



## Métodos de la estadística multivariada para la agrupación hidroquímica de aguas naturales y mineromedicinales de la Sierra del Rosario

Patricia González Hernández<sup>(1)</sup>, Margaret Suárez Muñoz<sup>(1)</sup>, Juan R. Fagundo Castillo<sup>(1)</sup>, Clara Melián Rodríguez<sup>(1)</sup>, Barbara Luna Saucedo<sup>(2)</sup>, Maritza Llerena Portilla<sup>(3)</sup>, Joel Fagundo Sierra<sup>1</sup>

(1) Centro Nacional de Termalismo "Víctor Santamarina". Ave: 243, No 19815, Fontanar, Municipio Boyeros Ciudad Habana. C.P 19250. E-mail: [Patricia@rsrch.isctn.edu.cu](mailto:Patricia@rsrch.isctn.edu.cu)

(2) Centro Nacional de Investigaciones Científicas,. Avenida 25 esquina 158. Cubanacán, Municipio Playa. E-mail: [luna@química.cneuro.edu.cu](mailto:luna@química.cneuro.edu.cu)

(3) Instituto de Geografía Tropical. Calle G esquina 13. Municipio Vedado, Ciudad La Habana. E-mail: [Maritza@igt.edu.cu](mailto:Maritza@igt.edu.cu)

### Resumen

El análisis estadístico multivariado tiene la propiedad de reunir y sintetizar grandes conjuntos de datos y variables, en función de ciertos objetivos, con el fin de obtener informaciones sólidas, que permitan una mejor comprensión del fenómeno objeto de estudio. En las Ciencias de la Tierra estos métodos han sido denominados geomatemáticos por algunos investigadores. Su aplicación en el estudio de las aguas minerales, ha facilitado la interpretación del comportamiento hidrogeoquímico de los manantiales termominerales.

En este trabajo, aplicando métodos de la estadística multivariada, específicamente los métodos de análisis factorial y de análisis de cluster, se tiene como objetivo una agrupación de diferentes tipos de aguas de la Sierra del Rosario, así como de diferentes patrones hidroquímicos

característicos de esta zona de estudio.

Para ello se han tomado datos hidrogeoquímicos de muestras representativas correspondientes a un trabajo de exploración detallada llevado a cabo en el sector hidrogeológico San Diego de los Baños– Bermejales y a un muestreo realizado en el manantial Mil Cubres. Con los mismos fue creada una matriz constituida por 22 individuos y 14 variables (Temperatura, pH, TSS,  $\text{CaCO}_3$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{H}_2\text{SiO}_3$ ,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+ + \text{K}^+$ ). Las mediciones de los parámetros geoquímicos se realizaron "in situ". Los contenidos de  $\text{CO}_2$  y  $\text{H}_2\text{S}$ , así como la alcalinidad total ( $\text{HCO}_3^-$  y  $\text{CO}_3^{2-}$ ) fueron también determinados "in situ", mientras que los restantes macroconstituyentes ( $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$  y  $\text{Mg}^{2+}$ ) y componentes trazas se analizaron en el laboratorio antes de las 24 horas de tomadas las muestras.

Esto permitió la identificación de las diferentes facies hidroquímicas que forman parte de los miembros extremos de mezcla (aguas  $\text{HCO}_3\text{-Ca}$ ,  $\text{SO}_4\text{>HCO}_3\text{-Na}$  y  $\text{SO}_4\text{-Ca}$ ) y de las facies hidroquímicas que se manifiestan como resultado de la mezcla en diferentes porcentajes de los miembros extremos de mezcla (aguas  $\text{HCO}_3\text{-Ca}>\text{Na}$ ,  $\text{SO}_4\text{>HCO}_3\text{-Ca}$ ,  $\text{SO}_4\text{-Ca}>\text{Na}$ , y  $\text{SO}_4\text{>HCO}_3\text{-Ca}>\text{Na}$ ), todo ello relacionado con la geología y las características de la región objeto de estudio. También permitió la agrupación de los diferentes tipos de aguas y patrones hidroquímicos característicos de la Sierra del Rosario, y la visualización de las diferentes facies de mezclas presentes en esta zona de estudio. Obteniéndose dos factores significativos que dan como resultado estas facies hidroquímicas: un primer factor que relaciona los parámetros físico-químicos que caracterizan a las aguas bicarbonatadas de baja mineralización (acuífero somero de la Formación Artemisa; esquistos y areniscas de la Formación San Cayetano) y un segundo factor que relaciona los parámetros físico-químicos que caracterizan a las aguas del tipo sulfatadas cálcicas (del drenaje profundo).

#### Abstract

The multivariate statistical analysis has the property to gather and to synthesize big groups of data and variables, in function of certain objectives, with the purpose of obtaining solid information that allow a better understanding of the phenomenon. In the Sciences of Earth these methods has been denominated geomathematical by some investigators. Their application in the study of the mineral waters has facilitated the interpretation of the hydrogeochemical behavior of the thermal mineral springs.

In this work, the multivariate statistical methods, specifically the factorial analysis methods and the cluster analysis are used, with the purpose of grouping different types of waters of the Rosario's Sierra, as well as of different hydrochemical patron characteristic of this area.

To complete the main objective, it was analyzed an hydrogeochemical data of representative samples corresponding to a work of detailed exploration carried out in the hydrogeological sector of San Diego de los Baños– Bermejales and to a sampling carried out in the spring Mil Cumbres. With this data a matrix constituted by 22 individuals and 14 variables was created (the variables chosen were Temperature, pH, TSS,  $\text{CaCO}_3$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{H}_2\text{SiO}_3$ ,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+ + \text{K}^+$ ). The measure of the geochemical parameters was carried out "in situ." The contents of  $\text{CO}_2$  and  $\text{H}_2\text{S}$ , as well as the total alkalinity ( $\text{HCO}_3^-$  and  $\text{CO}_3^{2-}$ ) were also determined "in situ", while the macrocomponents ( $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ) and trace components were analyzed in the laboratory before the 24 hours of having taken the samples.

This allowed the identification of the different hydrochemical facies which are part of the extreme members of mixture ( $\text{HCO}_3\text{-Ca}$ ,  $\text{SO}_4 > \text{HCO}_3\text{-Na}$  and  $\text{SO}_4\text{-Ca}$  waters) and of the hydrochemical facies that are manifested as a result of the mixture in different percentages of the extreme members of mixture ( $\text{HCO}_3\text{-Ca} > \text{Na}$ ,  $\text{SO}_4 > \text{HCO}_3\text{-Ca}$ ,  $\text{SO}_4\text{-Ca} > \text{Na}$ , and  $\text{SO}_4 > \text{HCO}_3\text{-Ca} > \text{Na}$  waters). This different groups obtained are related with the geology and the characteristics of this region. It is also group different types of waters and hydrochemical patron characteristic of the Rosario's Sierra, and the differentiation of the facies of mixtures present in the area. There are obtained two significant factors that give, as a result, these hydrochemical facies: a first factor that relates the physical-chemical parameters that characterize the bicarbonated waters of low mineralization (aquifer of the Artemisa Formation; schists and grit of San Cayetano Formation) and a second factor that relates the physical-chemical parameters that characterize the calcic sulfated waters (of the deep drainage).

## Introducción

El Balneario San Diego de los Baños se encuentra ubicado junto al pueblo del mismo nombre, municipio Los Palacios, provincia de Pinar del Río (Cuba), en las coordenadas N 315 037 E 256 008 (Fig. 1). Comprende tres

manantiales fundamentales: El Tigre, El Templado y La Gallina, los cuales brotan en el cauce del río San Diego de los Baños a unos 7 Km aguas arriba alumbran, en el propio cauce, varios manantiales denominados M1, M2, M3, M4 y M5 (BM, Fig 1) (coordenadas N 318 351 E 250 468). Todas estas manifestaciones termales están asociadas a las fallas Pinar y San Diego de los Baños. La primera falla separa las zonas estructuro-faciales Guaniguanico y San Diego de los Baños, mientras que la segunda falla separa las dos subzonas en que se divide la unidad tectónica Guaniguanico (Sierra de los órganos y Sierra del Rosario).

#### Métodos de estadística multivariada

El análisis estadístico multivariado tiene la propiedad de reunir y sintetizar grandes conjuntos de datos y variables, en función de ciertos objetivos, con el fin de obtener informaciones sólidas, que permitan una mejor comprensión del fenómeno objeto de estudio (González, 1997). En las Ciencias de la Tierra estos métodos han sido denominados geomatemáticos por algunos investigadores. Su aplicación en la Hidroquímica ha permitido determinar el comportamiento de los acuíferos cársicos (Valdés et. al., 1981) y de otros tipos de acuíferos ante ciertas acciones externas (Pulido-Bosch, 1989; Dimitrov et. al., 1993). En el caso de las aguas minerales, su empleo ha facilitado la interpretación del comportamiento hidroquímico de los manantiales termominerales (López-Chicano y Pulido-Bosch, 1996). En uno u otro caso, se han podido estimar modelos de comportamiento, transformar las variables y reducir dimensionalidad, comprobar o rechazar hipótesis y dar respuesta a diferentes aspectos de carácter práctico.

Entre los métodos de análisis multivariados, los más comúnmente empleados han sido los siguientes:

- Análisis de cluster o de enjambres, para agrupar las variables hidroquímicas que tienen características similares. Este método consiste en seleccionar las mayores similitudes (correlaciones) entre parejas de variables (Pulido-Bosch, 1989). De esta manera, se forman las primeras parejas que presentan una mayor similitud entre las mismas. A continuación se genera una nueva matriz de similitud que ya incluye las agrupaciones previas, como si se tratara de una sola variable (Callejón et al., 1984). Una vez agrupadas todas las variables, el gráfico que resulta (dendograma) permite visualizar en forma rápida los grupos de variables afines y distinguirlos de aquellos que apenas guardan relación con el resto.

- Análisis factorial con el objetivo de encontrar nuevas variables (denominadas factores), que serán una combinación lineal de las primeras y permiten estudiar con más facilidad, y sin pérdida significativa de la información original, el fenómeno a analizar, su estructura, y evidenciar relaciones entre variables, entre relaciones, o entre ambas (González, 1997). El procedimiento en el análisis factorial parte de la obtención de  $n$  autovalores y autovectores (componentes principales) de la matriz  $n \times n$  correlaciones y de la estimación de la fracción de la varianza total de los

datos explicados por cada uno de tales componentes (Pulido-Bosch, 1989).

- Análisis de series temporales, para estudiar la variación en el tiempo de los datos. En el caso de la Hidroquímica, estos pueden ser de salinidad, mineralización o de algún ion de interés. Estas técnicas permiten estudiar la evolución de la variable en un determinado intervalo, determinar tendencias y hacer pronósticos, así como analizar las relaciones temporales entre las variables.

- Análisis correlatorio y espectral, para identificar la estructura de las series cronológicas empleadas, bien de forma individual o bien como relación causa-efecto, con lo que se puede deducir el mecanismo de funcionamiento del acuífero e interpretar el fenómeno de salinización del mismo (González, 1997).

## Materiales y Métodos

Los procesos de interacción agua-roca y de mezcla de aguas que tienen lugar en el sistema hidrotermal de San Diego de los Baños, responsables de las diferentes facies hidroquímicas allí presentes son estudiados bajo la óptica de los análisis matemáticos multivariados, principalmente, análisis de cluster y análisis factorial.

Del total de muestras correspondientes al trabajo de exploración detallada llevado a cabo en el sector hidrogeológico San Diego de los Baños– Bermejales (117 datos) y el muestreo realizado en el manantial Mil Cubres (19 datos) por Fagundo et. al. (1993), fueron seleccionados 22 datos representativos de los tipos de aguas presentes en esos sitios. Con los mismos fue creada una matriz constituida por 22 individuos y 14 variables (Temperatura, pH, TSS,  $\text{CaCO}_3$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{H}_2\text{SiO}_3$ ,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+ + \text{K}^+$ ). La misma fue procesada mediante el sistema Statistica (Stat Sof Inc., 1993).

Las mediciones de los parámetros geoquímicos se realizaron "in situ" mediante pHmetro y medidor de temperatura y potencial redox (Eh) modelo HI-8424 de la marca HANNA. Los contenidos de  $\text{CO}_2$  y  $\text{H}_2\text{S}$ , así como la alcalinidad total ( $\text{HCO}_3^-$  y  $\text{CO}_3^{2-}$ ) fueron también determinados "in situ", mientras que los restantes macroconstituyentes ( $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$  y  $\text{Mg}^{2+}$ ) y componentes trazas se analizaron en el laboratorio antes de las 24 horas de tomadas las muestras. Las marchas analíticas se efectuaron mediante las técnicas analíticas estándar (APHA, AWWA, WPCF, 1992).

## Resultados y Discusión

En la figura 2 se presentan los resultados de la clasificación mediante análisis de cluster, realizada utilizándose el método de unión simple y como medida de similitud, la distancia euclidiana. Se aprecian fundamentalmente

dos grupos. El primer grupo numérico está formado por las variables Temperatura,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$  y  $\text{H}_2\text{SiO}_3$ , lo cual indica el aporte del ambiente de las aguas minerales del tipo sulfatadas cálcicas. El segundo grupo está asociado a las variables pH,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{Cl}^-$ , TSS,  $\text{Mg}^{2+}$ ,  $\text{CO}_2$  y  $\text{Na}^+ + \text{K}^+$ , lo cual es indicativo de la influencia tanto del acuífero cársico somero desarrollado en la Formación Artemisa (aguas bicarbonatadas cálcicas), como del medio no cársico representado por las areniscas y esquistos de la Formación San Cayetano (aguas bicarbonatadas sódicas).

En el dendrograma correspondiente a la clasificación de los casos (Fig. 3), se observan tres grupos separados entre sí a un nivel de similitud muy bajo, en los cuales se encuentran agrupadas las aguas de los manantiales El Tigre, El Templado, P19 y La Gallina (aguas sulfatadas bicarbonatadas cálcicas y sulfatadas cálcicas); los manantiales de Bermejales junto dos muestras del pozo 1, una del pozo 4 y otra del pozo 12 (aguas sulfatadas cálcicas y sulfatadas cálcicas sódicas), y varias muestras correspondientes al manantial Mil Cumbre, una muestra del pozo y muestras de los pozos 2, 3, 4 y 13 (aguas bicarbonatadas cálcicas y bicarbonatadas cálcicas sódicas). Finalmente, se presentan agrupados entre sí y con una disimilitud muy acentuada con respecto a los restantes grupos, dos muestras del pozo 4 (aguas bicarbonatadas sódicas).

A través del análisis factorial de variables (Fig. 4), se llega a los siguientes resultados:

- Aparecen definidos 2 factores que controlan las propiedades químico-físicas de las variables analizadas y que explican el 48.4 y 24.1 % de varianza total que implica un 72.5 % de la varianza acumulada. Para el total de datos utilizados (N=22), se consideran significativos los coeficientes de correlación mayores de 0.6 entre las variables y los factores (Tab. I). Así, presentan correlaciones (directas) significativas con el primer factor las variables TSS,  $\text{CO}_2$ ,  $\text{SO}_4^{2-}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ , Temperatura,  $\text{CaCO}_3$ ,  $\text{H}_2\text{SiO}_3$  y pH con signo negativo (en correspondencia con los equilibrios químicos), mientras que con respecto al segundo factor presentan correlaciones (inversas) significativas  $\text{CO}_3^{2-}$ ,  $\text{HCO}_3^-$ ,  $\text{Cl}^-$  y  $\text{Na}^+ + \text{K}^+$ .

- El primer factor (F1) está relacionado con el drenaje profundo del medio cársico (calizas fértidas de la Formación Artemisa), al cual se asocian fundamentalmente las variables señaladas junto a la Temperatura. En su conjunto, estas variables son características de las aguas termominerales.

- El segundo factor (F2) parece representar en la parte negativa del eje, el medio no cársico (Formación San Cayetano), al cual se asocian las variables  $\text{Na}^+ + \text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{HCO}_3^-$  y  $\text{CO}_3^{2-}$ . La variable  $\text{H}_2\text{S}$ , no relacionada con ninguno de los dos factores, puede ser interpretada a partir del aporte de las piritas y calcopiritas (polimetales) presentes en la Formación San Cayetano.

Tabla I. Correlación entre los factores y las variables seleccionadas y porcentaje de varianza extraída por los factores principales.

	Factor 1	Factor 2
Temperatura	.670300	-.332906
PH	-.694345	-.451322
TSS	.894376	-.352898
CO <sub>2</sub>	.711977	.235433
HCO <sub>3</sub>	-.297662	-.600000
CO3	-.501287	-.783327
CL	-.050714	-.859324
SO <sub>4</sub>	.938829	-.145137
Ca	.958817	.005728
Mg	.899041	-.205828
Na+K	-.133469	-.856495
CACO <sub>3</sub>	.963609	-.027339
H <sub>2</sub> S	.139491	-.543645
H <sub>2</sub> SiO <sub>3</sub>	.786784	-.314189
Expl.Var	6.774545	3.374026
Prp.Totl	.483896	.241002

En la figura 5 se muestra la posición que ocupan los datos con respecto a los dos factores principales. En general, se aprecian 3 grupos de aguas con composiciones extremas, ubicadas el primero, en la parte positiva del eje correspondiente al factor I (representado por las aguas mineromedicinales del pozo 1 (P1 237-320); el segundo situado en la parte negativa del eje correspondiente al factor II, constituido por muestras del pozo 4 (Bp4 103-206) (agua bicarbonatadas sódicas), y un tercer grupo, ubicado en la parte

positiva de este mismo eje, con muestras del manantial Mil Cumbre y de los pozos 2 y 3 (Mil Cumbre, P2 90m y P3 0-90m) (aguas bicarbonatadas cálcicas).

Entre estos tres grupos principales se sitúan las restantes muestras, lo cual indica contribuciones en mayor o menor proporción de los miembros extremos. Así las aguas de los manantiales El Tigre, El Templado y La Gallina se ubican entre las posiciones que ocupan las aguas sulfatadas cálcicas y bicarbonatadas cálcicas, mientras que los manantiales de Bermejales, el Pozo 12 y una muestra del pozo 4 (145-195 m), las cuales se corresponden con aguas con un relativamente alto contenido de sulfato, calcio y sodio, se sitúan en un espacio intermedio entre los sitios que ocupan las aguas del tipo sulfatadas cálcicas y las aguas del tipo bicarbonatadas sódicas. Finalmente, el pozo 13 del tipo de agua bicarbonatado, cálcico, sódico, ocupa una posición entre las aguas bicarbonatadas cálcicas y las bicarbonatadas sódicas. Todos estos resultados son coherentes con los obtenidos por análisis de cluster y otros métodos empleados.

## Conclusiones

El análisis multivariado, específicamente los métodos del análisis factorial y análisis de cluster utilizados en el trabajo que se presenta, permite realizar una agrupación de los diferentes tipos de aguas de la Sierra del Rosario, así como de los diferentes patrones hidroquímicos característicos de la zona de estudio y la visualización de los diferentes tipos de aguas de mezclas presentes en la misma.

Por una parte si se tienen en cuenta las variables físico-químicas estudiadas, se obtiene dos factores significativos: un primer factor que relaciona los parámetros físico-químicos que caracterizan a las aguas bicarbonatadas de baja mineralización (acuífero somero de la Formación Artemisa; esquistos y areniscas de la Formación San Cayetano) y un segundo factor que relaciona los parámetros físico-químicos que caracterizan a las aguas del tipo sulfatadas cálcicas (del drenaje profundo).

Mediante el análisis de los casos (las aguas de los diferentes manantiales y pozos) se muestra que en la zona de estudio las facies hidroquímicas son las siguientes: bicarbonatadas cálcicas, bicarbonatadas sódicas y sulfatadas cálcicas como miembros extremos de mezcla; y bicarbonatadas cálcicas sódicas, sulfatadas bicarbonatadas cálcicas, sulfatadas cálcicas sódicas y sulfatadas bicarbonatadas cálcicas sódicas originadas por la mezcla de los diferentes porcentajes de los miembros extremos de mezcla antes mencionados.

## Bibliografía

APHA – AWWA – WPCF (1992). "Métodos Normalizados para el análisis de Aguas Potables y Aguas Residuales, Am. Public Assoc., Edición 17<sup>th</sup>. Editorial Grijalbo (Madrid), 4-1 – 4-235.

Callejón, S.M., A. Pulido-Bosch y P. Valenzuela (1984). "Aplicación de los análisis de cluster al estudio de características físico – químicas de aguas subterráneas". *Estudios Geol. (Madrid, España)*, 40: 193-200.

Dimitrov, D., B. Velikov and M. Machcova (1993). "Processing of ground water hydrochemical data by means of cluster and discriminant analysis". *Hidrogeología (Granada, España)*, 8: 25-39.

Fagundo, J.R., J. Rodríguez, J.M. Pajón, E. Franco, G. Benítez, A.C., J. Guerón, I. Abelló (1993). "Caracterización hidroquímica de las aguas del Pan de Guajaibón y otras áreas cársicas cercanas en la Sierra del Rosario". En: Libro de Comunicaciones. I Taller Internacional sobre Cuencas Experimentales en el Karst. Ed. H.J. Llanos, I. Antiguada, I. Morell y A. Eraso, Universidad Jaime I, Castellón (España). 43-53.

González, A. (1997). "Problemas de salinización en el acuífero litoral del occidente de Huelva". Ed.: Universidad de Huelva Publicaciones (España), 450 Págs.

López-Chicano, M. y A. Pulido-Bosch (1996). "Observaciones hidrogeológicas e hidroquímicas sobre manantiales termominerales de Alhama de Granada (Cordilleras Béticas. España)". *Geogaceta (España)*, 19: 134-137.

Pulido-Bosch, A. (1989). "Caracterización Hidroquímica del Campo de Dalias (Almería)". Ed.: Instituto Andaluz de Reforma Agraria, Junta de Andalucía, España, 256 Págs.

Valdés, J.R. Fagundo y J.M. Pajón (1981). "Aplicación de modelos numéricos el estudio hidrogeoquímico de las aguas del río Cuyaguatete". *Ingeniería Hidráulica. La Habana*, 2 (3): 215-286.

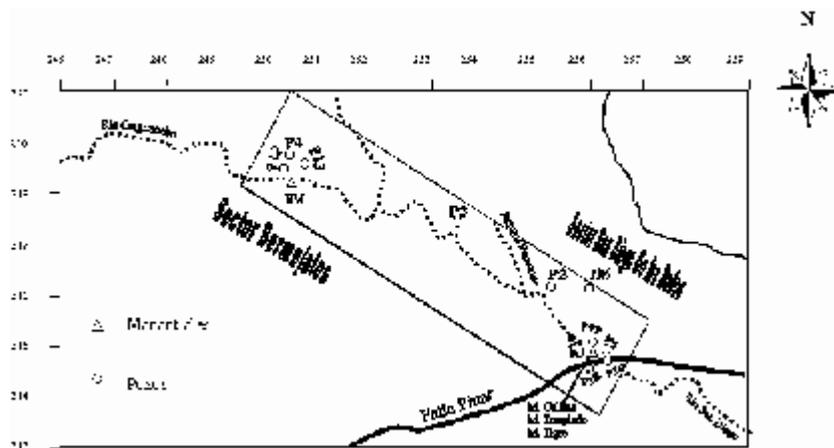


Figura 1. Sitios de muestreo en el sector hidrogeológico San Diego de los Baños-Bermemjales

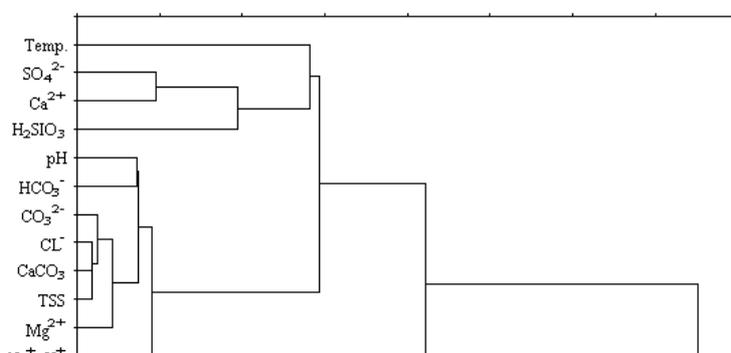




Figura 5. Distribución de los datos en el plano definido por los factores I y II.

- [Breve historia de los Congresos de Geología en Cuba.](#)
- [Junta Directiva Nacional](#)

[\[PRESENTACION\]](#) [\[POR TEMATICA\]](#) [\[POR AUTOR\]](#) [\[CORREO ELECTRONICO\]](#)  
[\[MAPA del SITIO\]](#)

