Application and Statistical Validation of a Linear Mixing Model: an Approach to Vent Fluid Geochemistry

Villanueva-Estrada, R.E.[a], Prol-Ledesma, R.M.[a], Torres-Alvarado, I.S.[b], Canet, C.[a]

[a]Departamento de Recursos Naturales, Instituto de Geofísica, Universidad Nacional Autónoma de México, Ciudad Universitaria, Delegación Coyoacán, 04510 México, D.F., México. [b]Centro de Investigación en Energía, Universidad Nacional Autónoma de México, A.P. 34, Temixco, 62580 Morelos, México

ruth@geofisica.unam.mx

Keywords: hydrothermal fluid, statistics, end-member composition.

ABSTRACT

A linear mixing model has generally been used to calculate the thermal end-member composition of a mixed hydrothermal vent fluid when the concentration of a chemical tracer (e.g. magnesium) can be considered zero for the thermal end-member. However, it is necessary to evaluate the feasibility of the linear regression model that is used to calculate the intercept value. After a statistical evaluation of the linear model application, it was found that the calculation of the chemical composition of the thermal end-member using a linear mixing model is feasible. However, a statistical evaluation of this model is necessary before its application. In the examples presented in this paper, it was considered that all of the Mg in the vent fluid originates from the seawater and is added to the vent fluid through a mixing process occurring beneath the seafloor discharges (Bischoff and Seyfried, 1978). Further, at temperatures around 85°C, the magnesium reacts with silicate species to form a Mg-hydroxi-silicate precipitate (Bischoff and Seyfried, 1978). The proposed validation of this regression method was performed in the following steps: (a) establishing the relationship between the chemical variables (Mg vs. concentrations of other dissolved chemical elements) using correlation and determination coefficients, (b) using residual plots as a diagnostic tool of the linearity and presence of possible outliers, (c) statistical validation of the linear regression equation, and (d) proposal of a confidence interval for the Y-intercept calculated using the regression model. For the cases presented here (both coastal hydrothermal systems off the coast of Mexico), it was concluded that Ca, Si, Mn and Ba are supplied by the hydrothermal system, since Na, Cl, SO_4 and K are all present due to a seawater mixing process occurring immediately beneath the vent discharges.

1. INTRODUCTION

The study of the geochemistry of geothermal fluids is useful to identify their origin and to quantify the processes that govern their composition and the associated chemical and mineralogical transformations of the rocks in contact with these fluids (Arnórsson et al., 2007). A serious problem with the fluid sampling of submarine vents, even in the shallower ones (depth < 200 m), is the mixing of hydrothermal fluid with seawater that occurs during the sampling operation. Although different sampling methods have been proposed to minimize the contamination with seawater in the vent fluid samples (Arnórsson et al., 2006; Johnson et al., 2006; Villanueva et al., 2006), seawater contamination is unavoidable.

To diminish seawater contamination and estimate the chemical composition of a thermal end-member (TEM – a fluid representative of a deep geothermal reservoir) in a hydrothermal vent system, a linear mixing model can be applied based on magnesium content (Vidal et al., 1981; Von Damm, 1990; Scott, 1997; Fitzsimons et al., 1997; Glasby and Notsu, 2003; Prol-Ledesma, 2003; Prol-Ledesma et al., 2004). This model assumes a linear mixing between two different end-members: the TEM and seawater. The magnesium content in seawater is around 52.7 mmol/kg (1281 ppm), whereas the TEM is assumed to contain almost none (Scott, 1997). This assumption about the magnesium TEM composition is consistent with the experiments made by Bischoff and Seyfried (1978); which involved progressively heated seawater in contact with basalt at temperatures up to 350°C. At the end of this experiment, the resulting fluid was highly depleted in magnesium because it precipitated when it reacted with silicate species. In view of that, magnesium concentrations have been used to estimate the bulk chemical composition of the TEM (Von Damm, 1990; Fitzsimons et al., 1997; Gieskes, et al., 2002; Prol-Ledesma, 2003; Prol-Ledesma et al., 2004). The concentration of the TEM is calculated from the extrapolation of [Mg] = 0.0 in a [Mg] vs. [component] plot (where square brackets indicate concentration). The obtained values represent the composition of a fluid before mixing with seawater and being discharged to the seafloor. These values indicate concentrations, and this mathematical method is the most effective known technique of correcting for seawater entrainment during sampling (Von Damm, 1990). Thus, the chemical composition of the TEM is calculated using a linear mixing model, but it is necessary to evaluate the validity of the linear regression equation with statistical methods.

In this paper, we propose statistical validation for the application of a linear regression model for determining the chemical composition of the TEM in submarine vent systems. We illustrate its application using the chemical composition of two Mexican shallow hydrothermal systems: (a) Bahía Concepción in Baja California Sur (Prol-Ledesma et al., 2004), and (b) Punta Mita in Nayarit (Prol-Ledesma et al., 2002).

2. LINEAR REGRESSION MODEL

A linear regression model is useful in discerning the probable type of relationship between the dependent variables (Y) and a independent variable (X). Thus, it is applied to calculate the value of these dependent variables (Burke, 2001; Daniel, 2005; Pagano y Gauvreau, 2001). In relation to the linear dependencies, it is necessary to calculate the following aspects (Daniel, 2005):

a) Linear dependence parameters.

b) Linearity hypothesis of the linear dependence.

c) Error of the parameters of the linear dependence.

d) Confidence interval for the linear dependence.

All these aspects can be evaluated using the linear least-squares regression technique. The focus of this paper is on the most common type of linear regression model of the relationship between a dependent variable (Y) and an independent variable (X).

The application and validation of the linear regression model to calculate specific component concentrations of fluids produced by hydrothermal activity in the chosen localities is presented here.

Correlation and determination coefficients, verification of linearity, and the Student's *t* statistic test were used to evaluate the linear regression model.

3. EXAMPLE: CALCULATION OF THE THERMAL END-MEMBER IN TWO SHALLOW SUBMARINE HYDROTHERMAL SYSTEMS

The statistical validation of a linear mixing model applied to two shallow submarine hydrothermal systems was developed, and the calculated TEM compositions were evaluated accordingly. Fluid geochemical data from the Punta Mita and Bahía Concepción hydrothermal systems in Nayarit and Baja California Sur were tested. These data were reported by Prol-Ledesma et al. (2002, 2003), respectively. In both cases, the hydrothermal fluids were collected with efforts to avoid seawater contamination (details about the sampling methods are provided by Villanueva et al., 2006), and the chemical composition of the TEM was estimated from the Mg content using a linear mixing model (Prol-Ledesma et al., 2002, 2004).

3.1 Bahía Concepción

The residuals are plotted in Figures 1 and 2 for the Bahía Concepción hydrothermal system. The outlier test value was applied to suspected outliers. The criterion to accept or reject a suspected outlier was in agreement with the outlier test for linear-squares regression (Burke, 2001). For Bahía Concepción, the test value for the suspected outlier (Mn) was 1.08, as shown in Figure 1. According to the outlier test, the critical value should be <1.48 (n=4 and 95% confidence level). Therefore, this value is not an outlier and can be considered part of the statistical population to be used in the linear regression model.

According to the Pearson´s correlation coefficient test, there is a significant correlation when $|r|$ is ≥ 0.95 (n-2 degrees of freedom and 95 % confidence level). For Cl and K, there is not a significant linear relationship (0.90 and -0.63 respectively). Otherwise, the correlation coefficients are significant and negative linear relationships can be established for Ca (-1.00), Si (-1.00), Mn (-1.00) and Ba (-0.99). Hence, we can affirm that these chemical species originate in the deep hydrothermal system. On the other hand, Na (1.00) and SO_4 (0.97) have significant correlation coefficients that indicate positive linear relationships with magnesium concentration and, thus, marine origins.

The determination coefficient means that about 97 % of the total variation of Ca, Si, Mn and Ba can be attributed to hydrothermal activity. In the case of Na and SO_4, more than 94 % of the variation can be attributed to seawater mixing.

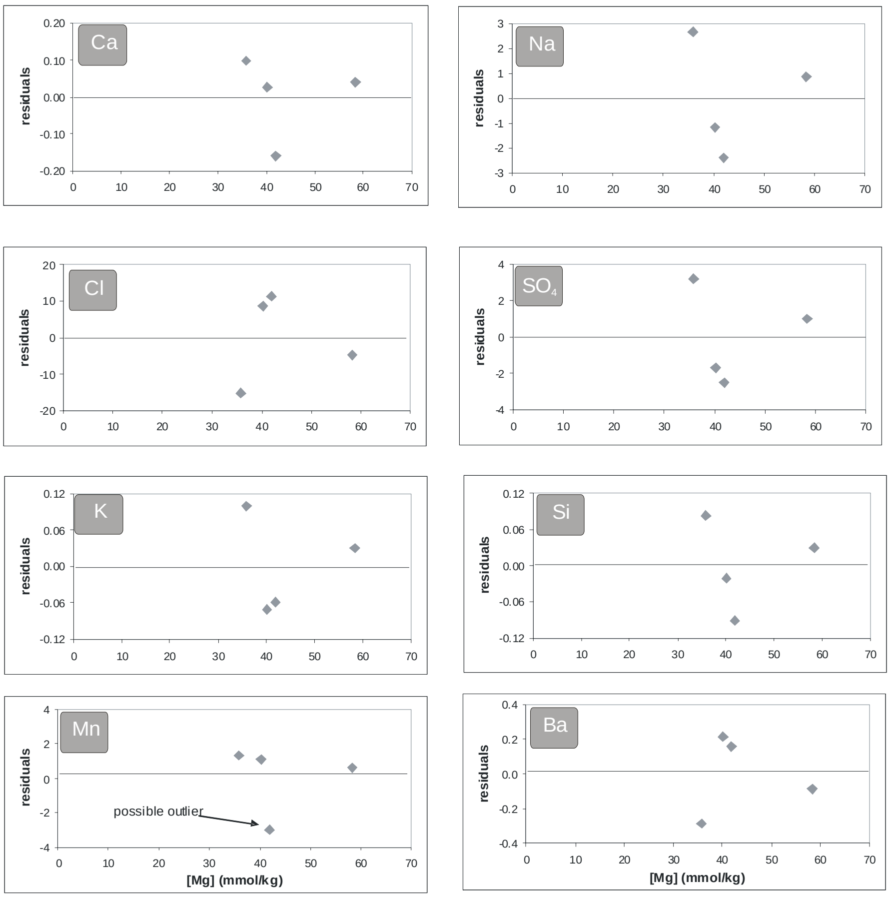

Figure 1: Plots of the residuals for vent fluid chemical data from Bahía Concepción (Prol-Ledesma et al., 2004). These plots represent the residuals (Yi-Ŷ) vs. [Mg] for: Ca, Na, Cl, SO_4, K, Si, Mg and Ba.

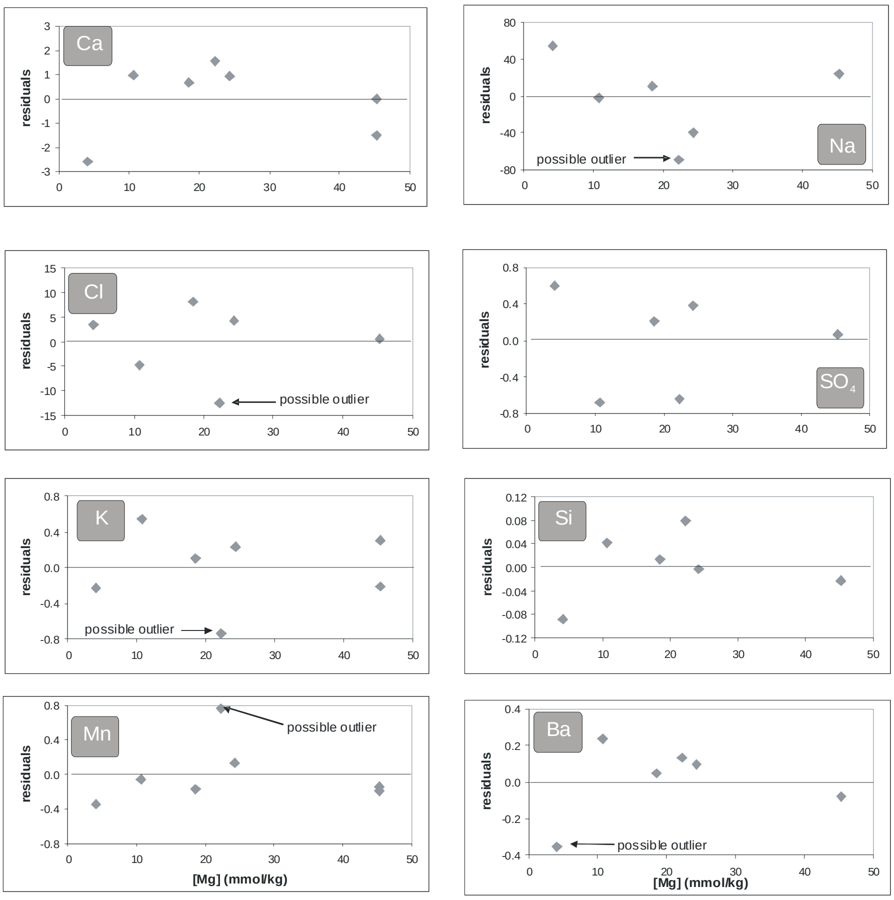

Figure 2: Plot of residuals for geochemical data of vent fluid from Punta Mita (Prol-Ledesma et al., 2002). These graphs represent the residuals (Yi-Ŷ) vs. [Mg] for: Ca, Na, Cl, SO_4, K, Si, Mn and Ba.

Although K and Cl accomplish the linearity condition, without curvature in the residuals plots (Fig. 1), these elements do not satisfy the Student's *t* statistical test, as the null hypothesis was accepted, or Pearson´s correlation coefficient test. For these reasons, the calculated value for Mg = 0 can not be calculated and neither can the confidence intervals for Cl and K.

3.2 Punta Mita

For the Punta Mita data set, a critical value of 2.08 (n=7) was obtained according to the outlier test for linear least-squares regression. The possible outliers for Na, Cl, K, Mn, and Ba were tested, and the results in Figure 2 show that the test values were below the critical value. Therefore, we can include them in the linear regression model.

From the Pearson´s correlation coefficient test, there is a significant correlation when $|r| \geq 0.754$ (Burke, 2001). Na, Cl, SO_4 and K have a significant, positive linear relationship (r> 0.95), and Ca, Si and Mn show a negative linear relationship (with r values of -0.99, -1.00 and -0.85 respectively).

The determination coefficient indicates that 97 % of the variability concentration of Ca, Na, Cl, SO_4, K, Si and Ba can be attributed to a deep hydrothermal source. However, only 72 % of manganese variation can be attributed to that source.

A linear regression model evaluation indicates that the relationship between Mg and the rest of the elements (Ca, Na, Cl, SO_4, K, Si, Mn and Ba) is linear and statistically acceptable for the calculation of TEM chemical composition. A confidence interval (95 % of confidence level) for all elements was calculated.

4. CONCLUSIONS

A linear regression model can be used to calculate the chemical composition of the TEM from the chemical analysis of vent water samples. However, statistical validation must be performed, according to the following steps:

a) Calculate the linear regression equation using the least-squares method after plotting the data.

b) Calculate the determination and correlation coefficients.

c) Verify the linearity with the residual plot by evaluating the possible outliers and applying the statistical test of the test value.

d) Evaluate the application of the linear regression equation (ANOVA or *t*-statistic).

e) Calculate the Y value and proposed and confidence interval for the calculated value.

A linear mixing model is adequate to calculate the chemical composition (Ca, Si, Mn, and Ba) of the TEM based on the magnesium content of the Bahía Concepción and Punta Mita hydrothermal systems.

5. ACKNOWLEDGMENT

Funding for this study was provided by IMPULSA-UNAM project and CONACyT project J-51127-I.

REFERENCES

Arnórsson, S., Bjarnason, A.O., Giroud, N., Gunnarsson, I., Stefánsson, A. Sampling and analysis of geothermal fluids. *Geofluids*, **6**, (2006), 203-216.

Arnórsson, S., Stefánsson, A., and Bjarnason, A.O.. Fluid-Fluid Interactions in Geothermal Systems. In: A. Liebscher, C.A. Heinrich (Editors), Fuid-Fluid Interactions. Reviews in Mineral & Geochemistry. **65** , (2007), 259-312.

Bischoff, J. L., Seyfried, W.E. Hydrothermal chemistry of seawater from 25° to 350°C. American Journal of *Science*, 278, (1978), 838-860.

Burke, S. Regression and Calibration. Statistics and data analysis. LCGC Europe online supplement, (2001).

Daniel, W.W. Bioestadística. Bases para el análisis de las ciencias de la salud. 4ª. Edición. Limusa Wiley. (2005).

Fitzsimons, MF., Dando, P.R., Hughes, J.A., Thiermann, F., Akoumianaki, I., Pratt, S.M. Submarine hydrothermal brine seeps off Milos, Greece: Observations and geochemistry. *Marine Chemistry*, **57**, (1997), 325-340.

Gieskes, J.M., Simoneit, B.R.T., Shanks III, W.C., Goodfellow, W.d., James, R.H., Baker, P.A., Ishibashi, J. Geochemistry of fluid phases and sediments: relevance to hydrothermal circulation in Middle Valley, ODP Legs 139 and 169. *Applied Geochemistry*, **17**, (2002), 1381-1399.

Glasby, G.P., Notsu, K. Submarine hydrothermal mineralization in the Okinawa Trough, SW of Japan: an overview. *Ore Geology Reviews*, **23**, (2003), 299-339.

Johnson, H.P., Baross, J.A., Bjorklund, T.A. On sampling the upper crustal reservoir of the NE Pacific Ocean. *Geofluids*, **6**, (2006), 251-271.

Pagano, M., Gauvreau, K. Fundamentos de Bioestadística. 2ª edición. International Thomson editors. México, (2001).

Prol-Ledesma R.M. Similarities in the chemistry of shallow submarine hydrothermal vents. *Geothermics*, **32**, (2003), 639-644.

Prol-Ledesma, Canet, C., Melgarejo, J.C., Tolson, G., Rubio-Ramos, M.A., Cruz-Ocampo, J.C., Ortega-Osorio, A., Torres-Vera, M.A., Reyes, A. Cinnabar deposition in submarine coastal hydrothermal vents, Pacific Margin of Central Mexico. *Economic Geology*, **97**, (2002), 1331-1340.

Prol-Ledesma, R.M., Canet, C., Torres-Vera, M.A., Forrest, M.J. and Armienta, M.A. Vent fluid chemistry in Bahía Concepción coastal submarine hydrothermal system, Baja California Sur, México. *Journal of Volcanology and Geothermal Research*, **137**, (2004), 311-328.

Scott, S.D. Submarine hydrothermal systems and deposits. In: Barnes, H.L. (Editor), Geochemistry of hydrothermal ore deposits. John Wiley & Sons, Inc., (1997).

Vidal, V.M.V., Vidal, F.V., Isaacs, J. D. Coastal submarine hydrothermal activity off Northern Baja California 2. Evolutionary history and isotope geochemistry. *J. of Geophysical Research*, **86**, (1981), 9451-9468.

Villanueva, R.E., Prol-Ledesma, R.M., Torres-Vera, M.A., Canet, C., Armienta, M.A., de Ronde, C.E.J. Comparative study of sampling methods and in situ laboratory analysis for sallow-water submarine hydrothermal systems. *J. of Geochemical Exploration*, **89**, (2006), 414-419.

Von Damm, K.L. Seafloor hydrothermal activity: black smoker chemistry and chimneys. *Annu. Rev. Earth Planet. Sci.*, **18**, (1990), 173-204.