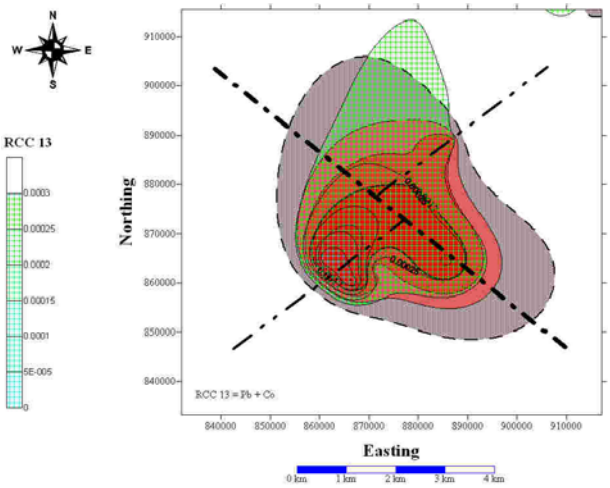


Figure 35. The lower concentRations of RCC 13 are a good approximation to the location of the ore body.

Conclusions and Recommendations from the Compositional Data Analysis

One very important effect of “opening” a dataset by using any of these transformations is that we get rid off all spurious correlations. The transformed data do contain unexpected correlations, but they are real.

Another important point is that we do not need to process the data separately (e.g. separating major oxides from trace elements), but can process the whole dataset taking advantage of the information contained in both groups.

From all the RCCs obtained so far, the RCC 8, RCC 9, and especially the RCC9/8 (ALR) were by far the most efficient one for targeting the copper mineralization.

The CRL transformed data did also provide for useful RCCs, especially if we concentrate in the higher correlations.

Finally, the IRL transformed data was effective for as long as the geochemist will “interpret” the coefficient and not plot them blindly. For example, a geochemist should know that elements like Pb and Co, usually concentrate bellow the ore body (inframinerals), and therefore while using RCC 13, the investigator should concentrate on the lower values as an indication of the location of the ore body. We have a similar situation with RCC 11. The investigator should know that a common effect of sodic metasomatism would be the lixiviation of MgO and K_2O; therefore, the geochemist should be looking for lower values of the RCC.

In general, I can state that transformed data are more effective for the location of the mineralized targets than the non-transformed dataset, and that the ARL method seems to be the most effective for processing this type of data. However, the geochemist should always use his background knowledge to help to decide the most efficient RCC for the studied area.

I recommend the use of the attached software CoDaPack for the processing of any type of “closed” dataset.

Factor Analysis

Factor analysis is a statistical data reduction technique used to explain variability among observed random variables in terms of fewer unobserved random variables called **factors**. It is useful to reduce the number of variables, by combining two or more variables into a single factor, thus "simplifying" the original dataset.

Factor analysis (FA) is especially useful in geochemistry when one has a known target or some other way to understand the meaning of the obtained associations. When failing this, the geologist is usually forced to "plot and see", and then to select the FA that he believes is the most useful for the studied area.

I processed both the initial dataset and the three transformed versions using SYSTAT SSPS 10.0 for Windows, but you can use any other statistical program capable of factor analysis.

Factor Analysis for the Initial Dataset

Figure 36 shows the plot for the initial dataset, while table 24 shows the principal components defined by the software.

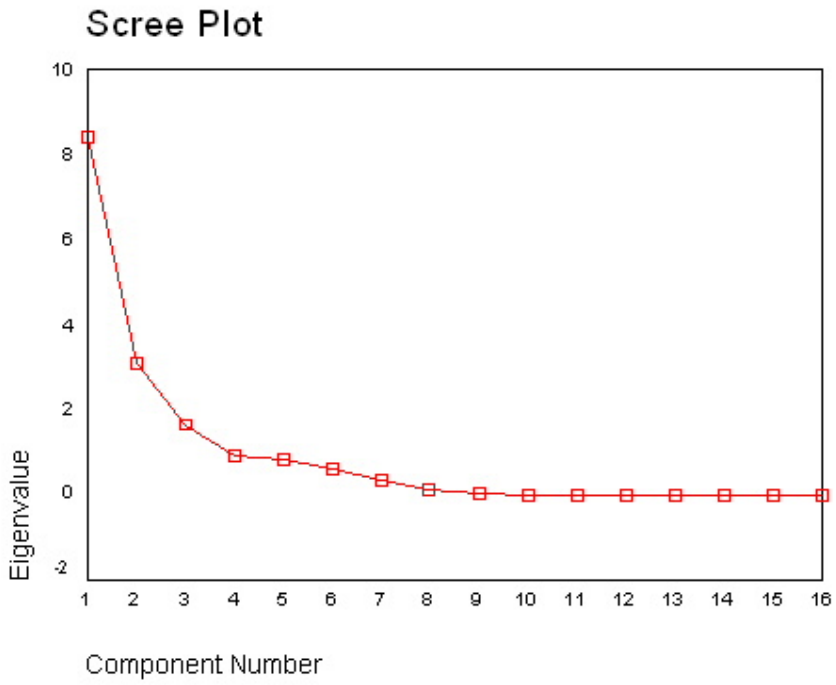


Figure 36. Scree plot for the initial dataset.

Table 24. Principal component analysis (PCA) for the initial dataset.

Component Matrix[a]

	Component		
	1	2	3
SIO2	-.109	-.948	.146
TIO2	-.195	-2.56E-02	.767
FE2O3	2.145E-03	.955	-6.28E-02
AL203	6.064E-02	.609	.458
MGO	.986	3.088E-03	-6.04E-02
K2O	.992	3.313E-02	-2.70E-02
CAO	.992	3.313E-02	-2.70E-02
NA2O	.992	3.313E-02	-2.70E-02
P2O5	.411	-.172	.310
CU	-.968	6.357E-04	9.417E-02
AS	-.968	6.357E-04	9.417E-02
PB	-.107	-.416	.176
CO	.992	3.313E-02	-2.70E-02
NI	.992	3.314E-02	-2.70E-02
SC	-.501	.807	-.216
LOI	-.377	-.310	-.788

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Equations 23 – 25 show the three FA components for the initial dataset.

Equation 23. FA 1 for the initial dataset.

$$FA_1 = \frac{SiO_2 + TiO_2 + Cu + As + Pb + Sc + L.O.I.}{MgO + K_2O + Na_2O + CaO + P_2O_5 + Co + Ni}$$

Equation 24. FA 2 for the initial dataset.

$$FA_2 = \frac{SiO_2 + P_2O_5Pb + L.O.I.}{Fe_2O_3 + Al_2O_3 + Sc}$$

Equation 25. FA 3 for the initial dataset.

$$FA_3 = \frac{Sc + L.O.I.}{SiO_2 + TiO_2 + Al_2O_3 + P_2O_5 + Pb}$$

Figures 37 – 39 show the effectiveness of these FA as a targeting tool for our ore body.

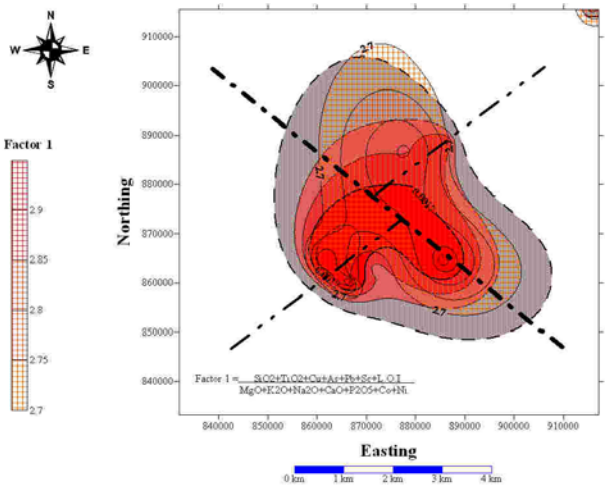


Figure 37. FA1 is a perfect match with the location of our ore body.

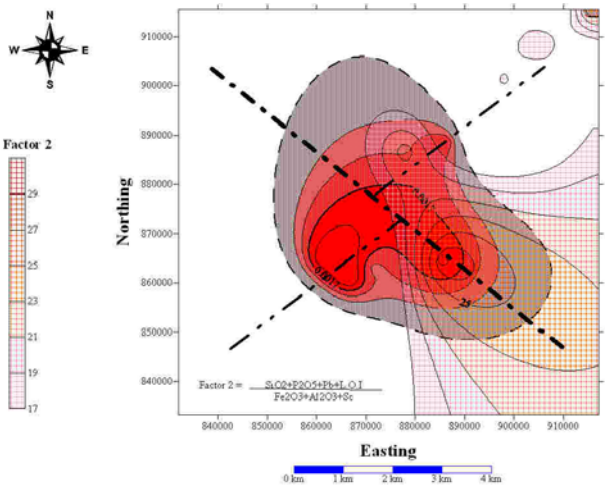


Figure 38. FA2 maps the southeast border of the ore body only.

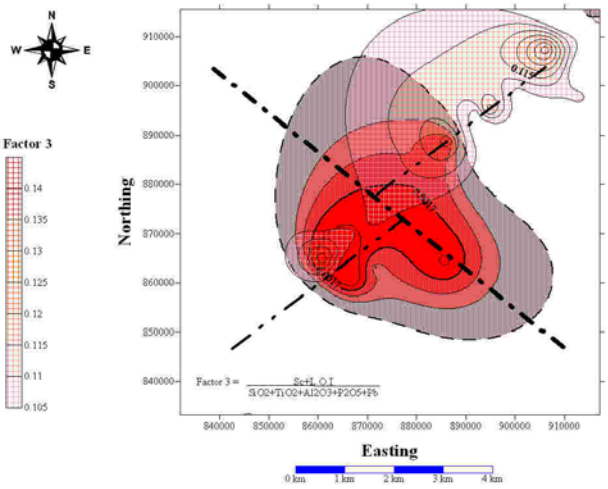


Figure 39. The fragmented distribution of FA3 renders him ineffective as a targeting tool.

<u>Conclusions and recommendations on the use of FA for the initial dataset</u>

For as long as we have a known target to test the obtained FA, this method offers better results than the RCC. It also allows for the combined studied of all the elements together.

FA1 and FA2 do contain the embedded correlations I introduced in the initial dataset, thus their effectiveness, especially FA 1, in mapping the location of the ore body.

The next question will be: ***Will the transformed data be any more effective in helping us locate our target?***

CRL transformed data

Figure 40 shows the scree plot for the CLR transformed dataset, while table 25 shows the principal components defined by SYSTAT.

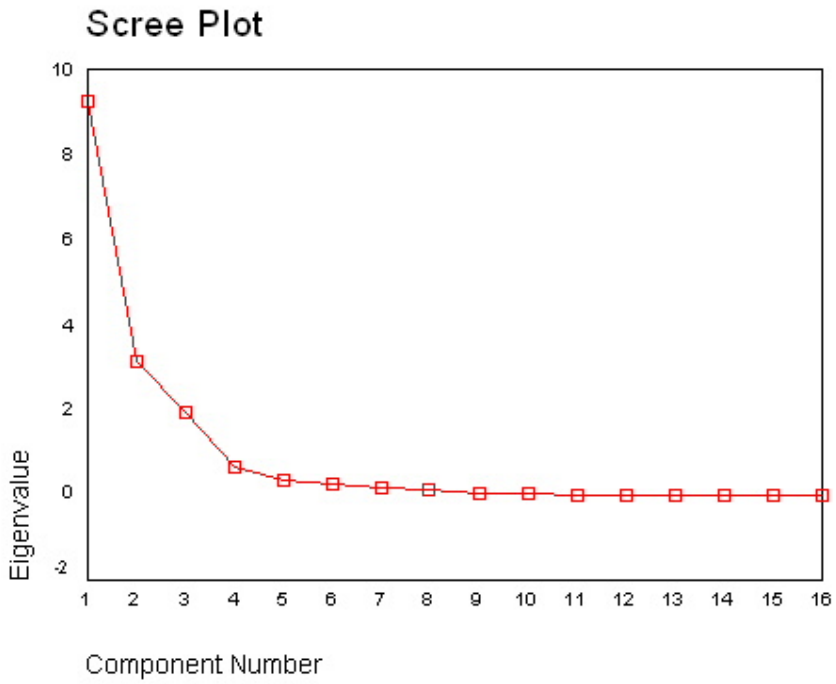


Figure 40. Scree plot for the CLR transformed dataset.

Table 25. Principal component analysis for the CLR transformed dataset.

Component Matrix[a]

	Component		
	1	2	3
SIO2	.955	-6.96E-03	.244
TIO2	.874	-3.23E-02	.207
AL203	-.403	-.746	.170
FE2O3	.910	.161	4.959E-02
MGO	-.597	-.705	.231
CAO	.771	.472	.364
K2O	.924	.231	.272
NA2O	.943	.180	.250
P2O5	.945	4.489E-02	.238
CU	-.782	.570	.242
AS	-.655	.687	.285
PB	.562	-.258	.301
CO	-.782	.570	.242
NI	-.782	.570	.242
SC	.532	.296	-.719
LOI	.361	.418	-.766

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Equations 26 – 28 show the three FA components for the CLR transformed dataset.

Equation 26. FA 4 for the CLR transformed dataset.

$$FA_4 = \frac{Al_2O_3 + MgO + Cu + As + Co + Ni}{SiO_2 + TiO_2 + Fe_2O_3 + CaO + K_2O + Na_2O + P_2O_5 + Pb + Sc + L.O.I.}$$

Equation 27. FA5 for the CLR transformed dataset.

$$FA_5 = \frac{Fe_2O_3 + CaO + K_2O + Na_2O + Cu + As + Co + Ni + Sc + L.O.I.}{Al_2O_3 + MgO + Pb}$$

Equation 28. FA6 for the CLR transformed dataset.

$$FA_6 = \frac{SiO_2 + TiO_2 + Al_2O_3 + MgO + CaO + K_2O + Na_2O + P_2O_5}{Sc + L.O.I.}$$

Figures 41 – 43 show the effectiveness of these FA as a targeting tool for our ore body.

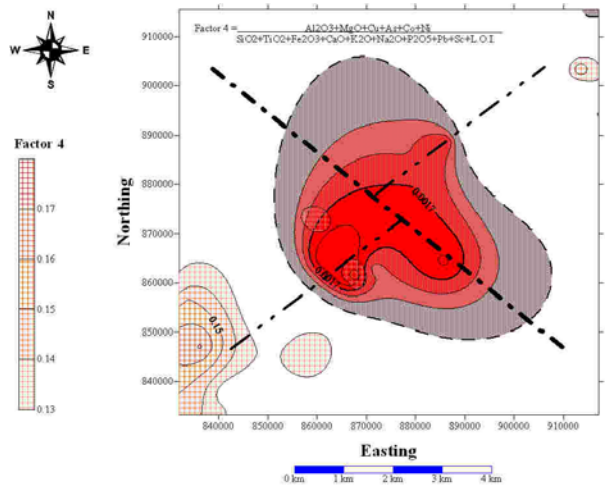


Figure 41. FA4 is not useful as a targeting tool.

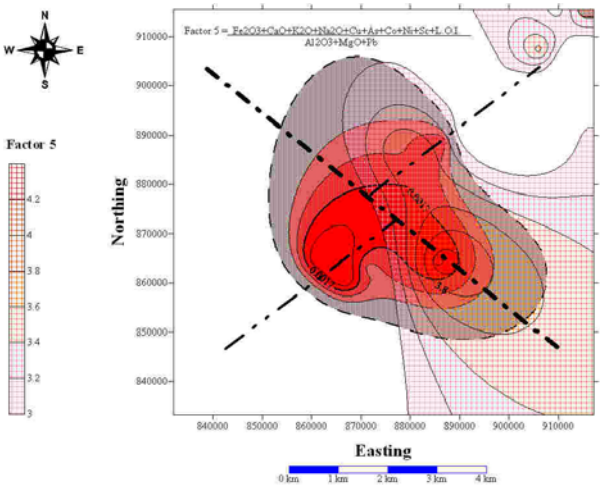


Figure 42. FA5 maps only the southeastern section of the ore body, very similar to FA2.

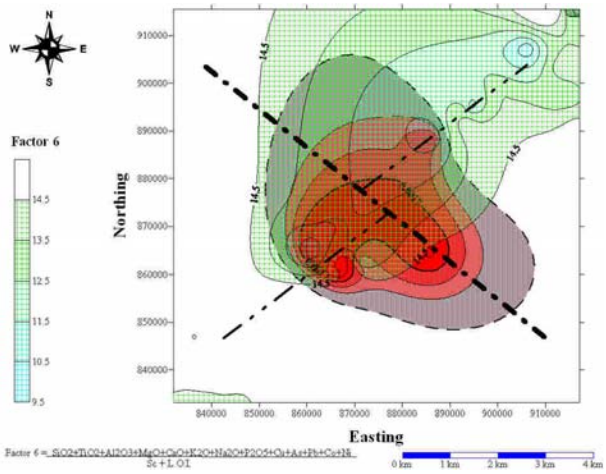


Figure 43. FA6 has the same fragmented distribution as FA3, thus it is not effective as a tool for targeting the ore body.

Factor Analysis for the ALR Transformed Dataset

Figure 44 shows the scree plot for the ALR transformed dataset, while table 26 shows the principal components defined by SYSTAT.

Figure 44. Scree plot for the ALR transformed dataset.

Table 26. Principal component analysis for the ALR transformed dataset.

Component Matrix[a]

	Component	
	1	2
SIO2	.995	-7.41E-03
TIO2	.981	2.411E-03
AL203	.758	-.210
FE2O3	.989	-4.51E-02
MGO	.990	-.101
CAO	.991	-6.67E-02
K2O	.992	-7.52E-02
NA2O	.992	-1.60E-03
P2O5	.993	-4.89E-02
CU	.822	.485
AS	.881	.421
PB	.940	-7.68E-02
CO	.917	-.308
NI	.917	-.308
SC	.760	1.080E-02
LOI	.483	.745

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Although table 26 shows two components, I will analyze only the second, which is a coefficient as shown in equation 29.

Equation 29. FA7 for the ALR transformed dataset.

$$FA_7 = \frac{Cu + As + L.O.I.}{Al_2O_3 + MgO + Ni + Co}$$

This factor contains the embedded relationship from the initial dataset, but because of the presence of other elements, its usefulness as a targeting tool is more limited, as shown in Figure 45.

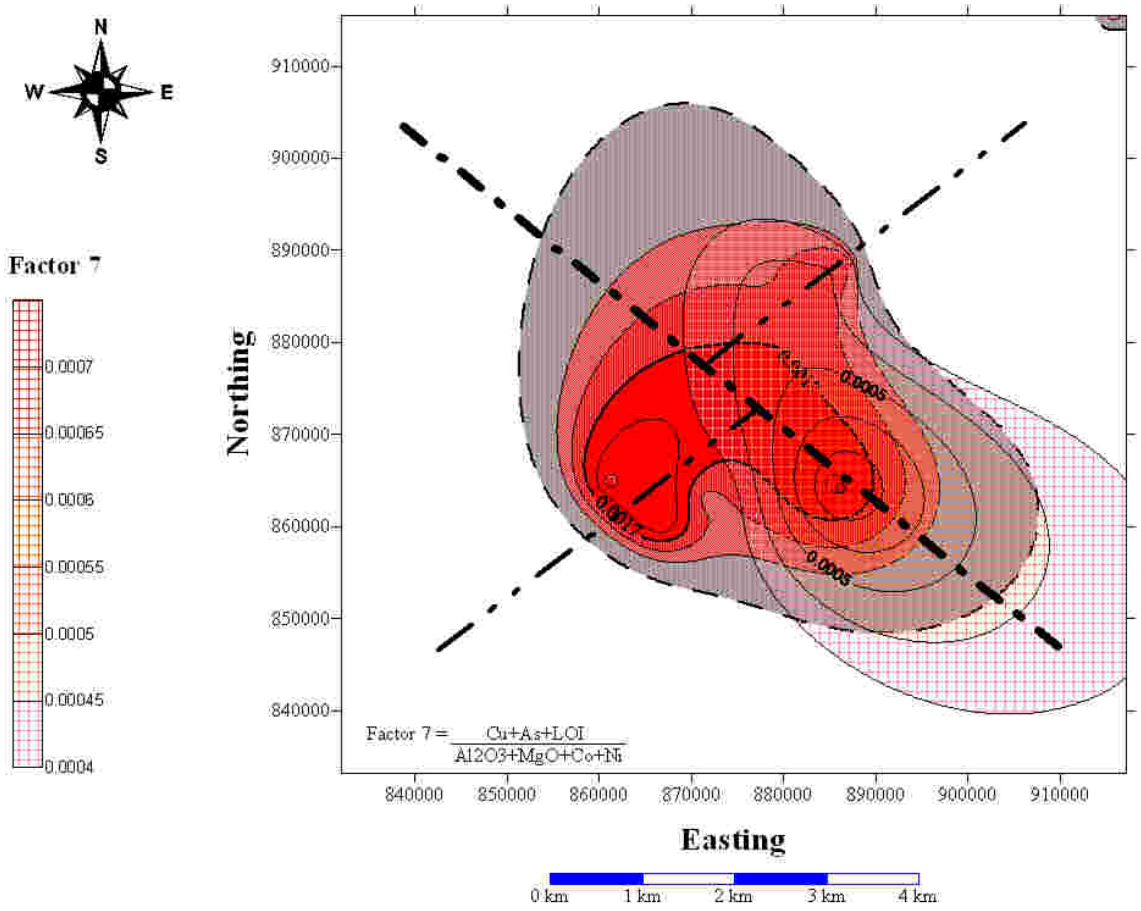


Figure 45. FA7 covers mostly the southeastern part of the ore body.

Factor Analysis of the IRL Transformed Dataset

Figure 46 shows the scree plot for the IRL transformed dataset, while table 27 shows the principal components defined by SYSTAT.

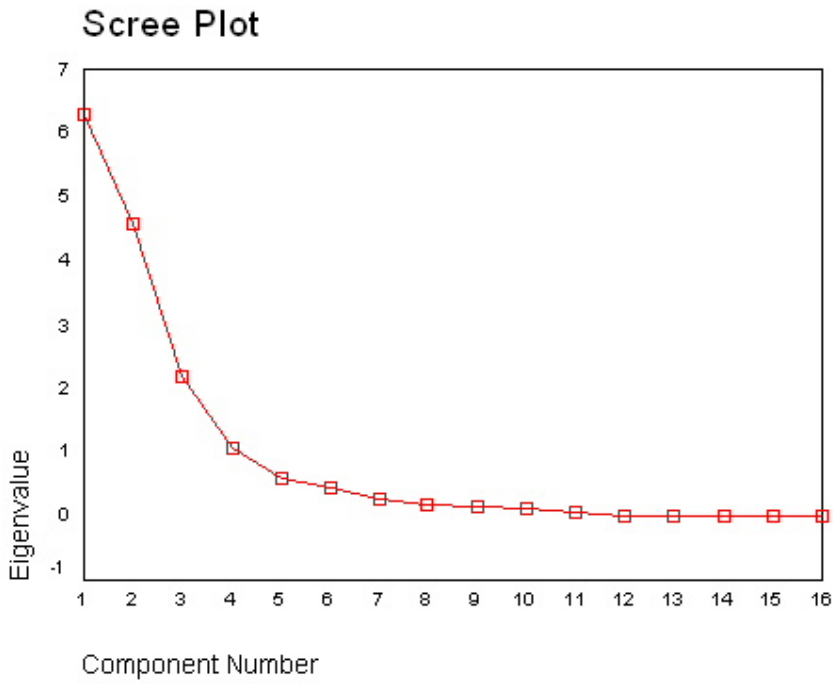


Figure 46. Scree plot for the IRL transformed dataset.

Table 27. Principal component analysis for the IRL transformed dataset.

Component Matrix[a]

	Component			
	1	2	3	4
SIO2	6.193E-02	.164	-.403	-.862
TIO2	-.979	-3.01E-02	6.620E-02	.119
AL203	.904	2.065E-02	8.512E-02	-2.36E-02
FE2O3	.893	-.405	4.749E-02	8.018E-02
MGO	.590	-.386	-.647	.117
CAO	.936	-.165	.204	5.129E-02
K2O	.645	3.910E-02	-.718	.104
NA2O	.904	7.459E-02	.296	1.683E-02
P2O5	.371	.810	-.349	.171
CU	.435	.838	-.198	.148
AS	.543	-.120	.480	-.282
PB	1.781E-02	-.963	.109	.113
CO	1.781E-02	-.963	.109	.113
NI	-.707	.440	-6.07E-02	.136
SC	.299	.697	.373	.266
LOI	.236	.552	.649	-.175

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

The fact that we have so many components as the result of the P.C.A., is an indication that we will not get good results this time. Equations 30 through 34 show the obtained factors.

Equation 30. FA8 for the IRL transformed dataset.

$$FA_8 = \frac{Al_2O_3 + Fe_2O_3 + MgO + CaO + K_2O + Na_2O + P_2O_5 + Cu + As + Sc + L.O.I.}{TiO_2 + Ni}$$

Equation 31. FA9 for the IRL transformed dataset.

$$FA_9 = \frac{SiO_2 + P_2O_5 + Cu + Ni + Sc + L.O.I.}{Fe_2O_3 + MgO + CaO + As + Pb + Co}$$

Equation 32. FA10 for the IRL transformed dataset.

$$FA_{10} = \frac{SiO_2 + MgO + K_2O + Na_2O + P_2O_5 + Cu}{CaO + As + Pb + Co + Sc + L.O.I.}$$

Equation 33. FA11 for the IRL transformed dataset.

$$FA_{11} = \frac{TiO_2 + MgO + K_2O + P_2O_5 + Cu + Pb + Co + Ni + Sc}{SiO_2 + As + L.O.I.}$$

Figures 47 through 50 shows the spatial distribution of these factors with respect to the location of our ore body.

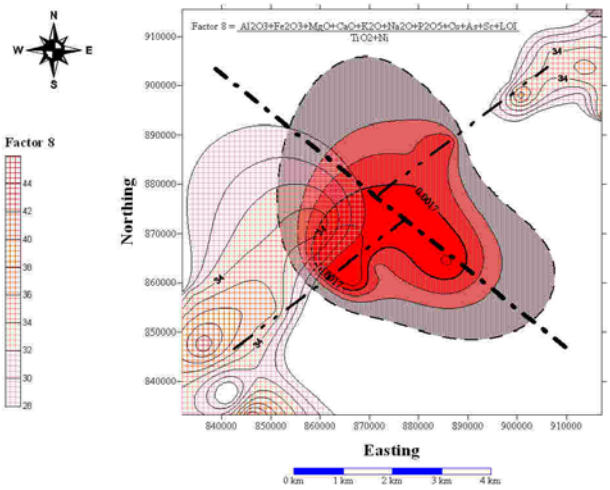


Figure 47. FA8 cannot target the ore body.

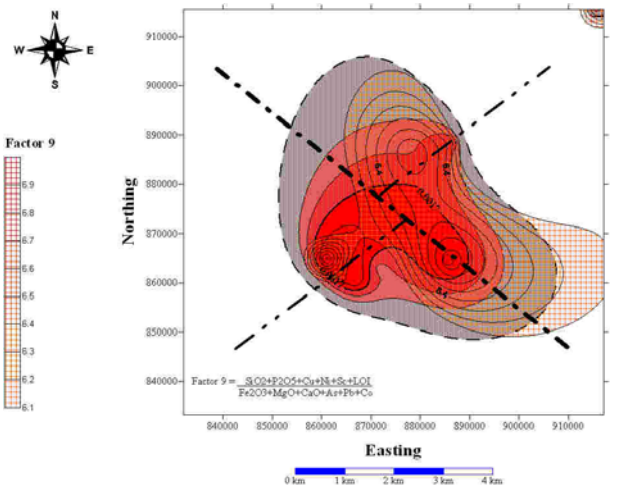


Figure 48. FA9 gives the best correspondence not only with the location of the ore body, but also with its internal composition.

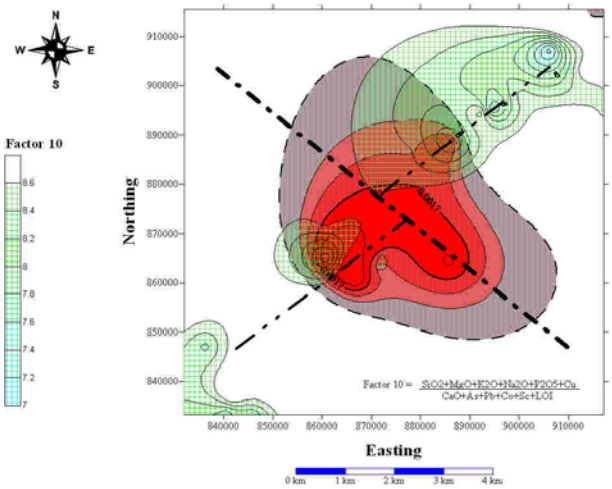


Figure 49. The fragmented character of FA10 makes it very difficult to use it as a targeting tool.

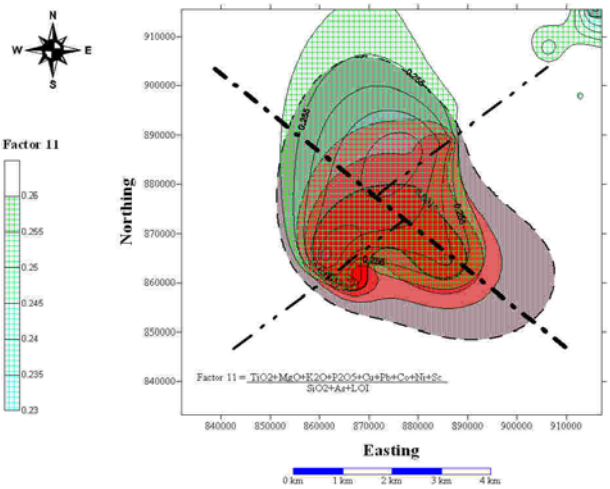


Figure 50. For as long as we had a known target to calibrate, the lower concentRations of FA11 are very good to locate our ore body.

Conclusions and Recommendations on the Use of FA for the Transformed Datasets

As I mentioned earlier, for FA to be most useful, one needs to have a known target to calibrate it. The factor analysis applied to the CLR transformed data gave us three factors, but only one (FA5) was useful for targeting the ore body.

The factor analysis of the ALR transformed data (Factor 7) was good in general, but the best factors were obtained from the ILR transformed data, specially Factor 9 that not only gave the exact location of the ore body, but also its internal structure. Another efficient factor was FA11, but it definitively required calibration based on a known target.

So answering the question from page 41, yes, the factor analysis of the IRL transformed data will be more effective than the factor analysis of the raw data as a tool for locating the ore deposit.

Dealing With Zero Values- a Summary

There is almost not a case in exploration geology, where the studied data does not includes below detection limits and/or zero values, and since most of the geological data responds to lognormal distributions, these "zero data" represent a mathematical challenge for the processing.

The method that I am proposing takes into consideration the well-known relationships between some elements. For example, in copper porphide deposits there is always a significant direct correlation between the copper values and the molybdenum ones. However, while copper will always be above the limit of detection, many of the molybdenum values will be "rounded zeros". In such a case, I will take the lower quartile of the real molybdenum values and establish a regression equation with copper, and then I will estimate the "rounded" zero values of molybdenum by their corresponding copper values.

One can apply this method to any type of data, provided we establish first their correlation dependency.

One of the main advantages of this method is that we do not obtain a fixed value for the "rounded zeros", but one that depends on the value of the other variable.

Are there any zeros in the house?

We need to start by recognizing that there are zero values in geology. For example the amount of quartz in a syenite is zero, since quartz cannot co-exists with nepheline (Trusova and Chernov, 1982). In binomial distributions, like for example the drilling of an ore body, you either will intersect the ore body (1) or not (0). Another common zero is a North azimuth, however we can always change that zero for the value of 360°. These are the "Essential Zeros" (Aitchison, 2003) or "Real zeros". They are not a problem for as long as their population does not respond to a lognormal distribution, since you cannot take the logarithm of a zero.

Then in geology, especially in geochemistry, we also have "Rounded Zeros". In some cases, laboratories report below detection limit (b.d.l.) as zeros or non-existent, while in most cases they just put the b.d.l. as the value for that parameter. These b.d.l. values are a similar problem to the "Rounded Zeros". Let us illustrate with the example proposed in Table 28.

Table 28. CLR transformed data for the used example. The original b.d.l. value was 0.5 ppm of Mo.

Au	Cu	Mo
0.23342	0.72138	0.0452
0.32663	0.61146	0.06191
0.12652	0.16198	0.71149
0.20133	0.28139	0.51728
0.20796	0.3302	0.46184
0.41506	0.51552	0.06942
0.20034	0.21824	0.58142

Au	Cu	Mo
0.22814	0.19547	0.57639
0.47232	0.48404	0.04363
0.21663	0.27648	0.50689
0.207	0.24005	0.55295
0.30235	0.24198	0.45567
0.20662	0.12838	0.665
0.25618	0.3309	0.41292

Au	Cu	Mo
0.13951	0.18003	0.68046
0.14029	0.12599	0.73372
0.12876	0.17744	0.6938
0.13442	0.28513	0.58045
0.22589	0.15577	0.61834
0.18861	0.13306	0.67834
0.26028	0.28088	0.45884
0.19831	0.31928	0.48241
0.37797	0.58109	0.04094
0.47982	0.46952	0.05065
0.17888	0.314	0.50712
0.26791	0.17539	0.5567
0.64782	0.3135	0.03868

Au	Cu	Mo
0.26629	0.29193	0.44178
0.50983	0.45371	0.03645
0.26377	0.25831	0.47792
0.50258	0.46207	0.03536
0.61358	0.3443	0.04212
0.34444	0.32052	0.33504
0.38402	0.1694	0.44658
0.24623	0.25965	0.49412
0.26424	0.17672	0.55904
0.42291	0.14872	0.42837
0.2371	0.18903	0.57387
0.27013	0.17273	0.55714
0.51564	0.45322	0.03113

The ternary diagram on Fig. 51 shows these results.

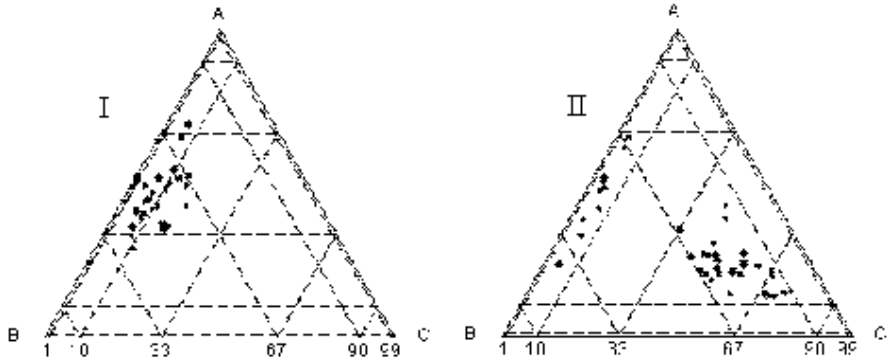


Figure 51. Ternary diagram of the studied data. To the left the CLR transformed data, to the right the same data after being centered using CoDaPack software,

It is clear that even centering does not solve the "problem" of the b.l.d. data, which remain grouped along the AB axis.

Zero, zero… What shall I do with you?

Geologists, even those that are not knowledgeable of compositional data analysis, have being dealing with this problem for quite some time (Kashdan *et al.*, 1979). One of the most frequently use technique is amalgamation (Aitchison, 1986). Amalgamation, e.g. adding Na_2O and K_2O, as total alkalis is a solution, but sometimes we need to differentiate between a sodic and a potassic alteration, and therefore amalgamation is not an option.

Pre-classification into groups is another solution, but it requires a good knowledge of the distribution of the data and the geochemical characteristics of the groups that is not always available.

Considering the zero values equal to the limit of detection of the used equipment, or substituting it by some other constant (e.g. half the limit of detection) will generate spurious distributions, especially in ternary diagrams as we show in Fig. 51.

Same situation will occur if we replace the zero values by a very small amount (Bacon-Shone, 2003) using non-parametric or parametric techniques (imputation). Even if we add the same small value to all of the analyzed parameters, we will get the same spurious distribution.

How do I deal with spurious distributions?

The method that I am proposing takes into consideration the existence of well-known relationships between some elements. For example, in copper porphide deposits, there is always a clear dependency between the copper values and the molybdenum ones (Fig. 52), but while copper will always be above the limit of detection, many of the molybdenum values will include b.d.l. values ("Rounded Zeros").

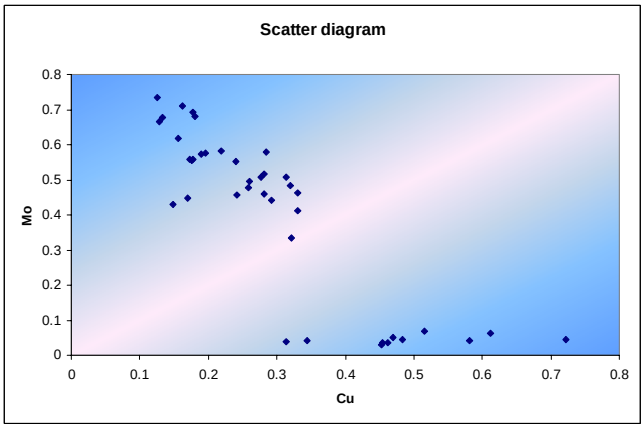


Figure 52. As in all Cu porphide deposits, there is a strong correlation between Cu and Mo values. We can use such correlation to estimate the values b.d.l. for Mo.

In this case, I will take the lower quartile of the real molybdenum values (Table 29) and establish a regression equation with copper, and then we will estimate the "Rounded Zero" values of molybdenum by those estimated from their corresponding copper values (Table 30).

Table 29. Values of the lower quartile of real data for Mo from this study.

Au	Cu	Mo
0.41506	0.51552	0.06942
0.34444	0.32052	0.33504
0.25618	0.3309	0.41292
0.42291	0.14872	0.42837
0.26629	0.29193	0.44178
0.38402	0.1694	0.44658
0.30235	0.24198	0.45567
0.26028	0.28088	0.45884
0.20796	0.3302	0.46184
0.26377	0.25831	0.47792

Table 30. Results of the regression analysis for the lower quartile of real molybdenum data from the studied case.

Regression Statistics	
Multiple R	0.792759158
R Square	0.628467082
Standard Error	0.079131742
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.084737701	0.084737701	13.53241	0.006231332
Residual	8	0.050094661	0.006261833		
Total	9	0.134832361			

	Coefficients	*Std. Error*	*t Stat*	*P-value*
Intercept	0.674594109	0.079027652	8.536178017	2.73E-05
X Variable 1	-0.954714886	0.259529113	-3.678642738	0.006231

Therefore, according to Table 30, we could use regression equation (34) in order to estimate the b.d.l. values for Mo.

Equation 34. Regression equation for Mo.

$$Mo = 0.675 - 0.955 * Cu$$

Fig. 53 shows that the obtained results are close to the predicted line.

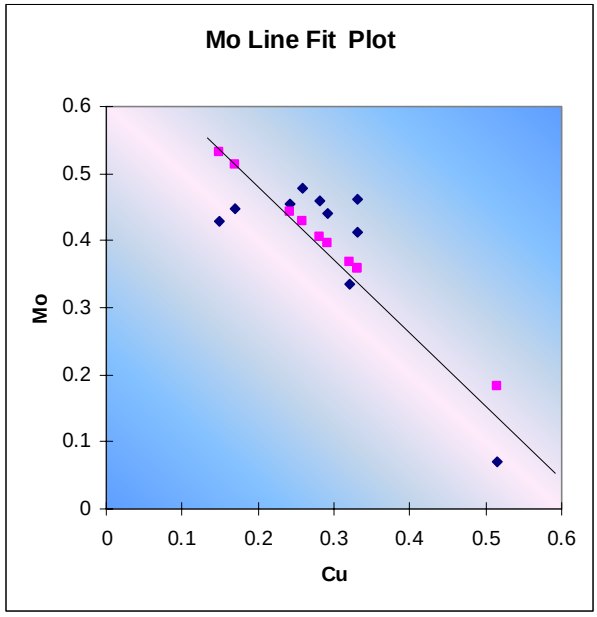


Figure 53. Molybdenum line fit plot showing a good correlation with the predicted values (in pink).

So, did we get ride of the spurious effect?

As Fig. 54 clearly shows, only one value of Mo was really close to zero, while the rest has now a value that is a geological reflection of the geochemical characteristics of the data.

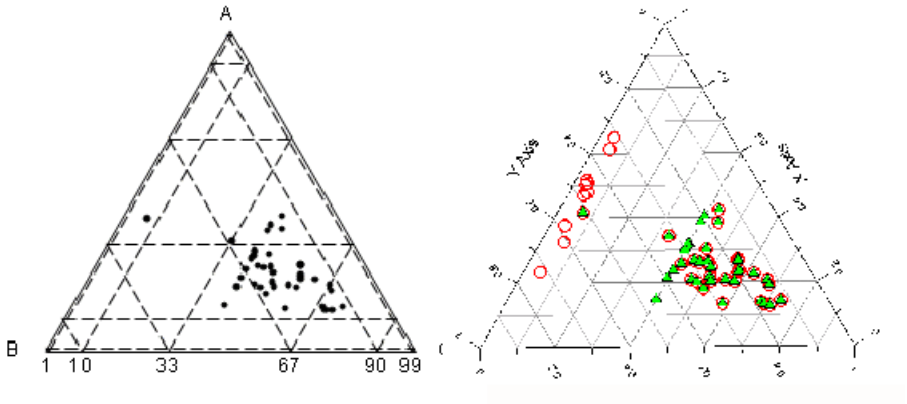


Figure 54. The ternary diagram of the left clearly shows that only one sample has low value of Mo. The ternary diagram to the right compares the original data in red with the new estimated values of Mo in green.

Conclusions and Recommendations

The treatment of "closed" dataset by normal statistical methods does create spurious correlations that lower the effectiveness of the obtained results. While there are ways to minimize this problem, like processing major oxides independently from trace elements, or using only the strongest correlations into the composition of the RCCs, I believe that the transformation of the initial dataset presents a better solution for the processing and interpretation of geological data.

For the estimation of the most efficient RCC, I propose the use of the ARL transformation, although the CLR is also effective.

When a target for the testing of the effectiveness of our coefficient is available, then we should use the factor analysis preferentially. I recommend the use of the IRL transformation to define the most effective combination of factors.

Finally, I introduced here a method for dealing with zero values. This method's main advantage is that we do not obtain a fixed value for the "Rounded Zeros", but one that depends on the real value of the other variable. The proposed method depends on the geological characteristics of the data, and therefore is less biased or random than other methods. It also presents a viable alternative to amalgamation and an effective way to deal with "Rssential Zeros" in a population.

The sequence of the method is as follows:

1. We transform the data using CoDaPack or other similar software.

2. We select the lower quartile of real data for the element with the b.d.l. values.

3. Within this dataset, we test the relationship between the elements with the b.d.l. values with one (or more) element without b.d.l. values. In most cases, these elements will correspond with well-established geological relationship like between Pb and Zn on polymetallic deposits, or between Au and Pb in hydrothermal deposits, or between Cu and Mo in porphyritic deposits, as in the case I presented here.

4. We establish the regression equation.

5. We then substitute the b.d.l. values by those estimated with the obtained equation of regression.

We can apply this method to any type of data, provided we establish first their correlation dependency. I would also like to see this method included as an option for dealing with zeros in the next version of CoDaPack.

References

Used in this text

Aitchison, J, 1986. The statistical analysis of compositional data. London: Chapman & Hall.

Aitchison, J. and J.W. Kay, 2003. Possible solutions of some essential zero problems in compositional data analysis. Proceedings of Compositional Data Analysis Workshop, 2003.

Bacon-Hone, J., 2005. Modeling structural zeros in compositional data. Proceedings of Compositional Data Analysis Workshop, 2003.

Kashan, A.B., Guskov, O.I. and A.A. Shimonsky, 1979. Mathematical modeling in prospection (original in Russian). Moscow: Nedra.

Trusova, I. F and V. I. Chernov. Petrography of magmatic and metamorphic rocks (original in Russian). Moscow: Nedra.

Suggested web links

3rd Compositional Data Analysis Workshop
http://ima.udg.edu/Activitats/CoDaWork08/

Bias arising from missing data in predictive models
http://www.scirus.com/srsapp/sciruslink?src=sd&url=http%3A%2F%2Fwww.sciencedirect.com%2Fscience%3F_ob%3DGatewayURL%26_origin%3DScienceSearch%26_method%3DcitationSearch%26_piikey%3DS0895435606002642%26_version%3D1%26_returnURL%3Dhttp%253A%252F%252Fwww.scirus.com%252Fsrsapp%252F%26md5%3D4868c0b11a79d4805f086c94316ef139

Compositional Data Analysis and Zeros in Micro Data.
http://www.scirus.com/srsapp/sciruslink?src=rpc&url=http%3A%2F%2Fideas.repec.org%2Fa%2Ftaf%2Fapplec%2Fv32y2000i8p953-59.html

Compositional Time Series: Past and Present
http://www.scirus.com/srsapp/sciruslink?src=rpc&url=http%3A%2F%2Fideas.repec.org%2Fp%2Fwpa%2Fwuwpem%2F0510002.html

Homepage of "compositions", an R package for compositional data analysis.
http://www.stat.boogaart.de/compositions/

Modeling Zeroes in Microdata.
http://www.scirus.com/srsapp/sciruslink?src=rpc&url=http%3A%2F%2Fideas.repec.org%2Fa%2Ftaf%2Fapplec%2Fv33y2001i3p383-92.html

New Features of CoDaPack. An User-friendly Compositional Data Package
http://ima.udg.edu/Activitats/CoDaWork05/CD/Session5/ThioHenestrosa-Tolosana-Gomez.pdf

Tectonic discrimination of basalts with classification trees.
http://www.scirus.com/srsapp/sciruslink?src=sd&url=http%3A%2F%2Fwww.sciencedirect.com%2Fscience%3F_ob%3DGatewayURL%26_origin%3DScienceSearch%26_method%3DcitationSearch%26_piikey%3DS0016703705009701%26_version%3D1%26_returnURL%3Dhttp%253A%252F%252Fwww.scirus.com%252Fsrsapp%252F%26md5%3De31d0170976cd137f78e7ce435aa9f32

Santiago Thió Fernández de Henestrosa.
http://ima.udg.edu/~thio/

Simcluster: clustering enumeRation gene expression data on the simplex space
http://www.scirus.com/srsapp/sciruslink?src=arx&url=http%3A%2F%2Farxiv.org%2Fabs%2Fq-bio%2F0703007

Some numerical considerations in the geochemical analysis of distal microtephra
http://www.scirus.com/srsapp/sciruslink?src=sd&url=http%3A%2F%2Fwww.sciencedirect.com%2Fscience%3F_ob%3DGatewayURL%26_origin%3DScienceSearch%26_method%3DcitationSearch%26_piikey%3DS0883292706001727%26_version%3D1%26_returnURL%3Dhttp%253A%252F%252Fwww.scirus.com%252Fsrsapp%252F%26md5%3Df9ab2e6094792c99aa80f30699cf5cb7

Statistical empirical index of chemical weathering in igneous rocks: A new tool for evaluating the degree of weathering
http://www.scirus.com/srsapp/sciruslink?src=sd&url=http%3A%2F%2Fwww.sciencedirect.com%2Fscience%3F_ob%3DGatewayURL%26_origin%3DScienceSearch%26_method%3DcitationSearch%26_piikey%3DS0009254107001052%26_version%3D1%26_returnURL%3Dhttp%253A%252F%252Fwww.scirus.com%252Fsrsapp%252F%26md5%3D8760181502f5d424d4ebb5fe21bf5ef6