Proceedings World Geothermal Congress 2020
Reykjavik, Iceland, April 26 – May 2, 2020

GQAnalyzer: An R Package for Geochemical Analysis of Geothermal Fluids and Gases

Oscar. García-Cabrejo, Juan P .Ávila, Laura E. Cárdenas and Laura A. Barrantes

Calle 4 Sur N° 15-134, Escuela Ingeniería Geológica, Seccional Sogamoso, Colombia.

Khaors@gmail.com

Keywords: Geochemistry, Geothermal Fluids-Gases, R Package, Geothermal Education.

ABSTRACT

This work introduces the R package GQAnalyzer that can be used to create the most important specialized plots employed in the analysis and interpretation of the chemical composition of hot springs, fumarole and well samples acquired in geothermal exploration and development programs. This package is composed of several functions that can be used from the command line and/or scripts, and a Graphical User Interface has been developed and included in the package to facilitate the geochemical analysis. In the case of geothermal fluids, GQAnalyzer has the capability to create user defined scatter and ternary plots, the Giggenbach Na-K-Mg geothermometer plot, and several plots used in the classification of the water samples including conventional and modified Piper, Durov and multirectangular diagrams. Fluid Mixing can be analyzed using the conventional Cl and SiO2 Enthalpy plots and the M3 algorithm based on a multivariate analysis of the composition of water. In addition, thirty-five geothermometers can be easily calculated from the chemical composition of water. For gases and steam samples, GQAnalyzer includes ternary diagrams (N2-He-Ar, N2-Ar-CO2, CH4-CO2-H2S, CO2-H2S-NH3), gas ratio geothermometers (CO2-Ar and H2-Ar, CO/CO2 and CH4/CO2), and the geothermometer grids (Fischer-Tropsh and H2-H2S, Fischer-Tropsh and CO2, Fischer-Tropsh and H2S). The implementation of the GQAnalyzer package in an open source programming language with great graphical capabilities such as R makes this package a good alternative to the widely used spreadsheets employed in the analysis of geothermal information. This package is offered under the GPL-3 license and can be used for teaching as well as research purposes without warranty for its use and/or the obtained results.

1. INTRODUCTION

An important part of any exploration program of geothermal resources involves the collection of samples of the chemical composition of geothermal fluids and gases. The analysis of this information via specific plots is a powerful tool to identify the geochemical processes controlling the composition of these fluids such as fluid mixing, reactions with reservoir rocks, and alteration patterns. Even today with advanced computers, these plots are the workhorse of the analysis of geothermal chemical compositions that are routinely used in exploration programs around the world.

There is a good review of the chemistry of geothermal fluids and gases by Nicholson (1993), where the link between chemical processes and mixing is presented via different types of plots. The same idea is followed by the encyclopedic work by Arnosson (2000), where a modern review of the geochemical techniques used in the study of the geothermal reservoirs. An important step along this line is the development of the spreadsheets to create most of plots used in the analysis of the chemical composition of geothermal fluids and gasses by Powell and Cumming (2010).

Although the spreadsheets created by Powell and Cumming (2010) provide an useful tool for the creation of the most important plots required in the analysis of the chemical compositions of geothermal fluids and gases, the platform in which these spreadsheets are developed (Excel (c)) is far from being a complete and versatile data analysis environment. Today most of the geothermal studies require the integration of information from multiple sources including but not limited to chemical composition of geothermal fluids, geology and geophysical surveys. This information must be analyzed in order to define conceptual models of the sites under study, formulate and tests different hypothesis about the origin and behavior of the geothermal systems. The previously mentioned tasks are facilitated with the possibility to do an interactive analysis of the available information and therefore a computational platform that support this type of analysis is required. The programming language R offers this type of interactive platform for statistical data analysis, which is released under a GNU GPL license and a large number of packages which makes it a viable option for the analysis of chemical composition of geothermal samples. Despite these advantages, to the best knowledge of the authors there is no specific R package aimed at the creation of the specialized plots commonly used in the analysis of the chemical compositions of geothermal fluids, and this is the starting point of this work.

This paper introduces the R-package GQAnalyzer that is designed to help in the creation of specialized plots used in the geothermal and hydrogeochemical studies. The programming language R was selected due to its open-source philosophy and the availability of state-of-the-art packages for visualization and statistical analysis. This enables an open-source distribution of the GQAnalyzer package while including the most recent advances in visualization and data analysis.

2. PACKAGE STRUCTURE

The R programming language was created by Ross Ihaka and Robert Gentleman in 1993 as a computational tool to help teaching basic statistical concepts. From this moment, R became the programming language of the statistical community and today is widely used in this programming language can be extended outside its original functionality via packages. According to Wickhan (2015), the R packages are the fundamental units of reproducible R code that include reusable R functions, documentation on how to use them, data and compiled code in a well defined and organized format that ensured its integration with other R packages. Currently there are over 12000 R packages covering from analysis of DNA information to educational surveys, but there is not a single package designed to help with the analysis of the chemical composition of geothermal fluids. For this reason, the package GQAnalyzer has been created.

The structure of the GQAnalyzer is based on the definition of a S3 class called geochemical dataset used to store the concentrations of the major and minor ions. All the operations and plots in the GQAnalyzer are based on the definition of this geochemical dataset variable. Two important calculations done when this variable is defined:

-The conversion from concentrations to meql

-Calculation of mixing ratio between fresh and a saline water of a given composition. An average seawater chemical composition is used as default, but the chemical composition of a different source (geothermal fluid) can also be specified. This calculation is helpful in the mixing analysis included in the package. Once the concentrations of the major and minor ions and trace elements are loaded in R, the geochemical dataset is created using:

my_data.gd ← geochemical_dataset (name = "MyData", data = concentrations.df).

where the concentrations of the major and minor ions are in the data frame concentrations.df. The main functions associated with the geochemical_dataset variable are shown in Table 1. The main strength of the GQAnalyzer package is the creation of several specialized plots for a single and multiple samples, which are based on the use of the ggplot2 library (Wickhan, 2016). The specialized plots currently available in the GQAnalyzer package are included in Table 2. These plots can be easily generated with the plot command as follows:

p1 ← plot (my_data.gd, type = "piper", measure = "conc")

p2 ← plot (my_data.gd, type = "multirectangular", measure = "conc")

p3 ← plot (my_data.gd. Type = "giggenbach")

where p1, p2 and p3 represents ggplot2 objects that can generate the specific plot. These plots can be shown using the print method. In detail, the creation of these specialized plots can be considered as a mapping from the concentrations space into an artificial space designed to highlight specific characteristics of our data. This conversion is done for each plot type using the transform function that creates a data.frame with the corresponding coordinates in the transformed space. This means that there is a transform function for each specialized plot included in the package. This data.frame is used in a conventional ggplot sequence of command to create the specific plot. For example, the creation of a Giggenbach plot can be done manually using the following sequence of commands:

- Calculation of the coordinates in the Giggenbach plot

my.data.giggenbach.df ← giggenbach_ternary_transform (my_data.gd)

- Plot the coordinates

p1 ← ggplot_giggenbach () + ggplot (aes(x = xc, y = yc), data = my.data.giggenbach.df)

where the command ggplot_giggenbach () defines the basic template of the plot.

The final plot is saved in the variable p1 and it can be displayed on the screen calling the method print from the command window. The same procedure is followed for other types of diagrams such as Piper and multirectangular that involves all the samples in a given geochemical_dataset. Additionally, it is possible to overload the subset operator [] in order to have access to the chemical composition of a chemical sample. The following code creates a Stiff diagram of the second sample included in the geochemical_dataset variable my_data.gd:

p4 ← plot(my_data.gd[2],type="stiff", measure = "conc")

The previously explained procedure to create different plots is based on the use of the R command line, which is the core of an interactive data analysis. However, it might be daunting for some users who are not used to write code or write scripts to automate the analysis processes. With this audience in mind, the package GQAnalyzer includes a Graphical User Interface designed to help in the creation of these specialized plot and the application of 24 geothermometers as shown in figure 1. The design is based on tabs that the user must follow from left to right. A short description of each tab is included in Table 3. A List of the aqueous geothermometers currently included in the package is shown in Table 4.

Table 1. Main functions included in the package GQAnalyzer.

FUNCTION	DESCRIPTION
geochemical_dataset	Constructor of the geochemical_dataset S3 class
print	Show the contents of the geochemical_dataset variable on the command window
summary	Show a statistical summary of the concentrations of the chemical species specified as input
plot	Creates a ggplot plot with the requested plot and other characteristics specified

Table 2. Specialized plots included in the package GQAnalyzer.

CATEGORY	PLOT
Single Sample	Stiff (Stiff, 1951)
	Radial
Multiple Samples	Ternary
	Schoeller (Schoeller, 1962)
	Piper (Piper, 1944)
	Modified Piper (Appelo and Postma, 2005)
	Durov
	Multirectangular (Ahmad et al, 2003)
	Na-K-Mg Geothermometer (Giggenbach plot), Giggenbach(1988)
	CAR-HAR Geothermometer (Giggenbach and Glover, 1992)

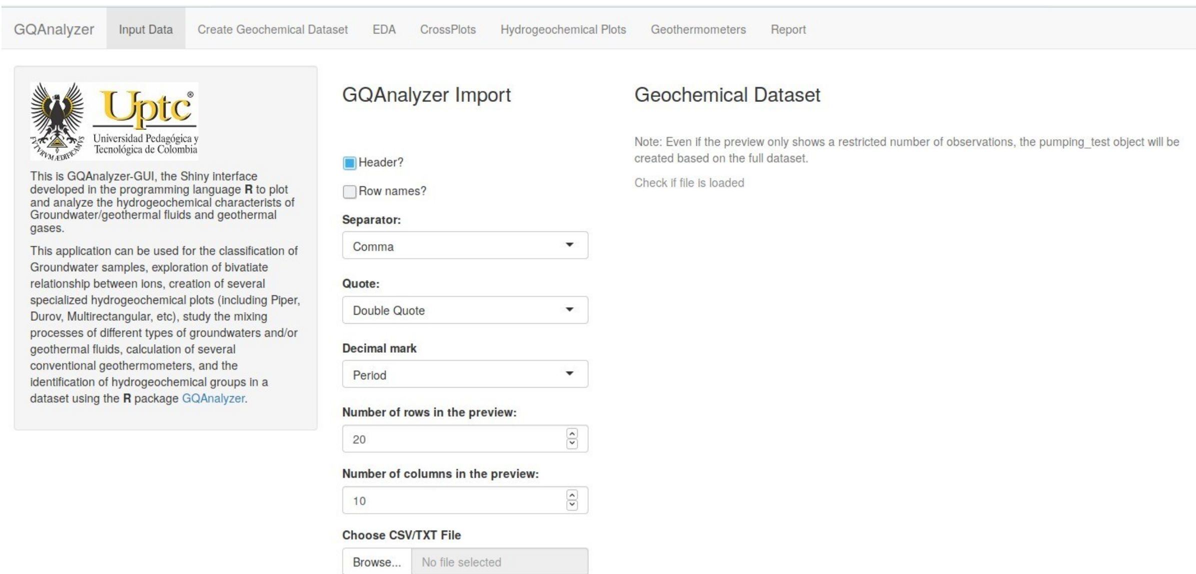

Figure 1. GQAnalyzer-GUI Screenshot.

Table3. Description of the Tabs of the GQAnalyzer-GUI.

TAB	DESCRIPTION
Input Data	Import data into R
Create Geochemical Dat	Create the geochemical_dataset variable
EDA	Creation of histograms, boxplots, empirical cumulative distributions plots, Quantile-Quantile plots
Cross-plots	Creation of scatterplots between the concentrations of different ions
Hydrogeochemical plots	Creation of the specialized hydrogeochemical plots for a single sample and all samples
Geothermometers	Calculation of 17 aqueous geothermometers
Report	Creation of simple reports with the corresponding plots in html/doc format

Table4. Type of geothermometers included in the package GQAnalyzer

TYPE	NAME
Na-K	Fournier-Truesdell (1973)
	Truesdell (1976)
	Fournier (1979)
	Tonani (1980)
	Arnosson (1983)
	Arnosson (1983)
	Nieva-Nieva (1987)
	Giggenbach (1988)
	Verma-Santoyo (1997)
	Can (2002)
	DiazGonzales-Santoyo-Reyes (2007)
	DiazGonzales-Santoyo-Reyes (2007)
K-Mg	Giggenbach (1988)
	Fournier (1991)
	Fournier (1991)
Mg-Li	Kharaka.Mariner (1989)
	Kharaka.Mariner (1989)
Si	Fournier (1977)
	Fournier (1977)
	Fournier-Potter(1982)
	Verma-Santoyo (1997)
	Verma-Santoyo (1997)
	Arnosson (2000)
	Verma (2000)

3. EXAMPLES

In this section some examples of plots created using the package GQAnalyzer are shown. Figure 2 shows an example of a ternary plot used in the identification of sources of geothermal gases from samples collected in the Eastern Cordillera of Colombia. Most of the samples belong to the group of geothermal gases. Figure 3 shows a Piper diagram created using the chemical composition of groundwater samples and geothermal hot springs from the Quaternary aquifer in the Central part of Boyaca, Colombia. There are no apparent trends that indicate mixing between groundwater and geothermal fluids in this area from this plot.

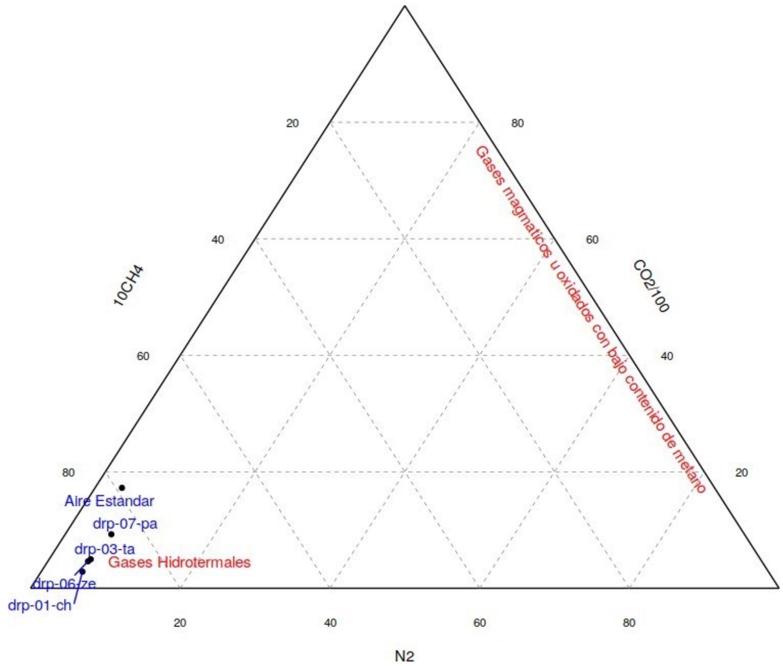

Figure 2. Example of a Ternary Diagram used to identify the origin of geothermal gases.

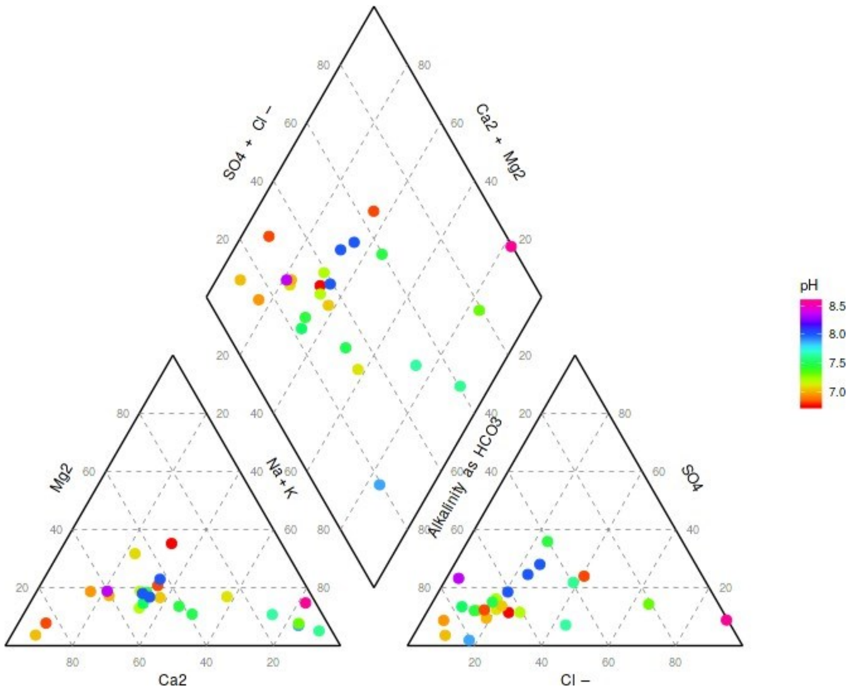

Figure 3. Example of a Piper Diagram of a set of groundwater samples in a Quaternary Aquifer in Colombia. (Piper, 1944)

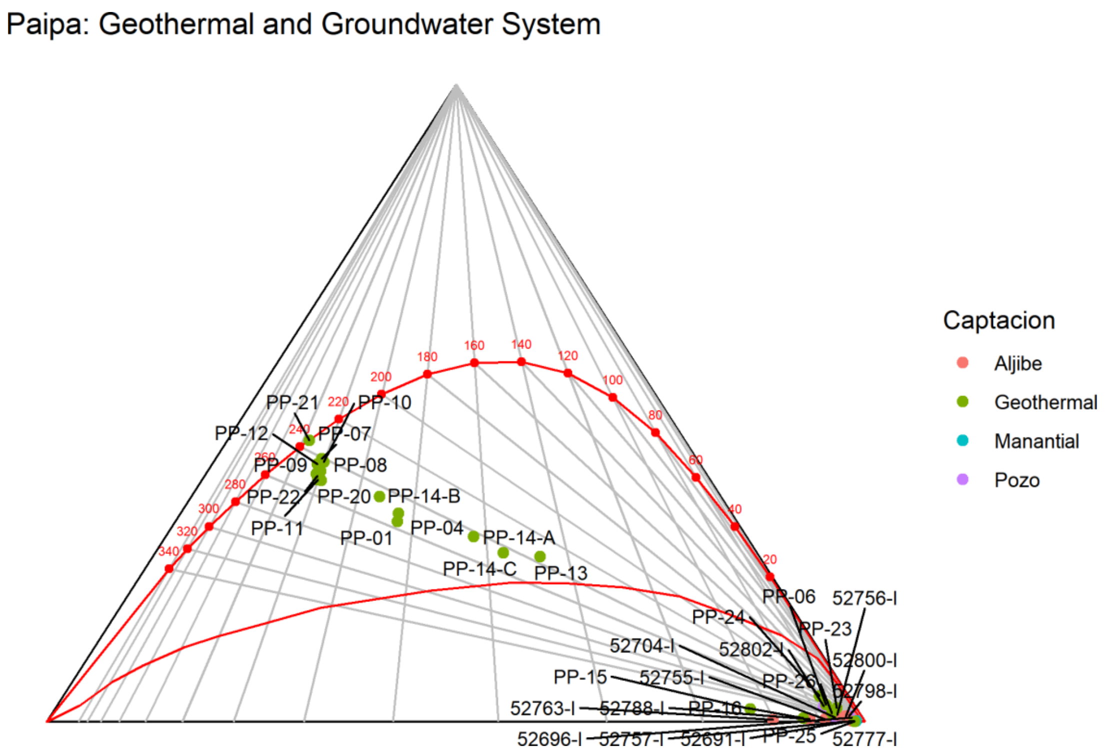

Figure 4. Example of a Na-Ca-Mg geothermometer plot of the Paipa Geothermal System, Colombia. (Giggenbach, 1988)

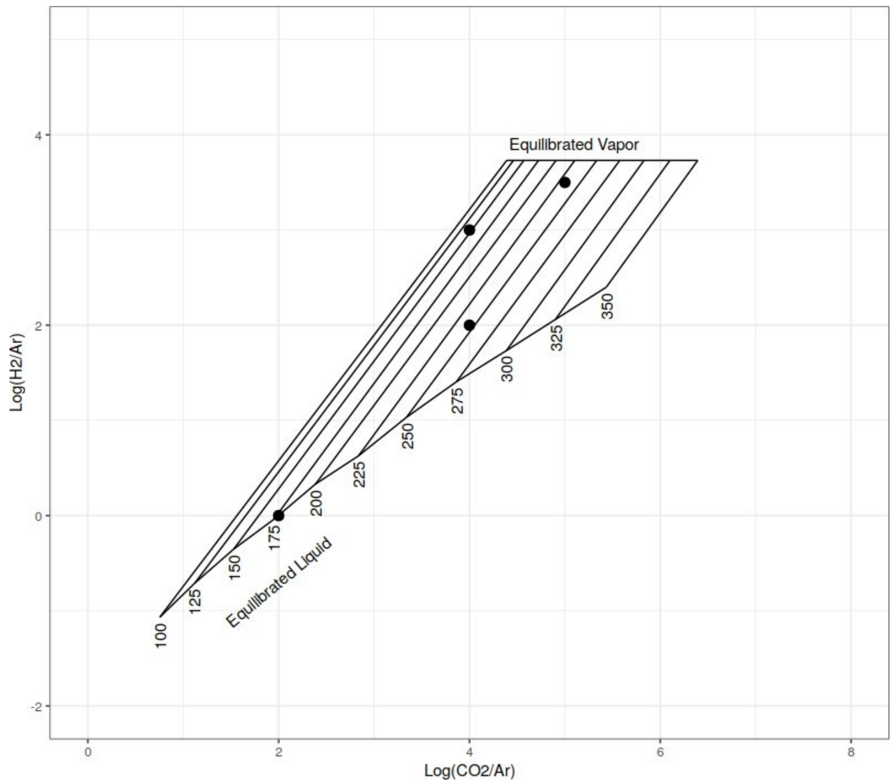

Figure 5. Example of CAR-HAR geothermometer with grid data from Giggenbach and Glover (1992).

3. DISCUSSION AND CONCLUSIONS

The GQAnalyzer package developed in the programming language R offers an open and extensible platform for the creation of specialized plots routinely used in the analysis of chemical composition of groundwater and geothermal fluids and gases. This package is distributed under a GPL License as are most R packages, which allows users to reuse the code and share their modifications to a broader audience. The audience of the GQAnalyzer package is composed of two large groups. The first group includes geoscientists with a basic knowledge of programming in R and the graphical library ggplot2 can extend the functionality of the package to include additional plots and/or additional analysis tools such as mixing calculations, Compositional Statistical Analysis, mapping using geostatistics, among others. The second group of users corresponds to those geoscientists who are not comfortable typing on the command line or writing scripts for whom a Graphical User Interface has been developed. This implies that the package can have a large audience in the Earth Sciences community. In addition, the package GQAnalyzer can be used as an instruction tool for classes in groundwater, geothermal resources and programming for Earth Sciences.

The package GQAnalyzer is an alternative to the commonly used spreadsheets proposed by Powell and Cummins (2010). The comparison between a statistical programming language such as R and a spreadsheet program such as Excel (c) is not fair. R has large graphical and statistical capabilities when compared to Excel (c) but its main disadvantage refers to the learning curve of R which is of intermediate difficulty. This issue is partially solved with the GUI included in the package.

To the best knowledge of the authors, the package GQAnalyzer is the first package completely designed in R for the analysis of the chemical composition of geothermal fluids and gases using the state-of-the-art plotting capabilities offered by this programming language. In addition, R has a large number of packages that can support additional specialized analysis such as geospatial analysis, mapping capabilities, multivariate statistical analysis, Compositional Data Analysis, Data Mining and Machine Learning. This makes it possible to integrate and analyze information from different sources as required in the exploration program for geothermal resources.

The package GQAnalyzer can be found at www.github.com/khaors/GQAnalyzer/.

REFERENCES

Ahmad, N., Sen, Z., & Ahmad, M., (2003). Ground Water Quality Assessment Using Multi-Rectangular Diagrams. Ground Water, 41(6), 828–832. https://doi.org/10.1111/j.1745-6584.2003.tb02423.x

Appelo, T. and Postma, V., (2005). Geochemistry, Groundwater and Pollution. Second Edition. CRC Press.

Arnórsson, S., Gunnlaugsson, E., Svavarsson, H., (1983). The chemistry of geothermal waters in Iceland. III. Chemical geothermometry in geothermal investigations. Geochim. Cosmochim. Acta 47, 567–577.

Arnórsson, S. (Ed.)., (2000). "Isotopic and Chemical Techniques in Geothermal Exploration, Development and Use. Sampling Methods, Data Handling, Interpretation," International Atomic Energy Agency, Vienna, 351p.

Can, I., (2002). A new improved Na/K geothermometers by artificial neural networks. Geothermics 31, 751–760.

Díaz-González, L., Santoyo, E., Reyes-Reyes, J., (2007). Tres nuevos geotermómetros mejorados de Na/K usando her- ramientas computacionales y geoquimiométricas: aplicación a la predicción de temperaturas de sistemas geotérmicos. Rev. Mex. Cienc. Geol. 25.

Fournier, R.O., (1977). Chemical geothermometers and mixing models for geothermal systems. Geothermics 5, 4150.

Fournier, R.O., (1979). A revised equation for the Na/K geothermometer. Geotherm. Res. Coun. Trans. 3, 221224.

Fournier, R.O., (1991). Water geothermometers applied to geothermal energy. In: D'Amore, F. (Ed.), Applications of Geochemistry in Geothermal Reservoir Development. UNITAR/UNDP Centre on Small Energy Resources, Rome, Italy, pp. 37–69.

Fournier, R.O., Potter II, R.W., (1982). A revised and expanded silica (quartz) geothermometer. Geotherm. Res. Coun. Bull. 11, 312.

Fournier, R.O., Truesdell, A.H., (1973). An empirical Na–K–Ca geothermometer for natural waters. Geochim. Cosmochim. Acta 7,1255–1275.

Giggenbach, W. F., (1988). Geothermal solute equilibria. Derivation of Na-K-Mg-Ca geoindicators. Geochimica et Cosmochimica Acta, 52(12), 2749–2765. https://doi.org/10.1016/0016-7037(88)90143-3

Giggenbach, W. F., & Glover, R. B., (1992). Tectonic regime and major processes governing the chemistry of water and gas discharges from the rotorua geothermal field, New Zealand. Geothermics, 21(1–2), 121–140. https://doi.org/10.1016/0375-6505(92)90073

Kharaka, Y.K., Mariner, R.H., (1989). Chemical geothermometers and their application to formation waters from sedimen- tary basins. In: Naeser, N.D., McCulloh, T.H. (Eds.), Thermal History of Sedimentary Basins. Springer-Verlag, New York, NY, USA, pp. 99–117.

Nicholson, K., (1993). Geothermal fluids: Chemistry and Exploration Techniques.

Piper, A. M., (1944). A graphical procedure in the geochemical interpretation of water analyses. Trans. American Geophysics Union Pp 914-923

Powell, T., & Cumming, W., (2010). Spreadsheets for Geothermal Water and Gas Geochemistry. Proceedings XXXV Workshop on Geothermal Reservoir Engineering, 9(1174-5878 (Print).

Schoeller, H., (1962). Leseaux Souterraines. Mason and Cie. Paris 642p.

Stiff, H. A., (1951). Interpretation of Chemical water analysis by means of pattern. Journal of petrol Technology Vol. 3 No 10 Pp 15-17.

Tonani, F., (1980). Some remarks on the application of geochemical techniques in geothermal exploration. In: Proceedings of 2nd Symposium Advances in European Geothermal Research, Strasbourg, France, pp. 428–443.

Truesdell, A.H., (1976). GEOTHERM, a geothermometric computer program for hot spring systems. In: Proceedings of Second U.N. Symposium on the Development and Use of Geothermal Resources 1975, San Francisco, CA, USA, pp. 831–836.

Verma, S.P., Santoyo, E., (1997). New improved equations for Na/K, Na/Li and SiO2 geothermometers by outlier detection and rejection. J. Volcanol. Geotherm. Res. 79, 9–23.

Verma, M.P., (2000). Revised quartz solubility temperature dependence equation along the water–vapor saturation curve. In: Proceedings of the 2000 World Geothermal Congress. Kyushu, Tohoku, Japan, pp. 1927–1932.

Wickham, H., (2015). R Packages. O'Reilly Media,

Wickham, H., (2016). Ggplot2: Elegant Graphics for Data Analysis. Second Edition. Springer Verlag.